Hojjatollah Farahani · Marija Blagojević · Parviz Azadfallah · Peter Watson · Forough Esrafilian · Sara Saljoughi

An Introduction to Artificial Psychology

Application Fuzzy Set Theory and Deep Machine Learning in Psychological Research using R

Springer

An Introduction to Artificial Psychology

Hojjatollah Farahani • Marija Blagojević
Parviz Azadfallah • Peter Watson
Forough Esrafilian • Sara Saljoughi

An Introduction to Artificial Psychology

Application Fuzzy Set Theory and Deep Machine Learning in Psychological Research using R

Springer

Hojjatollah Farahani
Tarbiat Modares University
Tehran, Iran

Parviz Azadfallah
Tarbiat Modares University
Tehran, Iran

Forough Esrafilian
Tarbiat Modares University
Tehran, Iran

Marija Blagojević
University of Kragujevac
Kragujevac, Serbia

Peter Watson
University of Cambridge
Cambridge, UK

Sara Saljoughi
Tarbiat Modares University
Tehran, Iran

ISBN 978-3-031-31171-0 ISBN 978-3-031-31172-7 (eBook)
https://doi.org/10.1007/978-3-031-31172-7

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

This book is dedicated to thinkers whose worthy thoughts never saw the light of day

Online Resources

Online resources and data are available to aid your study and help you for further understanding and in-depth learning.

Check it out and test yourself!

Preface

Don’t You Know the Statistics? Don’t be Afraid and Enter

In science we try to build up a picture of the world. With the advancement of science, certainty has been isolated day by day and dark clouds have cast doubt on its beauty and radiance, and the certainty of Holy Grails turned upside down. Scientists have worked hard to push certainty out of the heart of mathematics. This effort was not very successful, and the advancement of science did not only increase human knowledge of phenomena but also expanded our knowledge of the limitations of scientific knowledge.

It shook the body of certain scientific principles, and the use of most of the tools of world representation have been shown to be problematic.

Until the twentieth century, science dealt with the certainty of Aristotelian logic. Now there was a world in which something could be both A and Not-A. Thus the death knell rang for the dominant 2000-year-old Aristotelian thought and instead of a comprehensive description of the world by drawing fixed and unique maps whose lines were in harmony with the day-to-day world, science made a set of maps, maps that were different but which gave a flexible understanding of the world.

Science is emerging in the postmodern world where scientific reading is not a fixed machine, rather it is creative and dynamic and readers create meaning in their minds based on thoughtful or lived experiences. Therefore, an author is no longer the final reference and refuge of the work or the only true creator. The reader does not suffer in passive silence in the ruthless wave of information, rather he is the one who gives life to the text. The data is silent and vague, and the mind of the reader ignites and speaks to them. In this new scientific worldview, we are not neutral observers of phenomena, which is neither possible nor desirable, but have become participants in the world.

Bohr, one of the great thinkers of quantum physics, believes that measurement questions the world and that question changes the answers. So what we are seeing are not phenomena, but, as Heinberg puts it, the genius of physics. What we see is not reality, but reality that is intertwined with the method of studying it. So we can say that what we see is the product of the science of measurement. In other words, the process of recognizing a phenomenon causes it to occur, but to what extent? We do not know. With this view, certainty is slowly receding from the world of science, and uncertainty is rising above thought. Psychological scientists, who are skilled in crossing the boundaries of science and brain activity, are not spared from this dramatic change. They found inadequate behavior and adopted a qualitative methodology for trapping elusive, dynamic, and meaningful phenomena. As we get closer to reality, precision decreases and vice versa.

Although a group with flexibility and scientific breadth combined these two methodologies, this hybrid was not inherited certainty at all, and this method gave rise to inherent uncertainty of the phenomena, and the struggle between precision and reality continued.

David Appelbaum has written a beautiful book called *The Stop*, in which he made a subtle statement which we express here. In this book, he compares the speed of seeing and examining things as experienced by a blind person by touch. Although the blind person cannot see, he says that this not-seeing also has its advantages, because the blind person sees things that a seeing person does not because he moves cautiously and slowly and, as a result, he finds his way through obstacles with a different touch and method and gains a new understanding. Perception is a trace of a hidden meaning; a perception often deepens, albeit slowly. So it can be said that although fast methods of understanding have their advantages, standing and looking calmly and evolving perception is also important. More attention should be paid to the development of a new methodology and extension of the current one. Therefore, it is hoped that mind researchers, psychologists, and all those with a research mindset will use multifaceted methods.

Behavioral scientists therefore have a reliable basis for relying and looking at psychological phenomena to be able to grasp the pearl of truth, but there is doubt about the nature of truth. The transition from classical (nineteenth century) certainty to the uncertainty of the present century was necessary as understanding evolved. This transition was not straightforward but enabled us to get nearer to the truth.

Developing theories in quantitative and qualitative research is the main purpose of this book.

We know that researchers' inference involves probability and uncertainty, so we talk about decay of the Holy Grail of perceived truths. It should be said that probability is an uncertainty about the existence or non-existence of phenomena, while what is examined in this book is a representation of another type of uncertainty, that is, the degree of certainty about existence or non-existence of phenomena rather than naively assuming a phenomenon's non-existence.

Most of us, however, have the same uncertainty in terms of the likelihood of the results obtained by doing research in psychology. We are aware of research findings but do not always believe that its results present a true understanding of psychological problems. However, the goal is not to reduce their value in any way. Because uncertainty is everywhere!!

Inference was created to model human thought, in order to obtain more dynamic and consistent findings with reality by using approximate and fuzzy reasoning.

The science of psychology is full of linguistic variables that require an approximate, indefinite, and obscure method of their own kind so that we do not fall into imitation models in studying the psychological phenomena of these multi-faceted concepts. These multi-faceted concepts may be interpreted and examined in different ways which is fine. What is annoying, however, is the dominance of a particular method or a particular way of thinking and its application to the inference of data in research.

In this book, we have considered the methodology of approximate inference in psychological research from a theoretical and practical perspective. Quantitative variable-oriented methodology and qualitative case-oriented methods are both used to explain the set-oriented methodology which we call fuzzy psychology. As stated in the opening sentence of the book, it does not matter if you do not know much about mathematics or statistics, because statistical and mathematical intuitions are key here and they will be learned through practice. What is important is to understand the method and its application to new, dynamic, and elusive phenomena.

Finally, your comments on this book are very welcome, so please do not hesitate to share them with us.

In the end, remember Montagne's short but deep saying "What do we know?" The human world is indeed full of uncertainty, whose beauty we have not been able to define and explain.

Acknowledgments

We would like to thank the team at Springer Nature for their support in bringing this book to fruition. Our thanks go to Megan McManus for supporting our ideas at the outset and suggesting further topics to explore, Pradheepa Vijay for reviewing the content and her important comments about the layout of the book, Tessy Priya for her encouragement and help with final editing of the proofs and Brian Halm for co-ordinating and explaining the publishing procedures to us.

Contents

About the Authors

Hojjatollah Farahani was born on September 23, 1978, in Iran, and he is an Assistant Professor at the Tarbiat Modares University (TMU). He received his Ph. D. from Isfahan University in 2009, and he was a Postdoctoral Researcher in Fuzzy inference at the Victoria University in Australia (2014–2015), where he started working on Fuzzy Cognitive Maps (FCMs) under supervision of professor Yuan Miao. He is the author or co-author of more than 150 research papers and a reviewer in numerous scientific journals. He has supervised and advised many theses and dissertations in psychological sciences. His research interests and directions include psychometrics, advanced behavioral statistics, fuzzy psychology, artificial intelligence, and machine learning algorithms in psychology.

Email: h.farahani@modares.ac.ir

Marija Blagojević is an Associate Professor at the Faculty of Technical Sciences Čačak, Department of Information technology, University of Kragujevac. Born in 1984, she obtained her M.Sc. degree in 2007 and Ph.D. degree in Technical Sciences in 2014. Her current research interests include data mining, artificial neural networks, e-learning, and programming. With 15 years of experience in teaching and research activities, she is the author or co-author of more than 80 research papers, patents, and technical solutions and a reviewer in numerous scientific journals. She is also a reviewer of projects and study programs and a member of partner universities for Tempus and Erasmus projects. She was a member of the organizational and scientific committees of several conferences and a member of the Commission for self-evaluation at the faculty. She was project coordinator for one innovative project in the domain of technology transfer.

Parviz Azadfallah has a Ph.D. in Health Psychology and became qualified in several international academic courses in Neuropsychology and Psychodynamic Psychotherapies. He started his scientific activity at the Psychology Department of Tarbiat Modares University (TMU) in 1999. He conducted many research projects and supervised 45 PhD theses and more than 50 Master dissertations from 1999 to

2023. He is the author or co-author of more than 250 research papers. He is also the author of some books, and his last book, *Embodied Ego*, uses a developmental approach to explain mind-body integration. He founded the Neuropsychology Laboratory of TMU in 2017. He is also a member of the Cognitive Neuroscience Department at TMU since 2019. Most of his research projects are based on experimental designs. On the other hand, he conducted many projects with the aim of knowledge construction through fundamental qualitative and culture-bound studies.

Peter Watson holds three degrees in Mathematical Statistics including a Ph.D. (Manchester). He has been providing statistical support in various ways to the research at CBSU, the Cognition and Brain Sciences Unit in Cambridge (and its predecessor, the Applied Psychology Unit) since 1994 (and prior to that fulfilling a similar role at the MRC Age and Cognitive Performance Research Centre in Manchester). He is the co-author of over 100 papers and lectures at the University of Cambridge. He is a statistical referee for several journals including *BMJ Open* and the *Journal of Affective Disorders*. He has also been a major contributor of articles to the on-line CBSU statswiki web pages which receive upwards of 100,000 visits annually. He has also been secretary, since 1996, of the Cambridge Statistics Discussion Group and chair and meetings organiser for the SPSS users group (ASSESS) since 2001 and has also been a member of the Clinical Trials Advisory Panel for Alzheimer's Research UK.

Forough Esrafilian is an Assistant Professor at the Tarbiat Modares University (TMU). Her interests include psychopathology, cultural and cross-cultural psychology, and developmental cognitive psychology.

Sara Saljoughi was born on March 1, 1995. She received her Bachelor's in Electrical Engineering from the Iran University of Science and Technology. Afterward, she discovered her intense passion for Cognitive Science; as a result, she changed her academic direction and followed her passion. Right now, she is in the final stage of her Master's in Cognitive Psychology and looks forward to continuing her studies in a Ph.D. program. Her research interests include computational neuroscience, connectivity, biomarkers, and brain networks. She is keen to discover brain networks and biomarkers for psychiatric and mental disorders by ML approaches.

Chapter 1
In Search of a Method

H. Farahani et al., *An Introduction to Artificial Psychology*,
https://doi.org/10.1007/978-3-031-31172-7_1

1.1 What Is Artificial Psychology?

Artificial psychology (AP) is a highly multidisciplinary field of study in psychology. AP tries to solve problems that occur when psychologists do research and need a robust analysis method. Conventional statistical approaches have deep-rooted limitations. These approaches are excellent on paper but often fail to model the real world. Mind researchers have been trying to overcome this by simplifying the models being studied. This stance has not received much practical attention recently.

Promoting and improving artificial intelligence helps mind researchers to find a holistic model of mental models. This development achieves this goal by using multiple perspectives and multiple data sets together with interactive and realistic models. This comprehensive, holistic, and interactive view may lead to a new research line in the near future. AP can open up a new horizon for mind researchers from clinical to theoretical psychologists to find a more realistic model. This horizon is rooted in a multidisciplinary approach updating our view along with the development of the related sciences leading to the finding of new results even from old datasets and models. AP has some assumptions. Satisfying these assumptions helps find a more precise and deeper way of modeling for artificial psychologists.

The assumptions of AP are discussed here. First, we assume that the mind is filled with uncertainty. The uncertainty is the cost we are paying for living in the real world. We are usually trying to proceed through this uncertainty by considering the most certain fact as a truth. It is important to note that uncertainty not only occurs in nature but also in almost all man-made systems. Second, we assume that the mind is continuous. In other words, we assume a continuous consciousness in which the brain acts holistically and outputs behaviors discretely (Huette et al., 2012); therefore, there is not a sharp dividing line between emotion and cognition. The brain is a grey matter that constructs mental systems not separated by solid lines. These ambiguous areas are the ones mind researchers are trying to handle by the use of

statistical models. The third assumption is that the mind is a complex system; human mentality is made up of complicated systems. Even the simplest system and phenomena are complex. This complexity can be captured and interpreted by a dynamic model. The fourth assumption is that there is always a proxy between mind and data. It is not possible to study mental activities directly. Brain data needs to connect to some psychological constructs and behaviors. Therefore, we need to use multiple sources of data in a single model at the same time. Conventional statistical techniques use rigorous mathematical models. These models require comprehensive and complete data for analysis and prediction. In the real world, we are facing big, imperfectly measured data as well as nonlinear relationships in complex systems. The fifth assumption is that brain data is highly dimensional data. This implies that the dataset has many features even in small sample sizes. This problem commonly occurs in psychological research, especially in clinical, cognitive psychology, and neuroscience, where we need to deal with $P > n$.

In summary, psychologists need new analysis models to help them model complex mental systems. Artificial psychology uses intelligent models that satisfy these assumptions (Fig. 1.1).

One technique used in applied computing is to emulate the strategies involved in the intelligent systems or models for problem-solving. Intelligent models are related to the human way of thinking and interpretation. These models use fuzzy logic, artificial intelligence, and genetic algorithms both individually or together.

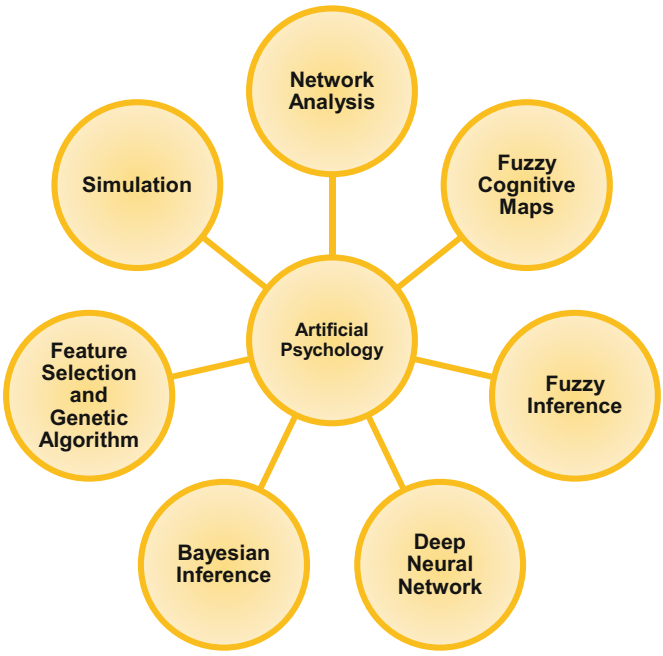


Fig. 1.1 Artificial psychology as a multidisciplinary field

1.2 In Search of a Method

Science aims to clarify concepts systematically, and its core is replicability. The ambient world is full of concepts, some of which are determinable or almost determinable, such as country borders and per capita income. Others, however, that are more prominent and studied in the social and behavioral sciences are different, such as depression, suffering, grief, love, and selective attention.

There is an infinite set of these concepts and they are very vague, scattered and elusive and assume highly diverse intertwined forms. The models designed by researchers to acquire an approximate understanding of these concepts represent the efforts of science to clarify them. Models are an approximate simplified understanding of reality and do not deal directly with reality. Models are always an approximation of reality outside of them. These models test theories and hypotheses about different concepts.

If psychological reality is vague and elusive, how should it be examined then?

Such a reality cannot be understood well by a single method. Therefore, research methods must develop like any other science, and there is no harm in employing a multitude of methods and, sometimes, methods of other disciplines for more accurate proximity to concepts.

Perhaps the principles and assumptions of many of the existing methods and future ones will probably require revision, even the methods introduced in this book. The dominance and hegemony of a particular method is remarkably alarming and dangerous.

It should be borne in mind that methods are not the only means of achieving reality, but they are the containers of reality, and the research findings take shape from them. Furthermore, these findings from the method occasionally become so extreme that they annoy some researchers.

It is noteworthy that the present book does not attempt to discard, ignore, or devalue past or existing methods but to overcome the fear of going beyond them. The research psychologists' shared fear might result in innovation. The aim is to find a way in which elusive concepts could be understood more clearly by researchers.

We know that wrong answers are more harmful than random answers because wrong, non-random answers mislead science systematically and significantly.

From Sigmund Freud's birth in 1856 until the writing of this book (2022), the earth has rotated 60,590 times, and research methods need to rotate as well. We know that psychology did not begin from Freud's birth, but Freud is cited since he has been called one of the scientific revolutionaries.

1.3 From *p*-value to *p*-war

Statistical significance plays a significant role in scientific research by linking data to hypothesis testing from the late mid-twentieth century (Gigerenzer et al., 1990). Currently, the most commonly used statistical measures in scientific studies, despite much criticism of its use, is the *p*-value (Lyu et al., 2020). In spite of the widespread use of the *p*-value in psychological research, various studies show that most researchers and students misinterpret *p*-values. This misinterpretation is rooted in *p*-value misuse, including the statistical significance hypothesis (Ziliak & McCloskey, 2008) and p-hacking. These are the main reasons for the confidence crisis and

reproducibility crisis in psychology research. The effect size and confidence intervals (CIS) can be regarded as alternatives to the *p*-value. Although the CIS can be used to improve statistical interpretation and inference, considering it an indicator of the effect size variations, its concept has also been misunderstood by researchers and its users (Lyu et al., 2020; Harrison et al., 2020; Greenland et al., 2016; Lyu et al., 2018; Morey et al., 2016; Cumming, 2013).

The study carried out by Lyu et al. (2020) on 1479 researchers and students in various fields in China revealed difficulty interpreting the *p*-value and CIS correctly, regardless of their academic degree and career stages. That is, 89% of them made at least one error in the *p*-value interpretation, and 93% made at least one error in the CIS interpretation. The level of misinterpretation in the significant and nonsignificant *p*-values and whether the CIS included zero or not was increased. Moreover, it is noteworthy that respondents were generally confident in their (incorrect) judgments.

These results indicate that researchers have misunderstood these crucial indicators of inferential statistics. This misunderstanding causes researchers to misinterpret, using classical statistics-based methods (assuming we are pleased with the *p*-value!), and these interpretations flow from different streams into the sea of psychological research findings (Harrison et al., 2020).

These interpretations include the following:

1.3.1 p-*value as Evidence to Confirm or Unconfirm a Null Hypothesis*

This issue is what is referred to as the illusion of certainty in Gigerenzer's research (2004, 2018), as it may provoke a crisis of confidence in the psychological research findings by encouraging researchers to reach a *p*-value of ≤ 0.05 as evidence of the existence of an "effect." One of the primary sources of this crisis is publication bias. This bias results from the fact that scientific journals welcome statistically significant results (p-value ≤ 0.05).

Chang et al. (2019) state that a *p*-value is a probability of obtaining an effect at least as extreme as the effect in the sample data, assuming the truth of the null hypothesis. Considering the *p*-value in classical statistics, despite its low statistical power in both single studies and meta-analyses, may result in distrust in the actual results of psychological research. Given the statistical power in the published studies, the frequency of statistical significance in those studies is suspiciously high (Francis, 2014).

Schmidt and Oh (2016) asserted that 90% of the research reports were significant, while the average total power was 0.4. What they found is strong evidence that these studies are questionable. Test power did not increase in questionable research practices until 1962, when Cohen emphasized the "low power" issue. What was the reason for this emphasis?

We know that in classical statistics, the *p*-value is a function of the effect size and the sample size, and the sample size seriously affects the power increase:

$$(p-\text{value}) = f(\text{Effect size}, \text{sample size}) \tag{1.1}$$

Low power leads to nonsignificant results. Nonsignificant results are as important as significant ones, but their nonsignificance made it difficult for researchers to publish their papers. This publication bias is called the file drawer problem, with studies less likely to reject the null hypothesis ending up unpublished in a file drawer. Therefore, researchers have earnestly strived to increase the power of their research. Maxwell (2004) has elaborated on this issue extensively. However, it has not always been possible to obtain a sufficient sample size, so maybe this is why researchers conducted questionable research. In other words, they conducted questionable research to obtain significant results to avoid the abundance of nonsignificant results due to low statistical power (Harrison et al., 2020).

1.3.2 Reverse Interpretation of the p-*value*

Another consequence of misinterpretation of the *p*-value, as Lyu (2020) states, is replication illusion. Many researchers avoid Bayesian thinking because of classical *p*-value-based statistics, the thinking that is the basis of classical inference. The reverse interpretation of the *p*-value is to consider 1 – *p*-value as the probability of successful replication of the result.

Despite these problems, the potential consequences of the lack of statistical thinking and ritual use of *p*-values have rarely been mentioned in the psychological research results, except in recent years (Lyu et al., 2020). The study of Farahani et al. (2021a, b, c) on a sample of 100 postgraduate and Ph.D. psychology students in Iran indicated that 95.7% of them make mistakes about the illusion of certainty and the replication illusion in the interpretation of the *p*-value.

The *p*-value is not well understood, and most researchers speak about it with a wrong mindset and perception. A *p*-value demonstrates the likelihood that the researcher's data will occur under the null hypothesis. This is obtained by calculating the likelihood of test statistics gained from the researcher's data (Indrayan, 2019).

It should be noted that the *p*-value is a ratio and a percentage. The *p*-value is the probability of a test statistic at least as big as the test statistic obtained from the data, assuming that the null hypothesis is correct.

Harison et al. (2020) summarized the shortcomings of using NHST (*p*-values) as follows:

1. Use of *p*-values without regarding the effect sizes and confidence intervals is not informative.
2. The potential for the use of "p hacking" by manipulating data and analyses deliberately to reduce *p*-values.
3. Simplistic dichotomous interpretations of *p*-values as either significant or nonsignificant.
4. Incorrect interpretation of $p > 0.05$ as no effect.
5. Misinterpreting statistical significance and taking it as clinical or practical significance.
6. Committing multiplicity by performing multiple statistical tests without adjusting the criterion *p*-value.

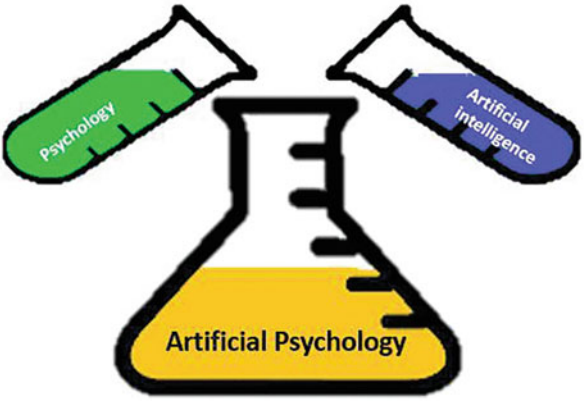
One way to improve *p*-value interpretation is to use clinical interpretation, practical interpretation, or practical significance. Apart from statistical significance, the effect size should be used for practical interpretation. Another point is that reporting inconclusive findings and null findings in articles is not harmful but valuable and strengthens the scope of scientific theories, but it should be borne in mind that what was said at the beginning of this chapter about the world not being black and white encourages researchers to choose another way to have accurate yet close-to-reality findings.

To design a different research model, a new conceptual framework is required with different measures, which will be discussed in detail in the second chapter.

Chapter 2
Artificial Psychology

H. Farahani et al., *An Introduction to Artificial Psychology*,
https://doi.org/10.1007/978-3-031-31172-7_2

2.1 Artificial Psychology



Artificial psychology was first proposed by Dan Curtis in 1963 as a theoretical discipline. Artificial psychology is a combination of psychology and artificial intelligence. As a comprehensive definition, psychology can be called the science of studying an individual's mental processes and behaviors.

Artificial intelligence also has a wide variety of definitions, and it may not be possible to provide a comprehensive definition that encompasses all of its dimensions; however, it is a science that deals with the design of intelligent machines and systems; systems that can perform tasks requiring human intelligence (Crowder & Friess, 2010).

Here, we return to artificial psychology. The developments in psychology and artificial intelligence as of 2022 have addressed the needs of researchers. Here, artificial psychology has a theoretical framework or is simply a theory on which the artificial psychology presented in this book relies to look at the world of psychology. In other words, artificial psychology uses artificial intelligence to design, train, test, and ultimately deploy methodological models in the psychological context. This representation of artificial psychology is shown in Fig. 2.1.

This theoretical framework has features borrowed from psychology, artificial intelligence, and the psychological contexts in question. As the basis of artificial psychology, this theory is interpretable and explainable in artificial intelligence-based psychology.

Artificial psychology in this book relies on the above theory for prediction and classification to provide robust, interpretable, and explainable models. Here, artificial psychology refers to the scientific application of this theory.

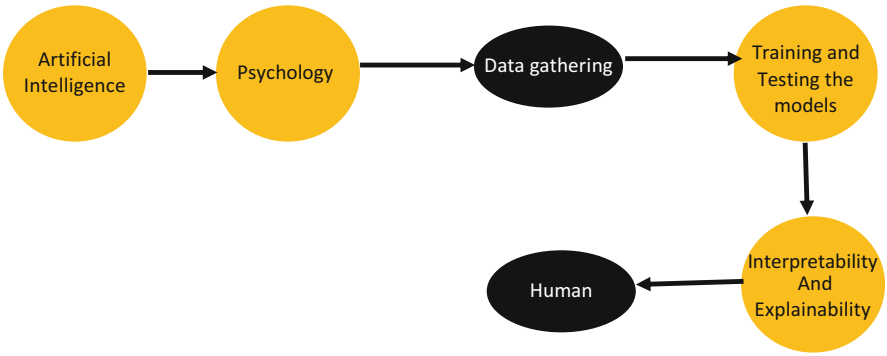
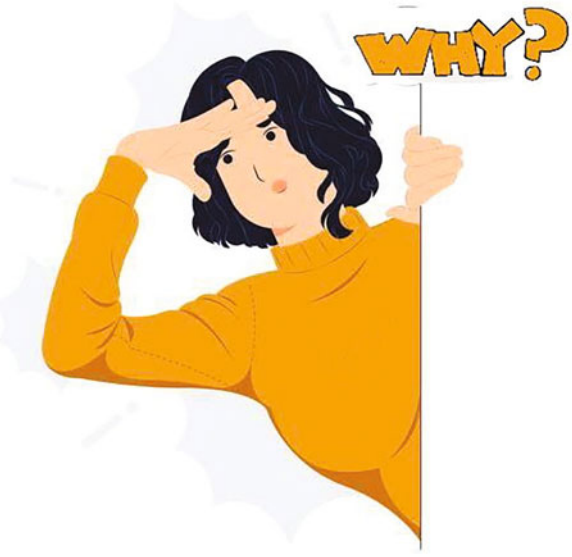


Fig. 2.1 Representation of artificial psychology

2.2 Why Artificial Psychology?

The question we have to address here is why artificial psychology is significant, as its application is becoming increasingly prevalent.

From the data analysis perspective, psychology mainly aimed to create models to infer human behavior for about a century. Researchers in behavioral sciences used the null hypothesis test to conclude and find the causality and underlying mechanisms of human behaviors.



For nearly two decades, machine learning has received increasing attention in scholarly studies as a branch of artificial intelligence (Lipkova et al., 2022). Machine learning is part of artificial intelligence and perhaps the most essential and practical part of artificial intelligence in artificial psychology. There are numerous definitions of machine learning, but the definition given by Samuel in 1959 is taken as the working definition (Buchanan, 2005).

Machine learning is a branch of study that enables computers to learn without being explicitly programmed. Perhaps the most prominent part of this definition is "without being explicitly programmed." In the case of obvious programming, the software receives the input data and generates the output according to the specific rules defined by the programmer (Fig. 2.2).

We know that these rules guarantee the explainability and transparency of the system as it is possible to understand exactly what those rules (algorithms) follow. Thus, knowledge of the algorithm answers "why," "what," and "how" in the model, and thus explainability is achieved. We will return to the topic of explainability later. The age of algorithms has given way to the age of data. Figure 2.3 indicates the short history of machine learning. Machine learning seeks to design models to predict unseen data with high accuracy. Fans of inferential statistics should not despair. Inferential statistics can still be used to test null hypotheses. These methods are used

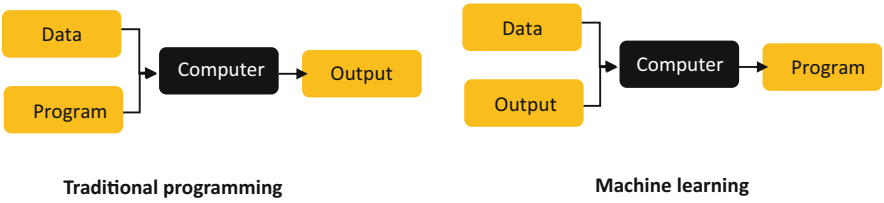


Fig. 2.2 Comparison of conventional programming and machine learning

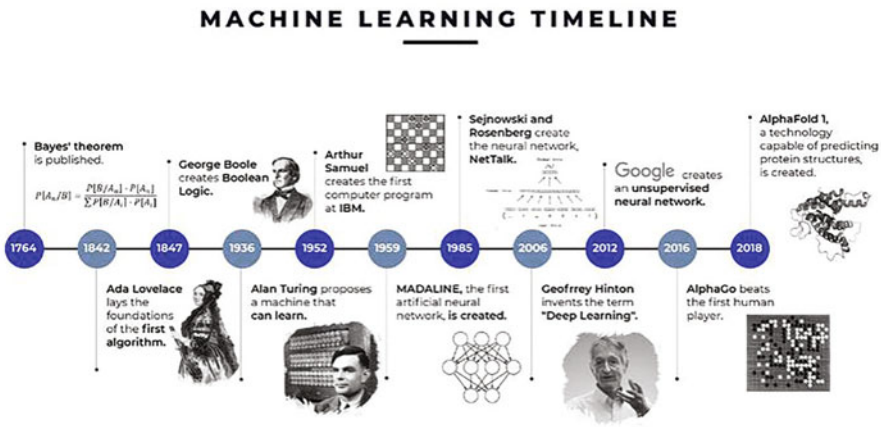
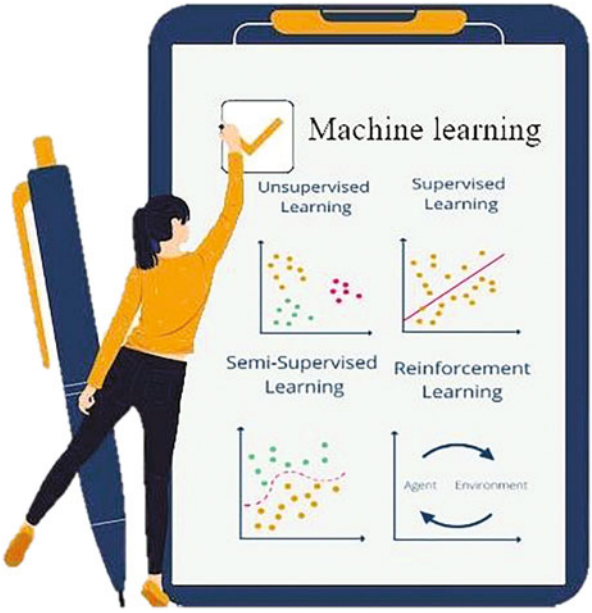


Fig. 2.3 Short history of machine learning. (From https://www.algotive.ai/blog/machine-learning-what-is-ml-and-how-does-it-work)



in inferential statistics to infer human behaviors. Behavioral inference means that the cause of behaviors and relationships between variables are explained, and the best possible insight into human behavior is obtained. These methods have limiting assumptions. The most significant limitation is that the generalizability of models is questionable.

In behavioral sciences and psychology, statistical inference is mainly based on statistical tests on aggregated data. The underlying assumption of these statistical tests is that what is estimated from the groups can be generalized to individuals. The generalization of group-based findings to single-case levels is interpretable as long as the studied process is ergodic. Ergodicity indicates that the studied effects are homogeneous among individuals and constant over time (Oliveira & Werlan, 2007). However, psychological constructs are organized over time in individuals, and in the case of individual exceptions, the group-based generalizability is not ergodic.

On the other hand, inferential models are utilized for samples and variables that are small to medium in size. The lack of multicollinearity is another assumption. Yarkoni and Westfall (2017) believe that one reason for choosing classical inferential statistical methods in psychology is the poor understanding of the tools that make successful predictions and the slow deployment rate of these tools after creation. Data and AI (Artificial Intelligence) scientists are able to target the limitations of classical inferential statistics, such as a large number of assumptions, generalizability, complexity, low number of input variables, and poor predictive power. Developments in recent years have targeted mainly the limitations of classical statistics. The models used for prediction create systems that incorporate advanced statistical and probabilistic methods. These methods learn from data to detect and extract

hidden patterns in order to predict unseen or out-of-sample data. These AI-based predictive models are called machine learning (ML). Machine learning has significant advantages over the models used for statistical inference. Machine learning is a relief from the cumbersome assumptions of inferential statistics. These assumptions work well on paper, but in real life, they are not generally met regarding the data. Machine learning is based on minimal a priori assumptions. In ML, there is no need to select variables because, in classical inferential statistics, multicollinearity is an erosive problem.

Machine learning can be utilized for high-dimensional data, where the number of rows is less than the number of columns in the dataset file. Highly correlated variables and small sample sizes are common problems in cognitive and clinical psychology and neuroscience. For example, consider predicting the disorder based on EEG-based brain features in people with clinical depression. Large samples and extraction of brain features are time-consuming, costly processes.

Machine learning provides good predictions even if the input variables are beyond individuals. It means the models based on machine learning can perform well with new data (unseen data) (Xin et al., 2018).

One of the problems of research and theoretical models in psychological sciences is their complexity. This complexity raises serious issues in modeling. Complex models involve the selection and use of many variables. Psychological phenomena and mental disorders are mainly complex because of their including subsystems (models) such as emotion, cognition, memory, behavior, and biology (Borsboom et al., 2021a, b). ML enables the building and testing of such models.

ML gives predictive models the ability to model individuals instead of groups. The advanced part of ML, called deep learning, allows one to study complex models and systems. The following is a brief description of machine learning types.

ML makes it possible to make predictions in the case of multiple variables, small samples, and complex or even unknown nonlinear relationships between variables to find hidden patterns. These unique features raise a new problem at heart: such complexities in the ML pose a severe challenge to interpretability, requiring expertise in interpreting these models.

Statistical and quantitative accuracy precludes qualitative and in-depth interpretation, so even if the accuracy of prediction models performed in the ML is higher than classical inferential statistics, the results might still be questionable. For instance, if we intend to identify depressed people in image-based deep neural networks, and in those images, the depressed people are wearing winter clothes and nondepressed people are wearing summer clothes, then the accuracy of prediction in their category is 100. Nevertheless, winter clothes are not related to depression. This shows that high accuracy is not necessarily a good result. The adage of "well-shaped apples are not necessarily tastier" applies here! In recent years, to overcome the problem of uninterpretability in ML, a new field has been developed in artificial intelligence and ML, which is called explainable artificial intelligence or ML (XAI or XML). The ultimate goal of XAI is to provide models to increase interpretability, transparency, and fairness.

The interpretable model is trained by minimizing cost functions such as weighted root mean square (RMSE). It should be noted that to achieve an interpretable model in ML, the cost function should be minimized to the extent that the complexity and fidelity indices allow. Fidelity can be considered the extent to which an interpretable model can approximate the original model. Let us take a closer look at these features.

The purpose of explainable artificial intelligence (XAI) and consequently explainable ML (XML), on which, as mentioned, the artificial psychology discussed in this book is based, has different features with the aim of attempting to address the inherent nature of the ML, which is a black box, to gain a better understanding of the decision-making process.

Adadi and Berrada (2018) state that explainable systems are also interpretable if humans understand their operations. This definition indicates that interpretability is closely related to the concept of explainability, although Gilpin et al. (2018) consider interpretability and fidelity as essential components for the explainability of models.

They believe a good explanation should be understandable to humans (interpretability) and accurately describe the model behavior in the entire feature space. Interpretability may contain features such as transparency and parsimony. Transparency shows that the explanation is not ambiguous, while parsimony means that the explanation is presented in a simple, concise pattern (Rudin, 2019; Zhou & Chen, 2018). Fidelity has features of completeness and soundness (Zhou et al., 2021). Completeness indicates that the given explanation describes the overall dynamics of the ML model, while soundness indicates how correct and reliable the explanation is (Fig. 2.4). This model is based on Markus et al. (2021).

Velez and Kim (2017) defined the machine learning environment interpretability as "the ability to explain or present in human terms." This definition may seem slightly vague. Let us clarify it in the form of an example.

Supposing that as the anxiety increases, the individual performance in the maths test first improves, then is stable, and finally decreases. Refer to Fig. 2.5.

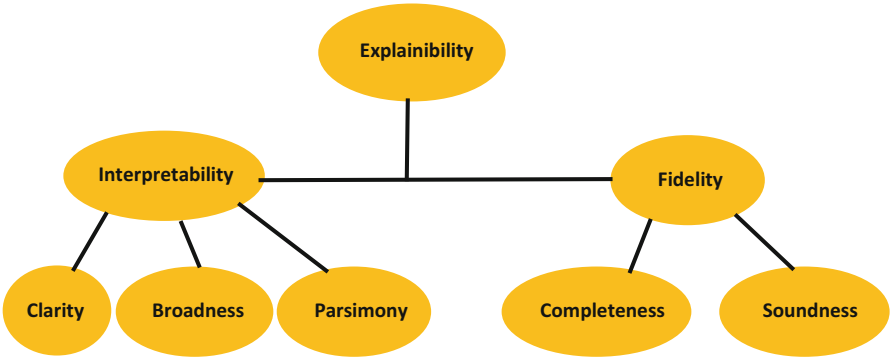


Fig. 2.4 Definition of ML explainability and related features

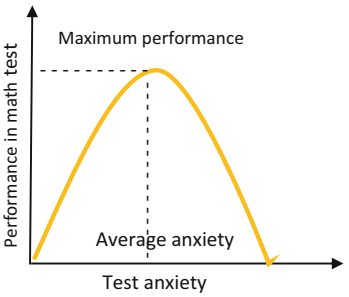


Fig. 2.5 Nonlinear relationship between test anxiety and math test performance

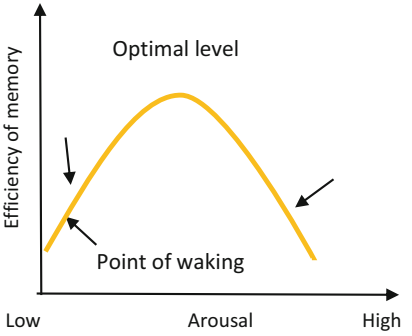


Fig. 2.6 Relationship between arousal and memory performance

Figure 2.5 shows the nonlinear relationship between test anxiety and maths test performance. This relationship is an inverted-U function.

In this example, interpretability can be understood as how the machine learning system predicts test performance by increasing stress scores. The explainability of this model refers to the fact that it is based on a theory such as Yerkes-Dadson's law. This general understanding is associated with knowledge discovery (Fig. 2.6). Simply put, explainability can be considered the possibility of understanding the mechanics of an ML model.

Figure 2.6 shows that Yerkes-Dadson's law demonstrates an inverted relationship between arousal and memory performance.

Gianfagna and Di Cecco (2021a, b) reveal that regarding the explainability of the machine learning model in practice, we are looking for how much the output will change if different values are placed in a feature that was not in the data.

Explainability requires checking whether the machine learning model is trained on a biased dataset. This feature is fairness, which is related to model validation. In addition, research models must be debugged to generate and discover knowledge. Model debugging ensures reliability and robustness. This debugging refers to displaying what is happening behind the scenes by examining the model. A small change in input data should not lead to a considerable change in output. This examination contributes to the stability and robustness of the model.

It should be borne in mind that the most complex function of the ML models is to acquire and increase a clear understanding of psychological processes, events, and systems studied in the psychological sciences. As previously discussed, the discovery of psychological knowledge is the primary purpose of explainable artificial psychology and cannot be interpreted by merely predicting and accurately classifying the application of these models in the psychological sciences. Extraction of scientific knowledge is not entirely done by prediction. Artificial psychology is required to create results with explainability and interpretability. Furthermore, another point that emerged in the data age is that the solid line between correlation and causality is fading. Correlation deals with the relationship of two or more variables that change together, i.e., they have a common variance, but this correlation does not indicate causality. In causality, one variable causes another (cause and effect).

In a town, as the number of storks increased, the number of infants also increased. This random association should not be considered a correlation because an illusory correlation causes it, and by removing the variable, such a relation would be removed as well. Moreover, what is considered a graver mistake is to make a causal inference from this phenomenon, meaning that the storks bring the babies to the town!

As noticed, this is a ridiculous interpretation. Therefore, explainability and interpretability can play a significant role by creating valid, debugged models to help discover scientific knowledge. Consequently, it tests the application of artificial intelligence in models with deep, mixed, and multilayered psychology, which is the primary purpose of this book and enables the explainability of the models (Table 2.1).

Interpretability, as the other part of the XAI theoretical framework, can be performed in three stages of the explainable ML modeling process. This feature could be observed in the form of three types or stages (Aslam et al., 2022) (Fig. 2.7).

Table 2.1 Summary of the types of explanations

The most important question	Type
(1) How does a model work?	(1) Global explanations
(2) Why does a model reach a prediction for a particular input? (Does this prediction refer to a specific feature in the data or a specific algorithm?)	(2) Local explanations
(3) Why does a model make a certain prediction instead of another for a particular data? (Why X and not Y?)	(3) Contrastive explanations
(4) How does an output change by tweaking the data and parameters of a model?	(4) What-if explanations
(5) How can the desired result be achieved by making the least change in the model? (regardless of understanding its internal structure)	(5) Counterfactual explanations

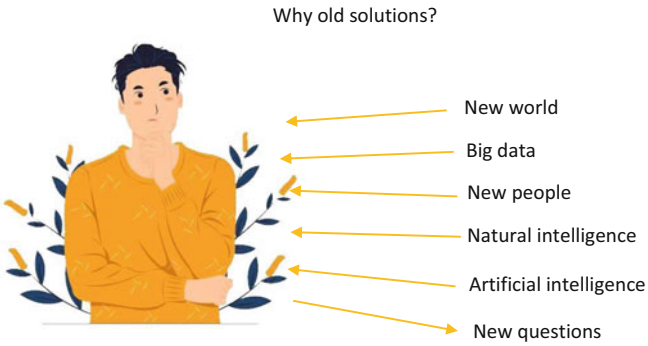
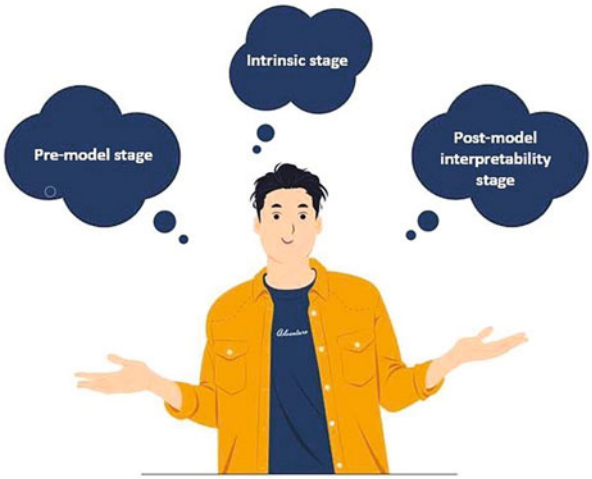


Fig. 2.7 Modern problems require modern solutions



1. Pre-model Stage
This type of interpretability has nothing to do with the model and is achieved before the implementation of the model and is solely based on data function, generally through data visualization and exploratory data mining such as exploratory factor analysis (EFA). This data interpretability is implemented prior to the formulation and implementation of the explainable ML model and plays a crucial role in determining the model. This type of interpretability is related to feature engineering. Psychologists are familiar with some of these methods in the form of descriptive statistics.

2. Intrinsic Stage
This interpretability refers to self-explanatory models that reinforce natural interpretability with their internal structure. Intrinsic models include basic models such as decision trees, generalized linear models, logistic models, and clustering models. It should be noted that natural interpretability is associated with cost, depending on the model's accuracy.

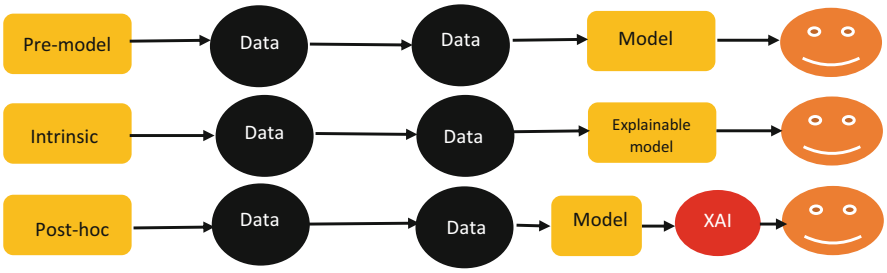


Fig. 2.8 Interpretability and models

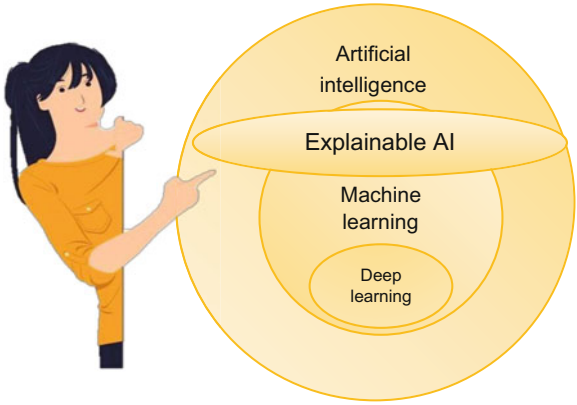


Fig. 2.9 Relationship among artificial intelligence, machine learning, and explainable artificial intelligence

3. Post-model Interpretability Stage
This kind of interpretability can be used for all black box models without the need to understand the internal structure. This interpretability can even be used for intrinsic models.

In short, Fig. 2.8 depicts the relationship among artificial intelligence, machine learning, and explainable artificial intelligence (Fig. 2.9).

Let us return to the ML classification, which is at the core of artificial psychology. Machine learning systems are generally divided into three categories, which can be briefly defined as follows. The basis of this division is the training type.

1. Supervised Learning
In this type of machine learning, the system uses data for active learning to map inputs to outputs. The reason this type of learning is called supervised is that there is a solution. The solution is called the target variable, and in psychological research, it is called the criterion (dependent) variable. This variable has a specific label in the data (for example, one hundred boys with ADHD and one hundred non-ADHD boys) and there are some features as predictors based on which predictions are made; for instance, the child's age, birth rate, and daily caffeine intake. In this example,

since the target variable has a separate label and is already known, supervised algorithms such as linear regression, random trees, boosted trees, and neural networks (ANNS) are used in data training. This type of learning is the most widely used type of ML in the psychological sciences (Vélez, 2021a, b).

Farahani et al. (2011) used a supervised learning algorithm in their study.

2. Unsupervised Learning

Training data is not labeled in this type of learning. In other words, the solution does not already exist, and the system learns to find the hidden pattern in the data (such as the PCA, K-means, and auto-encoder clustering algorithm). The target variable (criterion) is not determined in this algorithm. This part of machine learning is more about general artificial intelligence. An instance of this can be found in the brilliant study by Grazioli et al. (2021) to personalize treatment in children with ADHD using the clustering algorithm.

3. Reinforced Learning

This algorithm differs from the previous two algorithms and their intersection (Fig. 2.10). In this algorithm, there is no machine training on the available data. There is an agent here performing actions in an environment and receiving a reward for each action, and its goal is to find a policy and strategy to maximize the reward.

Deep learning algorithms should be considered here, which do not fit into these three divisions. Deep learning is a type of machine learning that runs with several hidden layers and yields the most successful findings and operations.

As described, there are three types of learning algorithms in the ML and deep learning systems, and there is no single solution or set of rules to tell which category needs more explainability than the rest. Explainability is associated with the interpretability of the algorithms in that category.

According to Fig. 2.11, there is a transversal and emerging necessity across the various domains of AI. In other words, although data-driven prediction and modeling are essential as a new field of artificial intelligence, it should be remembered that the principal basis of predictions in machine learning is correlation (Zhou et al., 2018). Furthermore, knowledge is generated when these predictions, classifications, and clusters are interpreted.

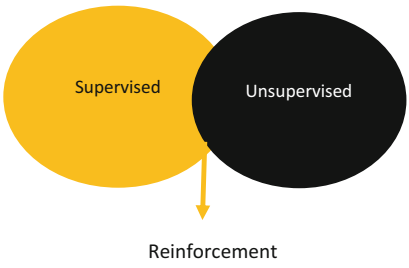


Fig. 2.10 Intersection of supervised and unsupervised algorithms

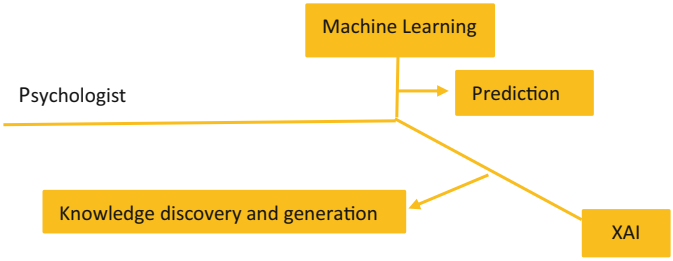


Fig. 2.11 Comparison of the purpose of machine learning and explainable artificial intelligence (XAI)

In summary, artificial psychology aims to apply explainable and interpretable artificial intelligence or, more precisely, explainable and interpretable machine learning to discover knowledge. The purpose of this new field of research is to study theoretical models using models based on explainable and interpretable artificial intelligence to be able to understand internal operations from a human point of view. This human-readable model combines accuracy with transparency and fidelity, which improves the models' weaknesses based on ML or AI. As a result, artificial psychology will be most helpful whenever models' ambiguity is reduced through using XAI, rule-based systems, and Bayesian models.

It is important to note that to reduce ambiguity using rule-based systems as well as the use of language variables that are most utilized in measuring structures, fuzzy inference systems (FIS) can be applied.

Furthermore, Bayesian Network Analysis (BNA) is an explainable and interpretable model based on prior knowledge and information. This prior knowledge helps to increase the transparency and fidelity and, thus, the explainability and

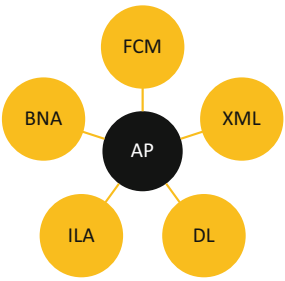


Fig. 2.12 The XML structure

interpretability of the model. This is one of the goals of artificial psychology. Fig. 2.12 illustrates the artificial psychology subfields.

2.3 Artificial Psychology in Practice

In addition to the theoretical knowledge, to become more familiar with the subfields of artificial psychology in this book, it is necessary to have the essential codes to run each algorithm of XML algorithms, fuzzy inference system, fuzzy cognitive map, and Bayesian network analysis using the R software, Python, and occasionally MATLAB. We know this book is not a software guide, but its purpose is to introduce the new field of artificial psychology and its sub-fields with examples.

R Software

For more than three decades, R (www.r-project.org) and Python (www.python.org) have facilitated the use of ML algorithms for everyone, as they are free and open-source. Recently, Julia (www.julialang.org) has emerged as a powerful and effective alternative. Since 1995, R has been a language and environment with various

statistical and graphical methods for classical statistics, machine learning, clustering, and predictive models and is highly expandable.

There are several freely available ML, FIS, FCM, and BNA packages in R. See the Comprehensive R Archive Network (CRAN) for more information.

Python Software

Guida Van Rossum developed Python software in 1991. It is widely used software with a high-level, object-oriented programming language. There are various Python packages for ML, FIS, FCM, and BNA.

The practice is further explained in the following. We do not look for coconuts on Coconut Island! The aim is to give a novel perspective, a novel function, and novel findings in psychological research.

Main Approaches for Modeling a Phenomenon

First, artificial psychology researchers must design their model to predict or classify the target variable. In this step, the researcher attempts to implement the desired model by examining the theory and research background. What are the predictors? Why are they chosen? What are the measuring tools? What is work innovation? This approach is called the knowledge-based or model-based approach in this book. In fact, in this approach, the initial, hypothetical model is designed through multiple sources, including a literature review, experience in the field, and interviewing the target and the expert group, using qualitative analysis (thematic analysis and content analysis) (Fig. 2.13).

The second approach, also called the data-driven or science-based approach, uses data as the model's source of inspiration. The present age has been called the age of data. Although Big Data has always existed, what is seen now is recorded Big Data,

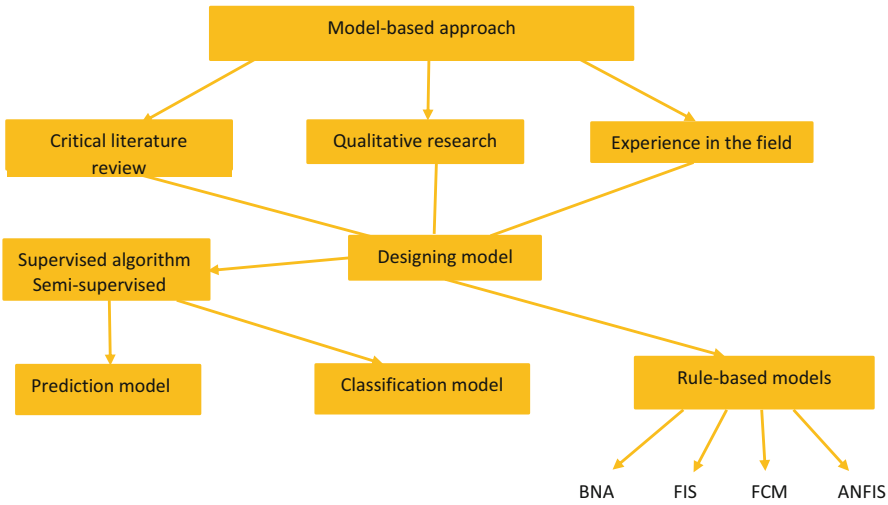


Fig. 2.13 The implementation process of a knowledge-based approach

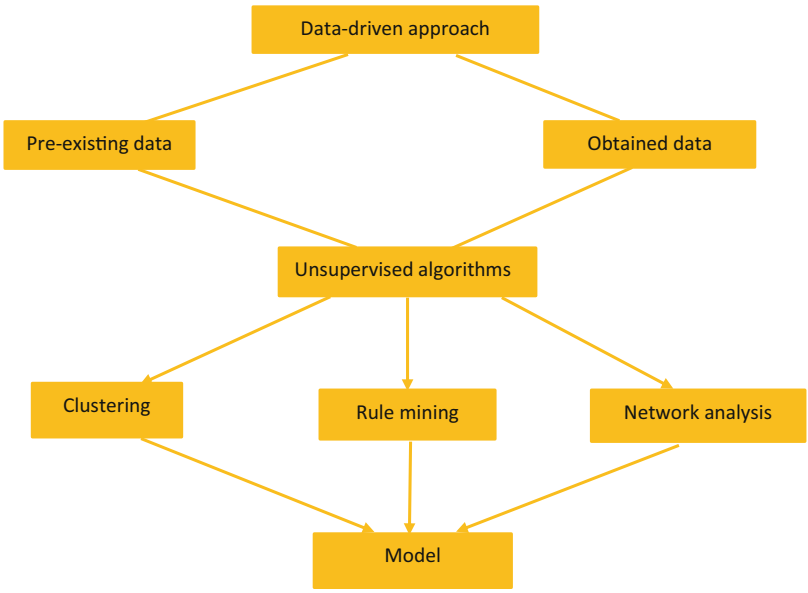


Fig. 2.14 Data-driven model process

and there are various ways to record Big Data. In psychology, data can be obtained from multiple sources, from surveys and clinical batteries that measure essential dimensions of human behavior using EEG, reaction times, and genetic and omics data (Velez, 2021a, b). Data recording through the Internet, national networks, and smartphones have led a data explosion that may cause confusion. In this situation, the artificial psychologist attempts to reach an explainable and interpretable model based on data-driven models. We know that an artificial psychologist can collect data for a specific purpose and use the collected data to design the final model, an interpretable and explainable model. The following figure shows the data-driven model process in detail (Fig. 2.14).

It is necessary to know that the explainability and interpretability of model-based and data-based approaches are particularly important. It should be noted that explainability and interpretability are slightly more straightforward in supervised models because they occur in a label-bound environment. Moreover, the explainability and interpretability of white-box algorithms such as linear regression, decision tree, and logistic regression are not essential as their algorithm is quite clear.

In the following, explainability and interpretability in the knowledge-based and data-driven models in artificial psychology are explained in detail.

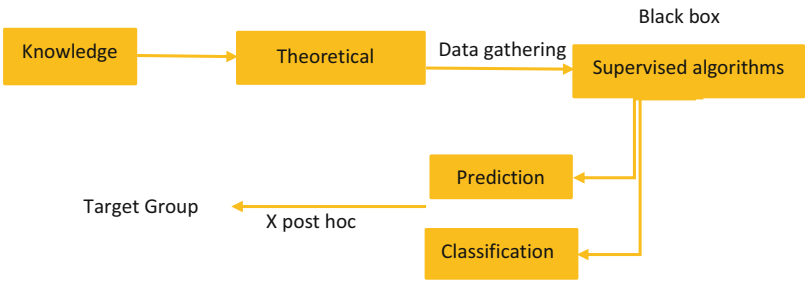


Fig. 2.15 An illustration of explainability and interpretability route in models resulting from a knowledge-based approach

2.4 Interpretability and Explainability in a Knowledge-Based Approach

As shown in Fig. 2.15, this explainable artificial psychology approach, which is the book's promised land, is of great importance.

In this approach, models are designed based on prior knowledge, and in quantitative studies, modern methods based on explainable machine learning, which is part of explainable artificial intelligence, are used. It should be noted that in this approach, rule-based models, including Bayesian Network Analysis, Fuzzy Inference System (FIS), Fuzzy Cognitive Map (FCM), and Adaptive Neuro-Fuzzy Inference System (ANFIS), are all based on rules obtained through theoretical foundations, literature review, and experts' opinions or raw data. Prior knowledge is their basis. Therefore, they are considered white boxes and are explainable and interpretable. However, we wish to call fuzzy logic models and fuzzy sets Gray Box because fuzzy logic sees gray land as its promised land! Hence, the resulting model is explainable (internal explainability). Supervised and black-box algorithms need to be explainable and interpretable. Model explainability is a significant criterion for performing computational methods.

Various methods have been proposed to explain the models resulting from black-box machine learning. According to Jaganathan et al. (2022), these methods are of two types: model-explainability-based and instance-explainability-based approaches. Model-based and instance-based explainability in relevant texts are known, respectively, as global and local explanations. On the other hand, these explanatory approaches can be model-specific or model-independent (agnostic). Unlike model-specific explanation, the agnostic explanation can be used for any ML model, and typically this explanatory model is used as post model or post hoc.

The Shapley Additive Explanations (SHAP) algorithm can be used for a general explanation in any ML model (Fig. 2.16).

Now it is time to deal with explainable and interpretable methods in models based on the second approach of artificial psychology, i.e., the data-driven. Unsupervised black-box models are challenging in terms of interpretation as most explainable methods in the XAI and XML require labeled data (Crabbe & Schaar, 2022). In other

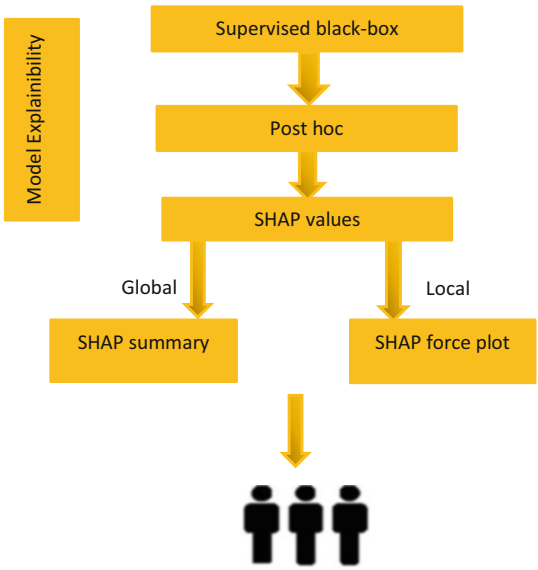


Fig. 2.16 SHAP algorithm

words, given the supervised algorithms for interpretability and explainability of algorithms in the second approach, several ways can be suggested based on the summary of one of the related sources. The model can be extracted from data based on unsupervised algorithms with a data-driven approach. Then, white-box algorithms (e.g., logistic regression or decision tree) or other supervised black-box algorithms or experts' opinions can be used to explain it on the basis of the target variable. The explainability of the result can then be given by, for instance, running random forest and/or support vector machine (SVM) and SHAP or feature selection. Figure 2.17 represents this process.

To sum up, in artificial psychology, test models are extracted and then explained and interpreted using two categories of supervised and unsupervised algorithms (semi-supervised and reinforced algorithms are not discussed in this book). Predictability and explainability with accuracy, precision, and human readability are the prime goals of this research field in psychology.

2.5 Achilles' Heel in Psychology

In recent years, several scholars have claimed that the theoretical foundation of psychology is shaky (Fielder et al., 2017; Eronen & Bringmann, 2021). They believe that psychological theories in general are of low quality. Therefore, instead of focusing on improving statistical methods, researchers should focus on developing better theories. In other words, the theory crisis is the Achilles' heel of psychology, and the theory crisis is more significant than the replication crisis. He argued that

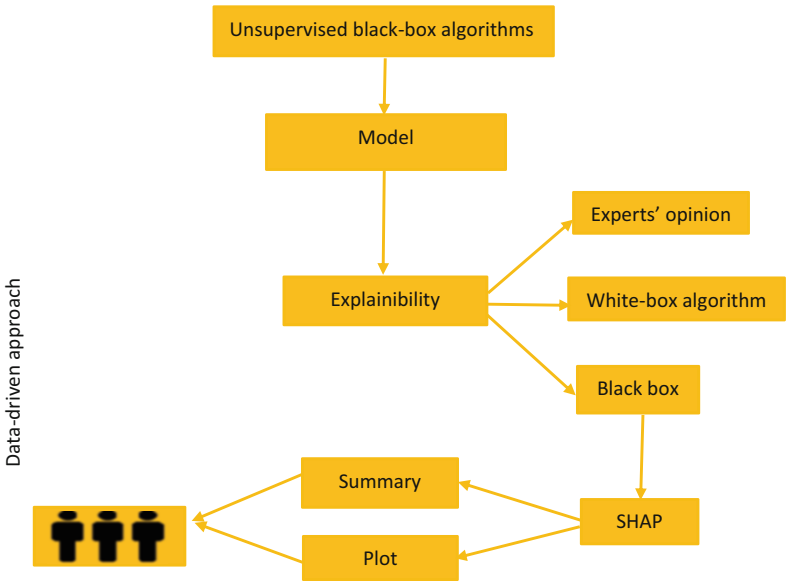


Fig. 2.17 Representation of the SHAP process

although psychologists are interested in developing new theories, these theories are not presented in a consistent manner. They are not tested, proven, or disproven. They simply "exist" until they are forgotten or set aside.

Some believe that this type of theorizing and degradation of theories has led to the crisis of reproducibility.

Theories are undoubtedly the most powerful scientific tools available to psychological researchers, whether they are grand theories like Freud's (revolutionary views about humans) or small theory-rich bubbles. It can be argued that theories are typically developed by a small group of theorists. This problem is referred to by Mischel (2008) as the "toothbrush problem," i.e., psychologists view theories as toothbrushes: no one (typically) uses someone else's toothbrush. It should be mentioned that this does not mean that psychology lacks necessary theories. It means that psychological sciences lack theorizing and methodologies.

Researchers have mostly focused on developing hypotheses based on deduction and then testing these hypothesis is collecting data and using inductive (statistical) methods. In other words, the scientific method has been based on the mixture of deductive and inductive approaches.

Deduction refers to what must be there, while induction refers to what is operative. In other words, theories are the result of the process of hypothesis-deductive, which relies on the assumption that science develops via constant scientific testing. This is the underlying assumption of most of the books in psychology. Using deduction, induction, and statistics is quite valuable and no one would undermine

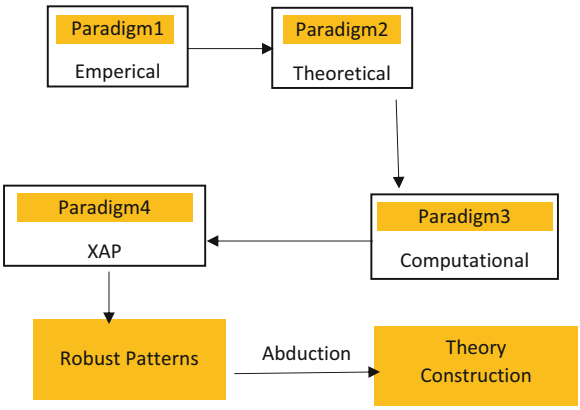


Fig. 2.18 Paradigm view on the XAP

its importance and role; however, this approach is limiting and mechanical and the process of theorization is more than procedural steps.

The aim of this volume is to practically implement a method that, by using Big Data, scholars' views, and interpretable and explainable artificial intelligence, encourages psychologists to develop theories. Hopefully, the proposed methods in the book would be taken as the input of the abductive approach to generate theories. Hopefully, the proposed methods in the book will be taken as the input of the abductive approach to generate theories.

Perhaps this could be called the fourth paradigm in the history of psychological science (Fig. 2.18).

Abduction can be seen as a transition from implausible explainability to plausible explainability, and thus plausibility is the main criterion. In deduction and induction, validity and uncertain reality are considered the main criteria, respectively. Therefore, in AP, robust patterns on the web are extracted using XAF, and these patterns are then used in the process of theory building using the abductive approach. For example, a psychologist assumes that all symptoms of depression are detectable in children with autism. Then, among 20 autistic children attending a particular educational center, there is one child who does not appear to be depressed and is happy. Assuming that our 20-child sample is a valid sample, we can inductively say that not all autistic children are depressed and deductively we can say that all autistic children are depressed. The autistic child who is happy is an anomaly in the population, which may lead to some questions regarding the abductive approach: Is the child really autistic? Does s/he have a different type of autism? Can s/he be classified in another part of the spectrum? Does the child suffer from another disorder?

Let us take another example. Suppose one wants to predict suicide attempts using a multilayer perceptron neural network (or even deep learning) by employing features such as emotional schemas, cognitive schemas, and ontological schemas. After data collection and model preparation, the output would represent the importance of each of these features in classification. By, then, using the necessary patterns, such as SHAP and LIME, these features become interpretable and

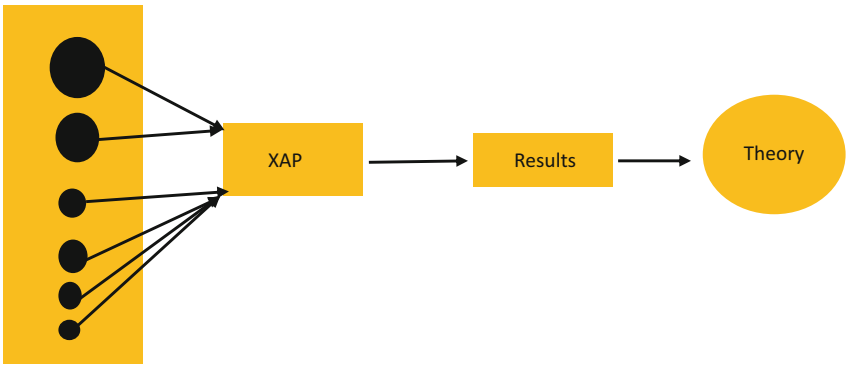


Fig. 2.19 Theory building based on XAP

causative. The derived patterns can be used as input for theorizing using an abductive approach (Fig. 2.19).

Is Big Data a new limitation in AP? How is it possible to use small samples and wish for development in AP?

Internet, digitalization, and the ability to store data on different digital platforms have made Big Data more accessible. Big Data has always existed, but what we did not have was the record of Big Data that makes the fourth paradigm, data exploration, a reality. In psychological research, paper-based questionnaires have changed into hyperlinks that can be easily distributed among participants and accessed online via smart devices. As technology rapidly evolves, it seems that Big Data is becoming much more accessible. On the other hand, some might say that studies with a small sample size are not appropriate for theory development because these studies may not be transparent and their results may not be replicable (Aguinis & Solarino, 2019; Pratt et al., 2019).

So, can we trust the conclusions? It should be said that, in practice, it is possible to find robust patterns in AP using relatively small samples. Unlike classical statistical approaches (or inductive approaches) that rely on hypothesis testing, the main concern with machine learning is overfitting, which can be easily overcome with the methods presented in this book. A review of the literature on machine learning, the core of this volume, shows that more than 100 academic papers have been written using "-iris data." This collection of data contains 150 observation cases involving five features (Dua & Graff, 2017). We know that in some areas of psychological sciences, data are quite expensive. However, there are studies that focus on these problems and difficulties and try to solve the problems of studies with small samples.

Chapter 3
Fuzzy Set Theory and Psychology

3.1 Fuzzy Set Theory and Psychology: Theoretical View

Psychology is the scientific study of mental processes and the behavior of individuals. Fuzzy thinking is an approach for studying the mind. This approach makes the assumption that the brain is a fuzzy inference engine and the mind is a collection of involute fuzzy micro-maps.

Fuzziness, indeterminacy, and overlapping are the main features of this mind, which are generally absent or even rare in classical psychology. Psychology is a science, and therefore psychologists try to correctly observe real psycho-systems or psychological phenomena, to measure and to assess data, to analyze them using the quantitative and qualitative methods, and to interpret results, and if the results have the features of replicability and reliability, they establish principles and then introduce an empirical theory. That is to say, to investigate real-world systems or phenomena, we connect them with a theoretical structure and basis (Seising, 2008).

H. Farahani et al., *An Introduction to Artificial Psychology*,
https://doi.org/10.1007/978-3-031-31172-7_3

This is the heart of science since without accepting this assumption we are not able to study reality. As Wolfgang Blazer (1982) believed, we create this connection between reality and theory and assume that it is possible. Without this vital assumption, it is senseless and meaningless to talk about empirical science. The main goal of this approach is to introduce new concepts and methods for studying the mind and psychological phenomena under the above assumptions.

3.2 The Gray World of Mind

We do not live in the world, but the world is living inside us. This kind of statement is not new but dates back many years. Buddha in the eleventh century B.C believed that phenomena were not completely black or white, nothing was fixed or permanent, and change was always possible. Based on this view, a problem is not a problem; rather, our reaction to that problem is our problem.

Uncertainty or vagueness is a wide part of human experience, language, and perception. Human perception is full of inaccuracy. The real world is not an abstraction; it is not perceived, well-defined and precisely calculated (Wierman, 2010). Vagueness and fuzziness as states of uncertainty are generally considered as a realization that our beliefs and representations of the world are unable to accurately predict future events in our environment (Mushtaq, et al. 2011). It is important to note that uncertainty can be present in many forms. Uncertainty is a cost to be paid for living in the real world (Wierman, 2010). Reviewing the related literature revealed that individual differences affect our dealing with uncertainty (Mushtaq et al., 2011). There are many sources of uncertainty that can impact the research in this field. This uncertainty is in clinical decision-making, disorder diagnosis, and in the measurement of psychological constructs. Here, it is worthwhile noting that the uncertainty is caused by psycho-researchers and participants contributing to the data gathering for research because the mind can be regarded as uncertain. Many mathematical methods exist for dealing with uncertainty, which is inherent in psychological research. A more recent class of these methods are called non-probabilistic methods. Among such methods, we focus on those from fuzzy set theory. This approach in psychological research extends the classical view of the mind (Fig. 3.1).

More formally, the theoretical systems are called independent systems (Blazer, 1982). This means that a researcher gathers data and builds a model as a very simple picture of the real system. Although it is enough in scientific work, this simplification results in a black and white, distorted, and static picture of the world. In other words, a researcher is looking for a straightforward and neat picture of a phenomenon, but

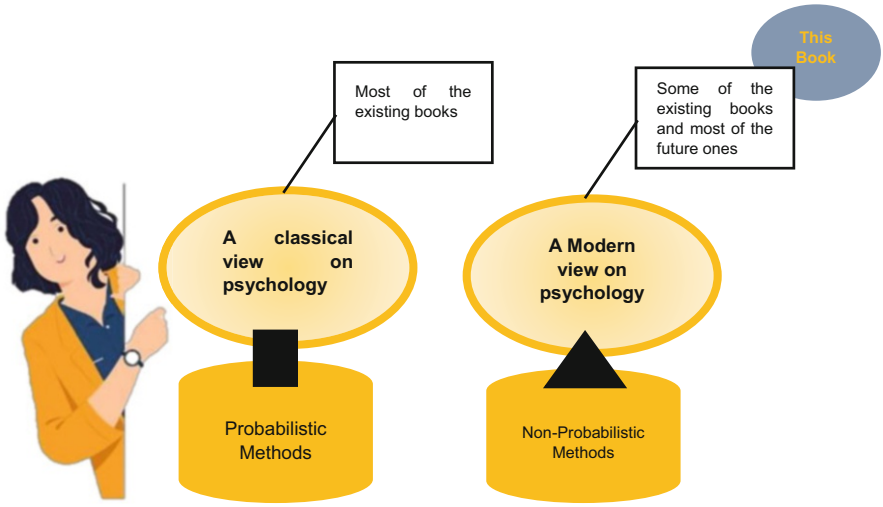
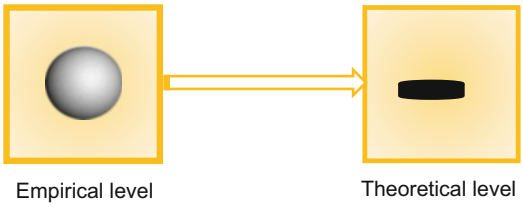


Fig. 3.1 The niche of this book in psychology

Fig. 3.2 From a reality to a model



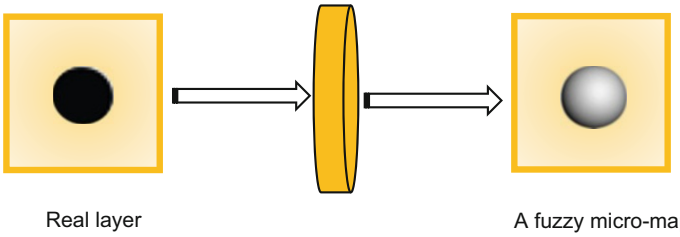


Fig. 3.3 From reality to a fuzzy micro-map

this view is not able to hook onto a dynamic and elusive psychological process. There is, therefore, a gap between reality and these theoretical systems (Fig. 3.2).

The main question which arises here is what the gap is and how to bridge this gap with reality. What is missed in this point of view is our perception of reality, which is between the empirical level and the theoretical level. In summary, a distinction between real systems and perceptions of these entities leads to a modification of the structuralism approach, which pertains to the empirical level. This modification can be attained utilizing fuzzy set theory. This new approach can change the previous model of studying the mental process (see Fig. 3.3).

In this new perspective, the real layer is maintained, but the former empirical level is replaced by a fuzzy layer. We agree with Seising (2008), who believes that the fuzzy layer is a subjective structure that is imposed by an observer's perception. This fuzzy layer implies that our perception of reality is vague, ambiguous, and uncertain. The fuzzy layer leads to a mind covered by a "fuzzy micro-map." A fuzzy micro-map is an individual's perception of a psychological event. A fuzzy psychologist tries to capture this map using a detailed interview with the individuals. A fuzzy psychologist draws this map by analyzing the story that has been told by the individuals. The main goal of this field is to introduce methods for gathering the fuzzy micro-maps and then to combine and aggregate them as a fuzzy combined map and make inferences using fuzzy set theory. In other words, "what we observe is not nature itself, but nature exposed to our method of questioning." This statement by Heisenberg is one of the most important statements in the history of science. Based on this view, the results from research in studying the mind are contaminated by the specified thinking which developed them. Our perception of a phenomenon is imposed by that thinking. This thinking is called the fuzzy layer, which is the foundation of this approach. The fuzzy layer yields imprecise and imperfect information, but this information is more dynamic, reasonable, and consistent with the real world. The main components of this model are the fuzzy layer and the fuzzy micro-map; the fuzzy inference system is produced from the application of the fuzzy set theory, and it leads to some innovative methods for making fuzzy inferences based on them. This is a consistent feature of this approach.

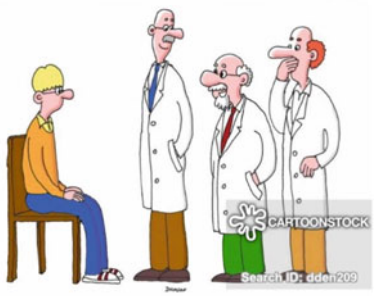
We need an impartial person for the judgment

3.3 The Fuzzy Logic Under Psychological View

In a classical view on psychology, we easily see that psychology as a science must describe, explain, and predict psychological phenomena. Obtaining this final goal is not always easy. In order to achieve this goal and to preserve the scientific nature of phenomena in psychology, statistical methods or Fisherian statistics are usually employed. Although these methods have influential effects on this science, they do not take into account that the mind is an overlapping, dynamic, and integrative map. Simply put, there is no sharp line among latent traits or psychological constructs because, under the assumption of this approach, the mind is a collection of fuzzy integrated overlapping schemas or maps. The boundaries among them are not crisp;

they are instead rather fuzzy and ambiguous like a spider's web. Fuzzy inference systems rely on these assumptions. In other words, these difficulties are rooted in the nature of a human's mind. Mentality does not consist of distinct parts but is a dynamic, flexible, and complicated whole.

Lack of complete information coupled with the imprecise and controversial nature of the mental process and its states leads to challenges in psychology, diagnosis, research, treatment, prediction, and classification, and in the construction of theory.



"Your psychosis is so original, that we're still brainstorming about a good name for it!"

The best and most precise descriptions of the psychological states and traits are made by individuals' linguistic terms. These terms are imprecise and vague. A given disorder may manifest itself quite differently, depending on its intensity and the individual characteristics of the patient. A single symptom may also correspond to different disorders (Torres & Nieto, 2006). On the other hand, some disorders are considered as non-specific, with, for example, the need for new theories to be developed to model mixed emotions. For example, some evidence indicates that people can feel happy and sad at the same time (Larsen & McGraw, 2011; see Fang et al., 2018a, b). Based on the classical view, normality (nondiseased) and abnormality (diseased) are mutually exclusive and even may be opposites. This view originated from Aristotelian logic, which held sway for around 2000 years in human reasoning, knowledge, and science. Everything must either be or not be. A predicate either belongs or does not belong to a given subject in a given respect at a given time. That is either A or $\neg A$. The modern formulation of the above rule is $\forall x(A(x) \bigvee \neg A(x))$.

Accordance with this emotionless view of classification, every proposition or state is of only two logical values: true or false (1 or 0). In the real world, however, not everything can be classified into either black or white; rather, the real world is colored by our perception. Therefore, most of the time, our mind's view of the world is gray. Let us demonstrate this with a simple psychological example using a statement that is an item of the Minnesota Multiphasic Personality Inventory (MMPI). This test and its different versions are one of the most commonly used and important tests in psychological settings and research. The statement "I think many people exaggerate their misfortunes to gain the sympathy and help of others"

is partly true and partly false, depending on the "states and situations" of individuals. Consider this statement as another example: "I am a famous person." If you are the president of a superpower country, this is true, but if, however, you are a shepherd in a small marginal village, then it is false. We are talking here of the relative fame of two sorts of people. Everybody is well-known (W) to some extent and unknown (N) to some extent. If you are famous, $W = 1$, and as each of us has just some degree of reputation, then $W < 1$.

$$W + N = 1$$

$$W + I + N = 1$$

(I) in the $W + I + N = 1$ denoted as "I don't know."

Let us pay attention to the diagnostic criteria of the obsessive-compulsive disorder in the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5). Criterion A states that the presence of obsessions, compulsions, or both are defined by the presence of two factors: First, by recurrent and persistent thoughts or impulses that are experienced, at some time during the disturbance, as intrusive and unwanted, and that in most individuals cause marked anxiety or distress; second, by the attempts of the individual to ignore or suppress such thoughts, urges, or images or to neutralize them with some other thought or action (i.e., by performing a compulsion). Compulsions are defined by the following:

1. Repetitive behaviors (e.g., hand washing, ordering, checking) or mental acts (e.g., praying, counting, repeating words silently) that make the individual feel driven to perform in response to an obsession or according to rules that must be applied rigidly.
2. Behaviors or mental acts aimed at preventing or reducing anxiety or distress, or preventing some dreaded event or situation; however, these behaviors or mental acts are not connected in a realistic way with what they are designed to neutralize or prevent or are excessive. Note: Young children may not be able to articulate the aims of these behaviors or mental acts. Criterion B states that the obsessions or compulsions are time-consuming (e.g., take more than 1 h per day) or cause clinically significant distress or impairment in social, occupational, or other important areas of functioning. Criterion C states that the obsessive-compulsive symptoms are not attributable to the physiological effects of a substance (e.g., a drug of abuse, a medication) or another medical condition. Criterion D states that the disturbance is not better explained by the symptoms of another mental disorder (e.g., excessive worries, as in generalized anxiety disorder, preoccupation with appearance, as in body dysmorphic disorder, difficulty discarding or parting with possessions, as in hoarding disorder, hair pulling, as in trichotillomania (hair-pulling disorder), skin picking, as in excoriation (skinpicking) disorder, stereotypies, as in stereotypic movement disorder, ritualized eating behavior, as in eating disorders, preoccupation with substances or gambling, as in substance-related and addictive disorders, preoccupation with having an illness,

as in illness anxiety disorder, sexual urges or fantasies, as in paraphilic disorders, impulses, as in disruptive, impulse-control, and conduct disorders, guilty ruminations, as in major depressive disorder; thought insertion or delusional preoccupations, as in schizophrenia spectrum, and other psychotic disorders or repetitive patterns of behavior, as in autism spectrum disorder). For many years, this disorder belonged to a bigger disorder class called anxiety disorder, but in 2012, it was reclassified. It is obvious that the nature of this disorder has not changed; rather, the definition has been changed and this kind of change may occur in the future as well. This example implies that the boundaries among psychological constructs are not well defined but are instead fuzzy and vague. This means that we face a partial inclusion of the categories (where we have at least two categories). This is a key component in fuzzy thinking according to Wierman (2010). Although fuzzy set theory has been used successfully in many fields of science for more than 50 years, there exist just a few research works, mostly articles, in psychology using this theory. While the existing approaches do not consider any practical aspect or software code that can be easily used by psychologists, they can be considered as a theoretical starting point for applying the fuzzy set theory in psychology, for example, Smithson (1982), Zetenyi (1988), Smithson and Oden (1999), Ragin (2000), Smithson and Verkuilen (2006), and Arfi (2010). Some recent examples of fuzzy set theory in psychology include the following:

- A fuzzy logical model of perception (see Oden & Massaro, 1978; Massaro, 1989; Massaro & Cohen, 2000; Martínez-Jiménez et al., 2018)
- Fuzzy set-based theory of memory and attention (Perfilieva & Vajgl, 2015; Terziyska et al., 2015)
- Fuzzy decision-making (see Khefacha & Belkacem, 2015)
- Fuzzy psychopathology (see Horowitz & Malle, 1993; Mosoiu et al., 2010; Ekong et al. 2013; Reinertsen et al., 2017; Ashish et al., 2018)
- Fuzzy consciousness (see Huette & Spivey, 2012)
- Fuzzy measurement and testing (see Stoklasa et al., 2011; Farahani et al., 2018, 2019)
- Fuzzy methodology (see Sugeno & Yasukawa, 1993; Stoklasa et al., 2014)
- Fuzzy epistemology (see Seising, 2008)
- Fuzzy clinical diagnosis (see Baig et al., 2011; Erin & Abiyev, 2019)

3.4 Why Fuzzy Logic Theory?

According to the literature, it is reasonable to argue that fuzzy logic and its derivatives are helpful to psychologists dealing with the ill-structured and ill-defined phenomenon. Many psychological constructs are not well-defined due to their nature. For example, categorizing patients into two distinct groups is often difficult because overlapping and partial inclusion is not possible. Our mind is continuous and full of overlapping fuzzy micro-maps. Thus, it is possible to face many cases,

which may belong to at least two categories. Finding a sharp and crisp line among psychological concepts in soft science is very difficult. The nature of continuity of mind implies that our mind is much more similar to overlapping cloud cubes in a rainy sunset than distinct sand particles on a shore. Categorization is much more difficult when we are too close to a threshold. In reality, many of the psychological traits such as emotion, cognition, and disorder are mixed and of different magnitudes. For example, an extrovert person may have some degrees of introversion and vice versa. Depressed persons may have high scores on only some symptoms, and in the other symptoms, the scores can be in the middle or low. The concept of the stage is important both in the old psychological theories of Piaget or Freud and in new theories that are being built and extended in cognitive psychology and neuroscience. These stages are not distinct and separate from each other but are instead overlapping. This continuity can be accommodated using a fuzzy psychological view. Psychological data and linguistic data are the same, and the fuzzy logic theory is a powerful tool for quantifying linguistic data. In psychology, information from individuals is collected using standardized tests or interviews; therefore, information is provided in words. The words are full of imprecise information that needs to be quantified. A concept can mean different things to different people. For example, when two people tick "very high" as a response to the item "I am happy with my life," they may have different mindsets of very high, although they chose the same answer. "Very high" does not represent a universal standard but varies depending upon the individual. You may feel like rating it 5 and your friend 10, even though it is very high for both of you.

Quantitative psychological research mostly relies on classical statistics in which testing the null hypothesis is essential. The null hypotheses are tested directly, and the alternative or research hypotheses are tested indirectly. In null hypothesis testing, the p-value is a criterion for making a statistical decision on acceptance (true) or rejection (false) of that null hypothesis. Most psychological researchers analyze their obtained quantitative data until a p-value smaller than 0.05 is observed; this may lead to an inflation of Type 1 error. Many scientists have criticized p-values (e.g.,

Simmons et al., 2016; Cumming, 2011), and they believe that p-hacking is widespread throughout science (Head et al., 2015a, b) where we increase the significance of "just significant results" (Leggett, 2013). This means that psychological researchers, like other researchers, are interested in publishing only the significant results, and we observe the surge of p-values and the file drawer effect. This means what is not significant statistically has been ignored and hidden (see Lakens, 2015). These challenging issues threaten the robustness of scientific knowledge. In fuzzy inference systems, the extent of alternative hypotheses is tested directly. That helps us to capture a more realistic picture of a mental event, process, and phenomenon. Fuzzy inference systems aim to rethink psychological theories and find new results and try to solve what is called a "crisis of confidence" in psychological research. The issue has drawn the attention of Bayesian statisticians (see Marsman & Wagenmakers, 2017; Wagenmakers et al., 2018). Clinical decision-making is a crucial activity in psychology; how do we make decisions in a clinical setting using fuzzy logic? Fuzzy inference systems focus on how a clinical psychologist can build a clinical model and make a fuzzy decision. Based on this method, a psychologist is capable of considering many factors for diagnosing a disorder or determining the efficacy and effectiveness of psychological treatment. In summary, the main goal of a fuzzy inference system is to capture the psychological concepts that are reflected in the linguistic concepts based on qualitative and quantitative research.

This approach bridges the depth of qualitative research and the precision of quantitative research. This bridge is interesting and of use in psychology. As we have already mentioned two-valued logic cannot adequately model fuzzy systems. The history of developing and paying attention to fuzzy thinking and theory is long and interesting. This journey represents the development of attempts to model "vagueness" in the history of science. What we are aiming to do here is to show the necessity of having this view in studying the mind and the need for fuzzy inference systems.

Is fuzzy inference a reasonable field of psychology? Reviewing the literature indicates that there are many pieces of research in which cognition and emotion have been investigated. Many of them have been done based on classical statistics and, therefore, they could test just simple models. For example, we can find that the cognitive system has an underlying interactive network (Stephen & Mirman, 2010; Spence et al., 2004) with meaningful noise conditions, which are useful in computation (Kello et al., 2008; Kello & Van Orden, 2009) and top-down constraints on perceptual processing (Motter, 1993; Spivey & Spirn, 2000; Gandhi et al., 1999; Ito & Gilbert, 1999; Lamme & Roelfsema, 2000). Although these research attempts are important and enlightening, opening our eyes to many complicated phenomena, it is worthwhile noting that there are still reasons for developing and extending fuzzy inference systems as a major research line. Almost all of this research is based on the linearity assumption, a condition that simplifies the world and may afford a way to reality and yield us a robust model. Huette and Spivey (2012) believe, however, that the most difficult part of transitioning to a descriptive and accurate model of consciousness is to abandon linear causality. The fuzzy inference system uses fuzzy logic to help us more closely represent human thinking and does not make an assumption of linear relationships. Huette and Spivey (2012) have extracted some

common points based on reviewing the research literature about consciousness. They also coined the term "fuzzy consciousness". These common principles include the following: (a) The everyday experience of consciousness is noisy, imperfect with at best partial information. Most of one's time is spent by moving through a space of concepts, percepts, and emotions, never quite fully reaching any pure concept in a context-free manner. This means that at any given time, consciousness is defined by the many thousands of environmental inputs in a natural environment, the constraints of many billions of neurons associated with previous learning experiences and a framework equipped with high-dimensional sensation and movement parameters. All of these variables combine to form what is reported as consciousness. (b) All relevant variables should be considered. This is based on the linearity assumption and classical statistics. Even considering all variables, at a given time, we may face some thoughts and behaviors that are unusual. For example, suppose that Shannon is a postgraduate student and she has been informed that her paper submitted to a journal was rejected yesterday and she must revise it and submit it to another journal soon. She is divorced and alone and her mother is so sick and her father passed away last year. You see her sitting alone and smoking in the door café of your university and you find her feeling depressed. These states of her consciousness are expected, but you may also see Shannon beginning to laugh suddenly. This is not an expected state of her consciousness, given she is depressed, but it is possible that she may have remembered a joke just for a millisecond. Under some researchers like Huette and Spivey (2012), we believe that a fuzzy consciousness can account for this surprising behavior. (c) Evidence for the fuzzy consciousness should be given. This new theory is a nonlinear and fuzzy system (Huette & Spivey, 2012). The continuity of the mind, which has been argued by Spivey et al. (2005), is a new view for omitting the crisp boundaries of the mind. Simply put, the mind acts continuously, but a given behavior may be taken as a discrete one, in isolation, by an external observer.

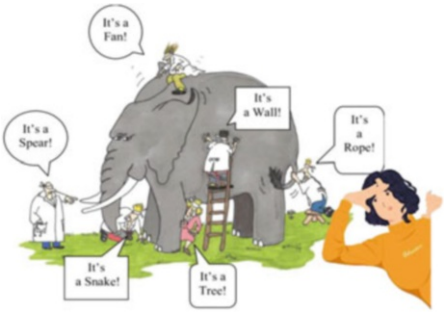
There is philosophical evidence for the importance and necessity of the fuzzy view of psychology; for example, Taylor (2022) presents a new philosophical view of perception and argues that perception is a messy cluster that is supported by mechanisms that don't always work. There is a fuzzy transition between the nonperceptual and the perceptual. This is an entirely new way of thinking about perception, completely unexplored in philosophy. There is some neurological evidence for the necessity of a fuzzy view on psychology. There is not much evidence in neurological research that fuzziness has been considered, although Rayan et al. (2015) indicated that there are shared performance impairments in cognitive control among patients with mood disorders (MDD and BD) relative to healthy controls. They also showed that the neurobiological bases underlying these seemingly shared dimensions of impairment are not as clear cut as we would like due to larger more diverse independent groups.

The nature of fuzzy logic is very close to the real world, due to the capacity of quantifying the linguistic variables, which we use for asserting our emotions and cognitive reasoning, judgment, and decision-making. It follows that a fuzzy inference system helps us to solve a "crisis of confidence" in psychological research, which is due to the use of classical statistics overlooking more complex relationships. One important assumption of a fuzzy inference system is the fuzzy layer. We

can think of this layer as representing our perception. It is important to note that perception is the core of cognition and emotion. Mao (2018) believes that an entity is not simple but complex and inter-related with other things. In other words, reality is the state of things, past and present, whether or not they are observable or comprehensible. Perception is the interpretation of the stimuli based on experiences we may have about them. That means that the brain is an active participant in constructing what we perceive. The perceptions that the brain creates are the result of an interaction between the signals received and what it does to them. To understand perception and the knowledge that we acquire through it, we must, therefore, enquire not only into the nature of the signals that the brain receives but also into the contribution that the brain makes and the limitations that its characteristics impose upon the acquisition of knowledge (Kant, 1908; Schopenhauer et al., 1859; as cited in Zeki, 2001). Zeki (2003) believes that the brain has evolved as a flexible tool for acquiring knowledge about both unambiguous and ambiguous conditions. The ambiguous conditions are not rare; rather, we commonly encounter them. The ambiguous condition is the condition in which two or more interpretations exist, each one of which has equal validity with the others. However, we can only be aware of one interpretation at any given moment (Zeki, 2003). This indicates that our perception is what we call a fuzzy layer. Based on Mao (2018), usually, a thing (T) is multilateral or complex, and so to hold on to its reality is difficult for human beings, where the complexity of the world implies the cognitive system on a thing (T) is itself complex, i.e., a system composed of many components that may interact with each other. Recently, some research has implied that we can experience mixed emotions (see Fang et al., 2018a, b; Keltner et al., 2019). Van der Heide et al. (2011) indicated that a human can experience (1) mixtures of emotions and/or (2) conflicting emotions, and Mao (2018) describes this ambiguity in a mathematical framework. In fuzzy inference, we formulate such a framework by applying fuzzy set theory and its derivatives.

3.5 What Is the Fuzzy Map?

Let us take a typical example accounting for the complexity of a cognitive system. It is the well-known fable "the blind men with an elephant."



This is a famous story in Buddhism, which implies the whole is more than its parts but we always tend to focus on the parts. In this story, there are six blind men

who were asked to determine what an elephant looked like by feeling different parts of an elephant's body. The men touched the elephant's leg, tail, trunk, ear, belly, or tusk and they respectively claimed it's like a pillar, a rope, a tree branch, a hand fan, a wall, or a solid pipe.

Similarly, Rumi (1207–1273) included it in his approach. In his retelling, "The Elephant in the Darkness," some people bring an elephant to be exhibited in a dark room. Many men touched the elephant in the darkness and described whatever they happened to touch. This fable is close to the famous statement of Werner Heisenberg (1901–1976). He stated that natural science does not simply describe and explain nature. It is, instead, part of the interplay between nature and ourselves. In this well-known fable, the men's perception of the part of the elephant is akin to a "fuzzy micro-map." Although there is a term which is called a cognitive map in the fuzzy literature, that is a map captured through experts investigating factors influencing a phenomenon. A fuzzy micro-map is a network that is obtained from a cognizant. In a fuzzy inference system, a cognizant is a person who produces a fuzzy micro-map of his or her perception of a phenomenon. His or her perception of a phenomenon is the same fuzzy layer and the theoretical structure as an output of the perception process is the fuzzy micro-map. A fuzzy psychologist tries to collect the fuzzy micro-map using a data-gathering method from the cognizant's point of view. In our elephant example, each fuzzy micro-map is the name of the part of the elephant mentioned above; therefore, all of the fuzzy micro-maps need to be combined into a final testable map. It is worth noting that the fuzzy micro-map is obtained from the data given by a cognizant. As we have seen, a cognizant is a person who has information and experiences of a system (each man touching a part of the elephant). Any human being may be a cognizant. The data are obtained using methods including storytelling by the cognizant, a transcription of the recorded detailed interview with the cognizant, MRI, EEG, QEEG, and EMG. The fuzzy micro-map is a readable mind map of a cognizant. After providing all the fuzzy micro-maps, a fuzzy psychologist combines them using an expert panel and provides a final map. This final map will be the basis for the fuzzy inference. Thus, fuzzy mapping helps us to find a real model of an elusive phenomenon. The fuzzy map has higher flexibility, an ability to capture more significant complex relationships, and more generalizability than traditional approaches.

The main purpose of research in psychology is to model the mind and define normality and abnormality, and these are not usually clearly separated. Most of the samples in psychological research are small and non-random; therefore, statistical inference is not a robust method for generalizing the results. The use of the P-value as the heart of null hypothesis testing in psychological research has been strongly criticized. Bayesian statistics require the specification of a prior distribution, which is not easy to define. Qualitative research can be an appropriate alternative. Although this methodology can be useful, its precision is somewhat vague. A fuzzy inference system, on the other hand, is a robust method for capturing and explaining the mind and its processes. This approach bridges the depth of qualitative research and the breadth of quantitative research and solves many difficulties by using different sources of information and integrating them to make inferences in a way that more closely resembles the reasoning of the mind.

3.6 Fuzzy Modelling of Psychological Systems

Neurosis is the inability to tolerate ambiguity (Sigmund Freud).

According to fuzzy logic, a concept is hardly ever completely true or completely false, but it is rather somewhere in between these two extremes.

A simple example of this difference is shown in Fig. 3.4 in which instead of a singleton number defining the number *x*', as we instinctively use in everyday logic, the concept of degree of similarity (or degree of truth or degree of membership) is introduced, defined by the membership function (mf) A x a: μ (). The interpretation of fuzzy sets () has arisen from the generalization of the classical sets to embrace the vague notions and unclear boundaries. It may not be always clear, if an element *x* belongs to a set *A*, or not. Thus, its membership may be measured by a degree, commonly known as the membership degree, taking a value from the unit interval by agreement.

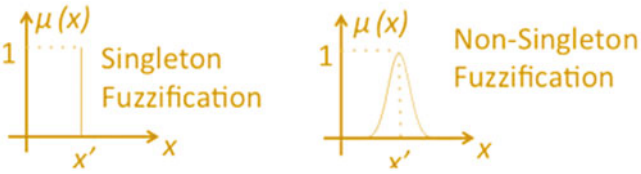


Fig. 3.4 A singleton and non-singleton fuzzification

Consequently, a fuzzy set A over the universe of discourse X is defined by function μA that matches each element of the universe of discourse with its membership degree to the set A

$$\mu A(X) : X \rightarrow [0, 1] \tag{3.1}$$

where $\mu A(X) = 0$ says that an element x definitely does not belong to a fuzzy set A and $\mu A(X) = 1$ says that x without any doubt is member of fuzzy set A. Higher value of $\mu A(X)$ indicates the higher degree of membership of an element x to a fuzzy set A. Each fuzzy set is defined by one membership function. A membership function maps each element of the universal set X into real numbers from the [0, 1] interval. We should emphasize that the universal set X is always a crisp set (). A fuzzy set can be defined as a set of ordered pairs:

$$A = \{(X.\ \mu A(X))|\mathrm{x} \in \mathrm{X} \wedge \mu A(x) \in (0.1)\} \tag{3.2}$$

When the universal set is finite, a fuzzy set constructed on this universal set can be expressed by counting the elements and their respective membership degrees:

$$A = \frac{\mu A(X1)}{X1} + \frac{\mu A(X2)}{X2} + \ldots + \frac{\mu A(X_n)}{X_n} \tag{3.3}$$

3.7 Properties of Fuzzy Sets

In this section, properties relevant for the next sections are examined. First, we define scalar and relative scalar cardinality. For any fuzzy set A defined on a finite universal set X, we define its scalar cardinality by the formula

$$\mathrm{card}(A) = |A| = \sum \mu A(x) \tag{3.4}$$

The scalar cardinality of a fuzzy set is a generalization of classical cardinality.

Elements of a universal set belong to the fuzzy sets with different membership degrees and therefore we cannot count elements of a set A, but their respective membership degrees should be summed. Some authors refer to $|A|$ as the sigma count of A (Dubois & Prade, 2005).

The relative scalar cardinality is defined by the formula:

$$\left\|A\right\| = \frac{\mathrm{card}(A)}{\mathrm{card}(X)} \tag{3.5}$$

where card(A) is defined in Dubois and Prade (2005) and card(X) represents the number of elements in X. These cardinalities are broadly used in areas such as

linguistic summaries. The third type of cardinality is fuzzy cardinality, expressed as an ordered pair defined as the number of elements belonging to a particular α-cut when the universal set is finite. Cardinalities are closely examined; e.g., scalar cardinality of a fuzzy set can be expressed as the area bounded by the membership function of fuzzy set and the x-axis. This approach is demonstrated on the trapezoidal fuzzy set (Dubois & Prade, 2005).

- Support: The support of a fuzzy set A is the finite set with the following property:

$$\text{supp}(A) = \{(x \in X \mid \mu A(x) > 0\} \tag{3.6}$$

- Core: The core of a fuzzy set A is the set with the following property:

$$\text{core}(A) = \{(x \in X \mid \mu A(x) = 1\} \tag{3.7}$$

  In the fuzzy sets literature, the term kernel is used as a synonym for the core.

- Height: The height is the highest value of the degree of membership of all elements in the considered fuzzy set A:

$$h(A) = h(A) = \sup\ \mu A(x).x \in X \tag{3.8}$$

- Normalized fuzzy set: Fuzzy set A is normalized, if the degree of membership of at least one element is equal to 1, i.e.,

$$\exists x \in X.\ \mu A(x) = h(x) = 1 \tag{3.9}$$

- Crossover point: The element xcp of a fuzzy set A that has a membership degree equal to 0.5 represents the crossover point, i.e.,

$$xcp = \{x \in X | \mu A(x) = 0.5\} \tag{3.10}$$

α-cut and strong α-cut: One of the important concepts used in fuzzy sets is the α-cut. The α-cut $A(\alpha)$ and its restrictive variant strong α-cut $A(\alpha+)$ are defined in the following way:

$$A(\alpha) = \{x \in X \mid \mu A(x) \geq \alpha\} \tag{3.11}$$

$$A(\alpha+) = \{(x \in X \mid \mu A(x) > \alpha\} \tag{3.12}$$

where $\alpha \in [0, 1]$.

The α-cut of a fuzzy set A is a set containing all the elements of set X whose membership degrees in A are greater than or equal to the specified value of α. This property is used in many areas, e.g., working with elements in a fuzzy set associated

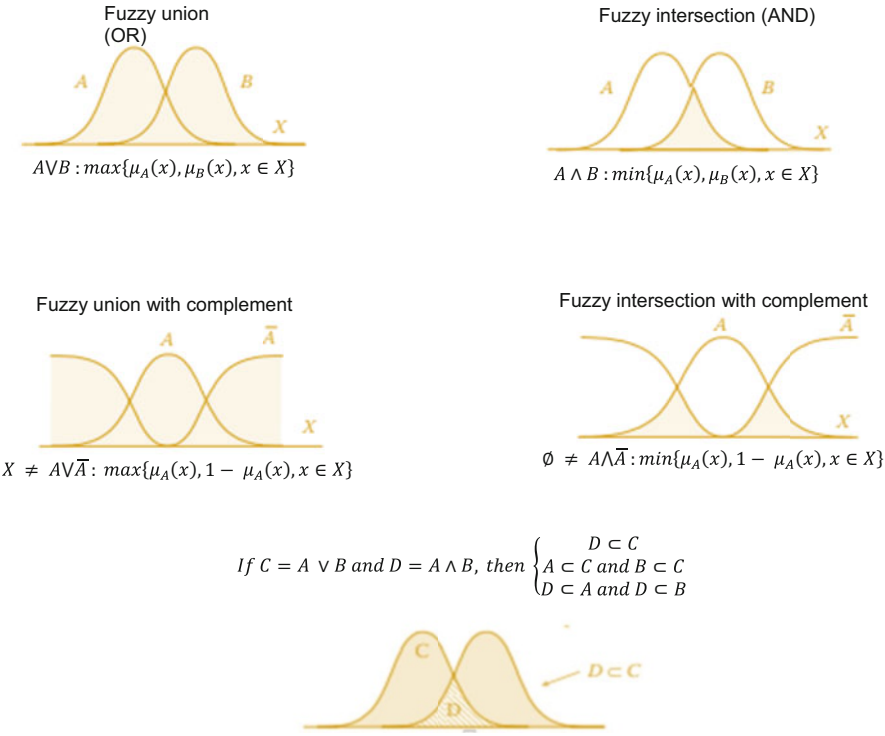


Fig. 3.5 Logical operations on fuzzy sets. The shaded areas represent the result of the operation. (Klir & Yuan, 1995a, b)

with a high likelihood of membership. A fuzzy set is convex if and only if (Pourabdollah et al., 2020)

$$\mu A(\lambda x + (1-\lambda)y) \geq \min(\mu A(x).\ \mu A(y)) \tag{3.13}$$

for all x and $y \in X$ and all $\lambda \in [0, 1]$. Convex and non-convex fuzzy sets are plotted in Fig. 3.5.

3.8 Types of Fuzzy Sets (Membership Functions)

Membership functions are classified into two main groups: linear and Gaussian or curved. All membership functions explained in this section are normalized fuzzy sets, i.e., with one element having a degree of membership equal to 1 (Hasan & Sobhan, 2020).

Triangular fuzzy set is defined by its lower limit a, its upper limit b, and the modal (highest) value m as

$$\mu_{A(x)} = \begin{cases} 1 & \text{for } x = m \\ \dfrac{x-a}{m-a} & \text{for } a < x < m \\ \dfrac{b-x}{b-m} & \text{for } m < x < b \\ 0 & \text{for } x \le a \bigvee x \ge b \end{cases} \tag{3.14}$$

Gaussian fuzzy set defined by the modal value (center) m and width k as

$$\mu A(x) = e^{-k(x-m)(x-m)} \tag{3.15}$$

The bell of the Gaussian function depends on the value k. If the value k is lower, then the bell is narrower.

Trapezoidal fuzzy set is defined by its lower limit a, its upper limit b, and the flat segment $[m_1, m_2]$ representing the highest value of height (3.16) as

$$\mu_{A(x)} = \begin{cases} 1 & \text{for } m_1 \le x \le m_2 \\ \dfrac{x-a}{m_1-a} & \text{for } a < x < m_1 \\ \dfrac{b-x}{b-m_1} & \text{for } m_2 < x < b \\ 0 & \text{for } x \le a \bigvee x \ge b \end{cases} \tag{3.16}$$

3.9 Practical Example Using R

The R programming language began in 1992 as an effort to create a special-purpose language for use in statistical applications. More than two decades later, the language has evolved into one of the most popular languages used by statisticians, data scientists, and business analysts around the world. R gained rapid traction as a popular language for several reasons. First, it is available to everyone as a free, open-source language developed by a community of committed developers. This approach broke the mold of past approaches to analytic tools that relied upon proprietary, commercial software that was often out of the financial reach of many individuals and organizations. R also continues to grow in popularity because of its

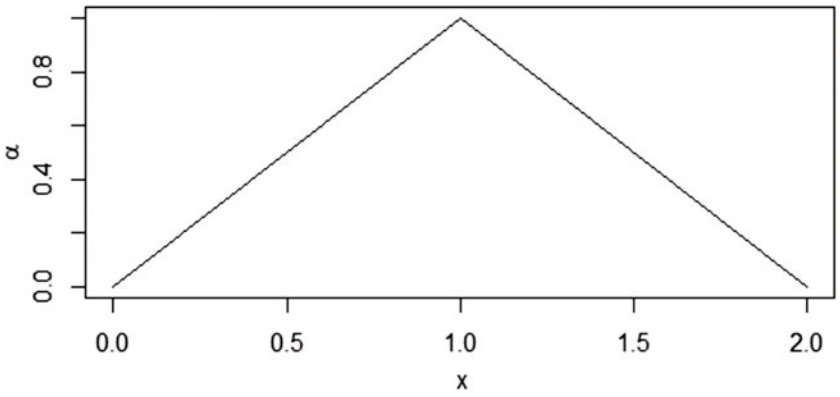


Fig. 3.6 R results for plotting of each fuzzy set element

Table 3.1 Fuzzy set parameters for linguistic terms

Linguistic terms (fuzzy set)	a	b	c
Very low	**0**	**1**	**2**
Low	**2**	**3**	**4**
Middle	**3**	**5**	**6**
High	**6**	**8**	**9**
Very high	**9**	**10**	**11**

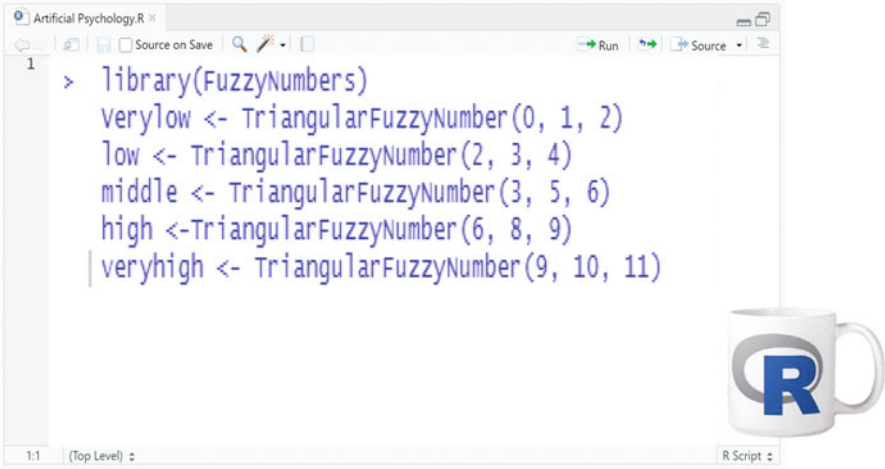
adoption by the creators of machine learning methods. Almost any new machine learning technique created today quickly becomes available to R users in a redistributable package, offered as open-source code on the Comprehensive R Archive Network (CRAN), a worldwide repository of popular R code. Figure 3.6 shows the growth of the number of packages available through CRAN over time. As you can see, the growth took off significantly over the past decade. it's also important to know that R is an interpreted language, rather than a compiled language. In an interpreted language, the code that you write is stored in a document called a script, and this script is the code that is directly executed by the system processing the code. In a compiled language, the source code written by a developer runs through a specialized program called a compiler, which converts the source code into executable machine language.

Suppose, we need to use a fuzzy set for demonstrating the degree of depression in a sample of multiple sclerosis (MS) ranged between 1 and 10 (Table 3.1). We consider the amount of depression in terms of a triangular fuzzy number for each linguistic terms, which may be, for example, responses by clinicians assessing the amount of various aspects of depression (denoted in the table as a, b, and c) in an individual.

Fuzzy set parameters of very low to very high for depression

Based on Table 3.1, we are going to define a triangular distribution for each fuzzy set element (ranging from very low to very high) and determine membership degree for various numbers. You can see the R codes as below in RStudio, an add-on interface to R (Listing 3.1).

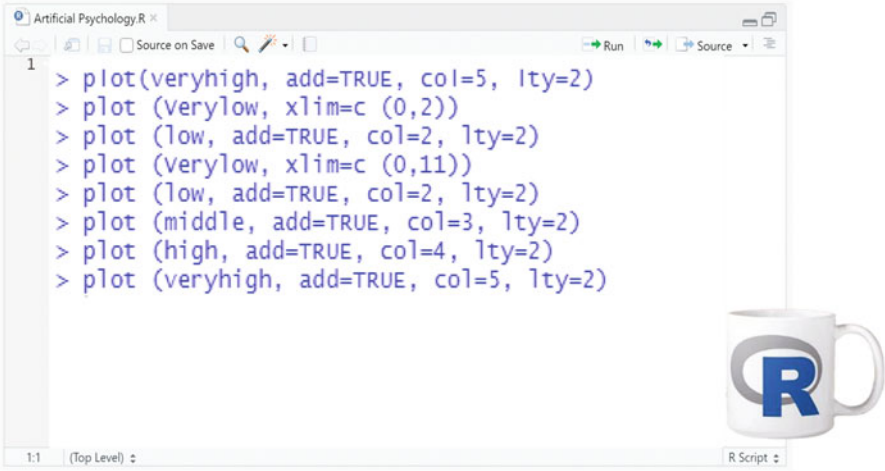
Listing 3.1 R codes for defining a triangular number for each fuzzy set element



Using plot (very low, xlim = c (0,2)), the plot of the triangular distribution is provided for the very low fuzzy set (Fig. 3.6).

Using the following codes, we can show the distributions of membership degree of all elements of the fuzzy set in the same plot (Listing 3.2).

Listing 3.2 R codes for plotting all of the fuzzy set



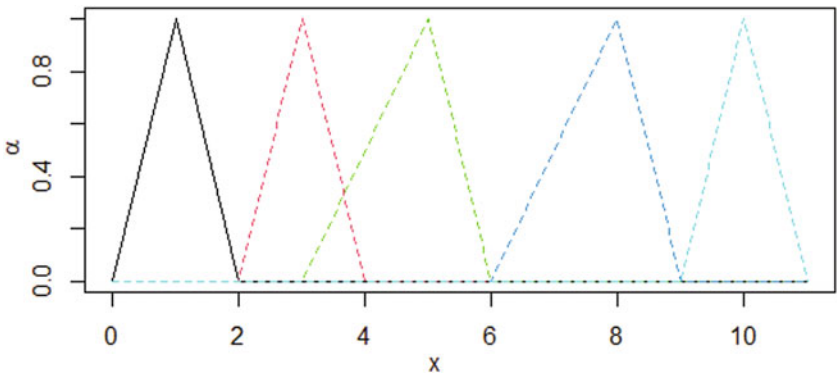


Fig. 3.7 R results of plotting of all elements of the fuzzy set

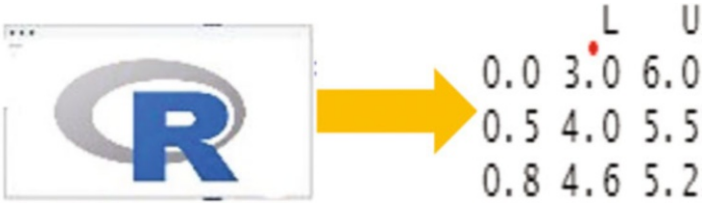


Fig. 3.8 The result of the code for middle

With a simple code, the support and core values can be accessed. The supports for very high are 9 and 11, and there are cores, which are 10 and 10 (Fig. 3.7).

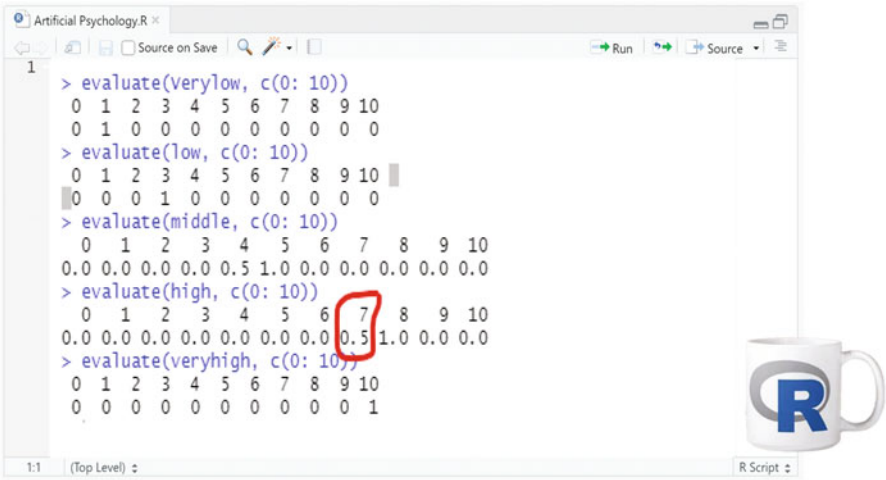
```
> supp(veryhigh)
[1] 9 11

> core(veryhigh)
[1] 10 10
```

For determining α-cuts, just need to use `> cuts <- alphacut(A, c(0, 0.5, 0.8))`, for every fuzzy set A. You can see the result of the code for middle (Fig. 3.8).

```
cuts <- alphacut(middle, c(0, 0.5, 0.8)
cuts
```

Suppose there is a patient suffering from MS who has a score of 7 on a depression scale. We wish to determine the degree to which this patient belongs to each fuzzy set element. Let us determine the membership degree for all scores from 1 to 10, including 7. To do this, we write a simple code using the function, evaluate() (Listing 3.3).

Listing 3.3 R codes for evaluation of the fuzzy membership function



As you can see, for a score of 7, membership degree is 0.5 only for high and 0 for the rest. This shows that there is a 50% chance of someone with a score of 7 on the depression scale being given a "high" rating by the clinicians. We can calculate membership degree for each of the depression scores between 1 and 10 for people rated by the clinicians as highly likely to be depressed.

```
> evaluate(high, seq(1, 10))
  1   2   3   4   5   6   7   8   9  10
0.0 0.0 0.0 0.0 0.0 0.0 0.5 1.0 0.0 0.0
```

3.10 Fuzzy Set Composition

A suitable tool for interpretation of the "AND" or connective (conjunction) in fuzzy logic are triangular norms (or short t-norms) (Nguyen & Walker, 1977). Relevant mathematical aspects of t-norms are discussed in depth in (Nguyen & Walker, 1977). Theoretically, there are an unlimited number of t-norms. The four basic and commonly used t-norms are given below (Wanga et al., 2022).

Minimum: $tm(\mu A1\ (x) \cdot \mu A2\ (x)) \quad = \min(\mu A1\ (x) \cdot \mu A2\ (x))$ (3.17)

Product: $tp(\mu A1\ (x) \cdot \mu A2\ (x)) = \mu A1\ (x) \cdot \mu A2\ (x)$ (3.18)

$$\textbf{Łukasiewicz t-norm}: tL\left(\mu A1\left(x\right)\cdot\mu A2\left(x\right)\right)=\max(0\cdot\mu A1\left(x\right)+\mu A2\left(x\right)-1) \tag{3.19}$$

$$\textbf{Drastic product}: td\left(\mu A1\left(x\right)\cdot\mu A2\left(x\right)\right)=\begin{array}{ll}0 & \text{for } \left(\mu A1\left(x\right)\cdot\mu A2\left(x\right)\right)\in[0,1)\\ \min(\mu A1\left(x\right)\cdot\mu A2\left(x\right)) & \text{otherwise}\end{array} \tag{3.20}$$

where $\mu Ai(x)$, $i=1, 2$, denotes the degree of membership of the element x in the i-th fuzzy set, Ai. An interesting t-norm is the nilpotent minimum t-norm defined as

$$tNm(\mu A1\left(x\right)\cdot\mu A2\left(x\right))=\begin{array}{ll}0 & \text{for } \mu A1\left(x\right)+\mu A2\left(x\right)\le 1\\ \min(\mu A1\left(x\right)\cdot\mu A2\left(x\right)) & \text{otherwise}\end{array} \tag{3.21}$$

The s-norm or t-conorm functions define a general class of disjunction operators. The following s-norm functions are correspondingly dual to the aforementioned t-norms (3.22)–(3.25):

$$\textbf{Maximum}: sm(\mu A1\left(x\right)\cdot\mu A2\left(x\right))=\max(\mu A1\left(x\right)\cdot\mu A2\left(x\right)) \tag{3.22}$$

$$\textbf{Algebraic sum}: sa(\mu A1\left(x\right)\cdot\mu A2\left(x\right))=\mu A1\left(x\right)+\mu A2\left(x\right)-\mu A1\left(x\right)\cdot\mu A2\left(x\right) \tag{3.23}$$

$$\textbf{Łukasiewicz s-norm}: sL\left(\mu A1\left(x\right)\cdot\mu A2\left(x\right)\right)=\min(1\cdot\mu A1\left(x\right)+\mu A2\left(x\right)) \tag{3.24}$$

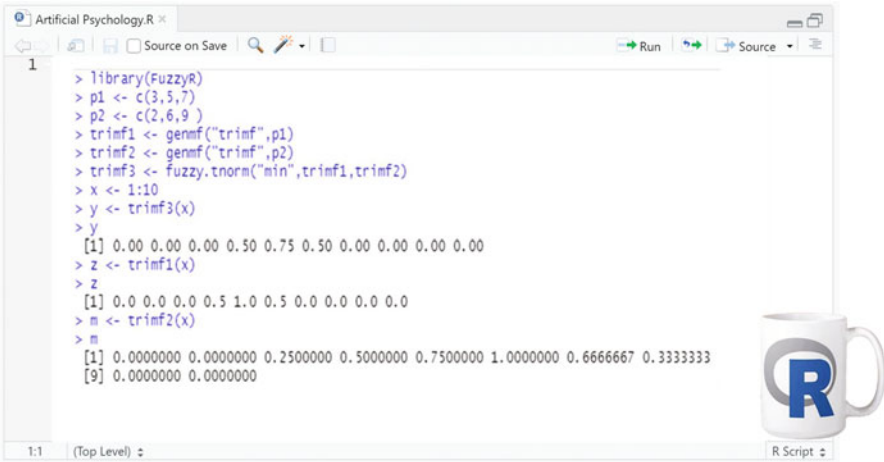
$$\textbf{Drastics-norm}: sd\left(\mu A1\left(x\right)\cdot\mu A2\left(x\right)\right)=1 \text{ for } \left(\mu A1\left(x\right)\cdot\mu A2\left(x\right)\right)\in[0.1)\max(\mu A1\left(x\right)\cdot\mu A2\left(x\right)) \text{ otherwise} \tag{3.25}$$

where $\mu Ai(x)$, $i=1, 2$, denotes the degree of membership of the element x to the fuzzy sets Ai.

3.10.1 Practical Example Using R

We can illustrate these fuzzy set terms with an example. Let us suppose two experts interviewed one MS patient and evaluated him/her as high and very high. If we want to quantify these linguistic terms and combine them based on fuzzy logic, we can use the following codes using the FuzzyR library (Listing 3.4)

Listing 3.4 R codes for linguistic terms and composite them based on fuzzy logic



We can use the code below for defuzzification.

```
> Crisp_value = defuzz(1:10, c(1,3, 5), "centroid")
```

This code defuzzifies scores between 1 and 10 based on a triangular fuzzy number using a "centroid." For aggregating the opinions of the experts in terms of a t-conorm, we write the codes as follows:

```
> p1 <- c(2,5,8)
> p2 <- c(3,6,8)
> trimf1 <- genmf ("trimf",p1)
> trimf2 <- genmf ("trimf",p2)
> trimf3 <- fuzzy.tconorm("max",trimf1,trimf2)
> x <- 1:10
> y <- trimf3(x)
> y
```

```
[1] 0.0000000 0.0000000 0.3333333 0.6666667 1.0000000 1.0000000
0.5000000 0.0000000
[9] 0.0000000 0.0000000
```

We conclude that the most prevalent scores are between 3 and 7.

3.11 Mamdani Fuzzy Inference System

Mamdani systems have proven to be a very useful tool for function approximation and control (Cao et al., 2001) – an aspect that has interesting potential applications in the social sciences too, which we will not discuss here. In this section, we analyze their potential usefulness as a tool for logical deductive inference to study, via simulation, the consequences, and behavior of a model defined by means of IF-THEN rules.

Mamdani systems are most often classified as a form of Approximate Reasoning, which has been defined as "the process or processes by which a possible imprecise conclusion is deduced from a collection of imprecise premises" (Pal & Mandal, 1991). This categorization, together with the fact that the core component of a Mamdani system is a set of IF-THEN rules, can easily mislead one to believe that Mamdani systems can provide the logical implications of the set of rules used to build them, even if only approximately. Without pointing at any particular example, it is not difficult to find cases in the literature that seem to be taking this assumption for granted, either implicitly or explicitly. Specifically, one might be tempted to believe that in a Mamdani system the joint truth of the premises guarantees the truth of the conclusions. In logics that admit degrees of partial truth, this expectation would read that if the inputs of the system are true to some degree (i.e., they satisfy the antecedents of the rules to some extent), then the outputs of the system should also be true to at least the same degree (i.e., they should satisfy the consequents of the rules to at least the same extent). This fallacious interpretation of Mamdani systems as truth-preserving inference machines is certainly – and fortunately – not shared by everyone, but is reasonably widespread and does permeate many simulation applications of the technique.

To be clear, Mamdani systems are not truth-preserving in the sense stated above; they can lead to very different results from those obtained if the IF-THEN rules embedded within are interpreted as proper logical implications. This fact has already been well established in the specialized literature of fuzzy logic – as the quote below shows – but, arguably, it does not seem to be so conspicuous in many practical applications of the technique.

The inference rule used by Mamdani systems "is not a logical inference, i.e., a procedure aiming at the derivation of new facts from some other known ones using formal deduction rules. No logical implication is inside and thus, no modus ponens proceeds" (Klawonna & Novák, 1996).

This section illustrates through several examples why the Mamdani method is not appropriate to explore the logical deductive consequences of a set of IF-THEN implication premises. More technical discussions of some of the aspects that we illustrate in this paper can also be found in the literature (Bodenhofer et al., 2007; Dubois & Prade, 1996; Hájek, 2013; Klawonna & Novák, 1996; Novák, 1994).

Mamdani fuzzy systems were originally designed to imitate the performance of human operators in charge of controlling certain industrial processes (Mamdani, 1974, Mamdani & Assilian, 1975). The aim was to summarize the operator's

experience into a set of (linguistic) IF-THEN rules that could be used by a machine to automatically control the process. Specifically, using such a set of IF-THEN rules.

A Mamdani fuzzy system defines a function f which generates numerical outputs $y = f(x)$ from (usually numerical) input values x. Here, we present a reduced and simplified exposition of the method. For a more complete and detailed presentation, the reader is referred to sections 11.4.1 and 11.4.2 in Zimmermann (2011) or section 11.4 and chapter 12 in Klir and Yuan (1995a, b).

Mamdani systems are composed of IF-THEN rules of the form "IF X is A THEN Y is B," such as "IF PRESSURE is HIGH THEN VOLUME is LOW." The IF part "X is A" is called the antecedent of the rule, and the THEN part "Y is B" is called the consequent of the rule. For simplicity in the exposition of the method and the examples, let us assume that X and Y (PRESSURE and VOLUME, respectively, in the example above) are numerical variables defined on real intervals. The examples we provide can be easily adapted to other input and output spaces, multiple inputs, or fuzzy inputs. Thus, henceforth variable X is assumed to be defined in a real interval that we call the input interval, whilst variable Y is assumed to be defined in a real interval that we call the output interval. Let us use lower-case letters x and y to denote specific values of the variables X and Y, respectively.

The symbols A and B (HIGH and LOW, respectively, in the example above) denote linguistic terms that are modeled as fuzzy sets defined on the input and output intervals, respectively. Fuzzy set A is defined by a membership function $\mu_A$that assigns a real value $\mu_A(x)$ between 0 and 1 to each element x in the input interval. The value $\mu_A(x)$ is called the degree of membership of element x in fuzzy set A and can be interpreted as the extent to which element x belongs to fuzzy set A. If the fuzzy set A represents a certain concept (i.e., "HIGH"), $\mu_A(x)$ can also be interpreted as the truth value of the proposition "X is A" whenever $X = x$ (e.g., the truth value of "PRESSURE is HIGH" whenever PRESSURE $= x$), represented as Truth Value $(X \text{ is } A \mid X = x)$. Likewise, fuzzy set B is defined by a membership function μB that assigns a real value $\mu B(y)$ between 0 and 1 to each real value y in the output interval.

Most often, Mamdani systems are composed of several IF-THEN rules. Naturally, each of the rules (which we index with subscript k) may use different fuzzy sets Ak and Bk. The antecedents and consequents can also be combined propositions that include the logical connectives AND or. A standard Mamdani system uses the following operations to compute the truth value of combined propositions:

$$\begin{gathered}
\text{Truth Value}\,(X \text{ is } C \text{ OR } X \text{ is } D|X=x) = \max\ (\text{Truth Value } X \text{ is } C|X=x), \\
\text{Truth Value}\,(X \text{ is } D|X=x) = \max\ (\mu C\,(x), \mu D(x)) \\
\text{Truth Value}\,(X \text{ is } C \text{ AND } X \text{ is } D|X=x) = \min\ (\text{Truth Value } X \text{ is } C|X=x), \\
\text{Truth Value}\,(X \text{ is } D|X=x) = \min\ (\mu C\,(x), \mu D(x))
\end{gathered} \tag{3.26}$$

The logical negation is implemented in a standard Mamdani system as follows:

$$\text{Truth Value}(X \textit{ is } \text{NOT } A|X=x) = 1 - \text{Truth Value}(X \textit{ is } A|X=x) = 1 - \mu A(x) \tag{3.27}$$

Leaving aside a possible fuzzification step, which is not relevant for our discussion, the algorithm that a Mamdani system uses to compute a numerical output y from a numerical input $X = x$, given a set of rules "IF X is Ak THEN Y is Bk," consists of the following steps:

1. Compute the degrees of consistency between observations (inputs) and antecedents of each rule. In this step, we evaluate the extent to which the antecedent of each IF-THEN rule is satisfied for a given input. The degree of consistency between an input or observation $X = x$ and an antecedent "X is A" is simply the degree of membership of x in the fuzzy set A, i.e., $\mu A(x)$. The result of this step is a number $\mu_{Ak}(x)$ for each rule "IF X is Ak THEN Y is Bk" (i.e., the degree of consistency between the input and each rule's antecedent). If $\mu_{\text{Ak}}(x) > 0$, the corresponding rule k is said to be "fired."
2. Truncate the fuzzy set in the consequent of each rule. The result of this step for each rule "IF X is Ak THEN Y is Bk" is the fuzzy set Bk truncated at the level $\mu_{\text{Ak}}(x)$, i.e., a set $\mu_{\text{output}k|x}$ such that

$$\mu_{\text{output}k|x}(y) = \min(\mu_{\text{Bk}}(y), \mu_{\text{Ak}}(x)) \tag{3.28}$$

3. Aggregate all the truncated fuzzy sets. In this step, the truncated fuzzy sets corresponding to each fired rule are aggregated to provide one single fuzzy set $\mu_{\text{Mamdani}|x}$ defined by the membership function:

$$\mu_{\text{Mamdani}|x}\ (y\ ) = \max\left[\mu_{\text{output}k|x}(y)\right] = \max[\min \mu_{\text{Bk}}(y), \mu_{\text{Ak}}(x)] \tag{3.29}$$

   The equation above clearly shows why Mamdani fuzzy systems are sometimes called max–min fuzzy systems.
4. Defuzzify the aggregated fuzzy set. The defuzzification step transforms the aggregated fuzzy set $\mu_{\text{Mamdani}|x}$ into one single crisp number. Standard Mamdani systems use the Centre of Gravity (COG) defuzzification method. This method returns the projection (on the horizontal axis) of the center of gravity of the area under the membership function $\mu_{\text{Mamdani}|x}$. If some input value is such that no rule is fired, the center of gravity for $\mu_{\text{Mamdani}|x}\ (y)$ cannot be calculated. In that case, some default output value can be considered, or the system can be readjusted to avoid that situation (e.g., by modifying the fuzzy sets Ak or by including new rules).

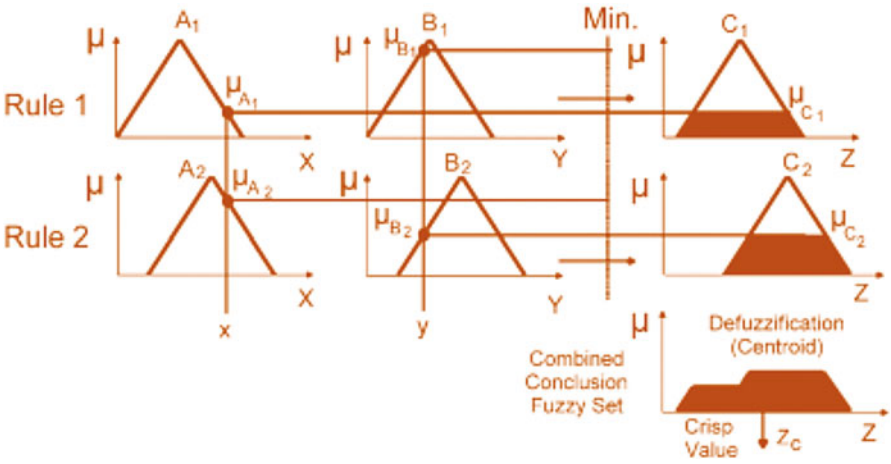


Fig. 3.9 Mamdani fuzzy inference system

Given that the defuzzification step has a large influence on the final function that the system provides, we will also consider here two other alternative defuzzification methods (Van Leekwijck & Kerre, 1999):

I. First of Maxima. This method returns the smallest value of y for which the membership function $\mu_{\text{Mamdani}|x}$ attains its maximum value.
II. Last of Maxima. This method returns the greatest value of y for which the membership function $\mu_{\text{Mamdani}|x}$ (y) attains its maximum value (Izquierdo & Izquierdo, 2017) (Fig. 3.9).

3.11.1 Mamdani Fuzzy System Steps

A Mamdani fuzzy system consists of different steps, which are explained below.

Step 1 Determining the case variables based on the research background and theoretical foundations of the researcher.

Step 2 Determining the fuzzy sets, through interviews with experts, theoretical logic, or the background of the conducted research. Fuzzy sets such as high, medium, low, or very high, high, medium, low, very low, or any fuzzy set that is suitable for the desired variables according to the researcher's opinion are determined. Fuzzy sets are linguistic terms.

Step 3 Determining the fuzzy membership functions, according to the previous step, the fuzzy membership function is determined for each fuzzy set.

In this field, you can get help from experts again; fuzzy membership functions are different functions that have been discussed in the previous sections. Triangular, trapezoidal, and Gaussian functions are among these functions. Each function is specified with parameters, which parameters are determined based on the range of variable measurement scores according to the researcher's opinion.

Step 4 Define IF-Then rules. These are the rules of the antecedent relationship and they specify the consequent. There are two solutions to determine the rules; one solution is based on the opinion of experts who actually formulate rules based on their experience in their work, for example, a health psychologist derives this rule based on his studies and professional experience. "If depression is high, stress sensitivity is average and life expectancy is average, then the quality of life is close to average." The linguistic terms "high," "average," and "close to average" are the same fuzzy sets; based on their type, the fuzzy membership functions are determined. The second solution is to explain the rules based on the data set in an exploratory manner. In this situation, Adaptive Neuro Fuzzy Inference System (ANFIS) or other algorithms such as Genetic Cooperative Classification Learning (GCCL) can be used. This issue is explained in detail below.

Step 5 Weighting the fuzzy rules, the researcher may be interested in assigning weights to the fuzzy rules. Weights indicate the degree of importance of those rules. It is standard to give all the rules the same weight, i.e., 1. According to the importance of each rule in the fuzzy system, different weights, however, can be determined and attributed to those rules. Weights can be obtained in different ways. One method is given in the example related to the fuzzy cognitive map (see FCM example section).

Step 6 In this step, all the truncated fuzzy sets obtained in the previous steps are aggregated. In fact, in this step, the shortened fuzzy sets related to the rules are aggregated with each other to provide a single fuzzy set. For this purpose, various methods are used. Max and Min operators are routinely used; that is why, Mamdani-type fuzzy inference systems are sometimes called max–min fuzzy systems.

Step 7 The last step is defuzzification of the aggregated fuzzy set. In this step, defuzzification, the aggregated fuzzy set is converted into a single crisp number one. There are different methods for defuzzification; the Center of Gravity (COG) method is routinely used. Different defuzzification methods affect the final function that the system provides. Therefore, the opinion of the researcher is important and he must decide on this matter.

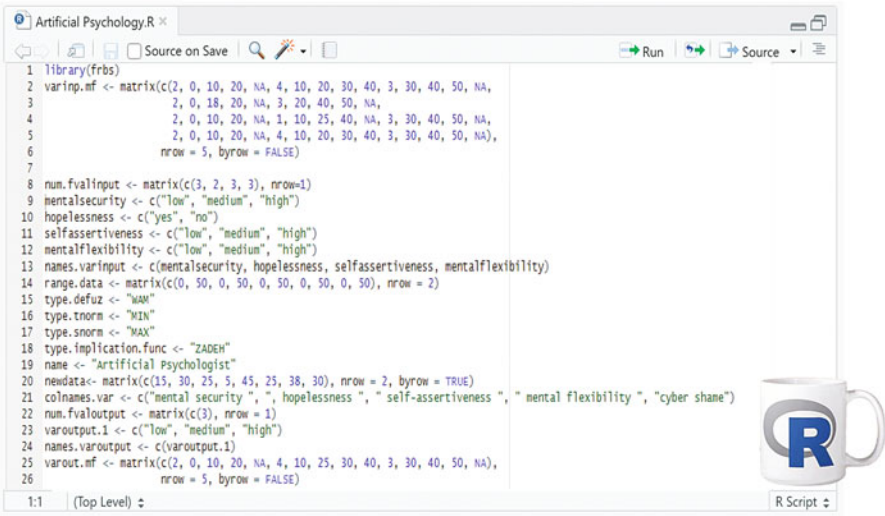
Table 3.2 The variable, fuzzy set, membership functions, and parameters to identify cyber shame

Variable	Fuzzy set	Membership functions	Parameter
Mental security (MS)	Low	Triangular	(0, 10, 20)
	Medium	Trapezoidal	(10, 20, 30, 40)
	High	Triangular	(30, 40, 50)
Hopelessness (HL)	Yes	Triangular	(0, 18, 20)
	No	Triangular	(20, 40, 50)
Self-assertiveness (SA)	Low	Triangular	(0, 10, 20)
	Medium	Triangular	(10, 25, 40)
	High	Triangular	(30, 40, 50)
Cyber shame (CS)	Low	Triangular	(0, 10, 20)
	Medium	Trapezoidal	(10, 25, 30, 40)
	High	Triangular	(30, 40, 50)

3.11.2 Practical Example Using R

Example 3.1 The artificial psychologist attempts to identify cyber shame based on mental security and hopelessness, self-assertiveness, and mental flexibility using the Mamdani-type fuzzy inference system (FIS). The psychologist first forms the parameter matrix of the fuzzy membership function based on expert opinion. The variable, fuzzy set, membership functions, and parameters are demonstrated in Table 3.2 (Listings 3.5 and 3.6).

Listing 3.5 R codes for the Mamdani fuzzy inference system



Listing 3.6 R codes for the Mamdani fuzzy inference system (continued)

```
type.model <- "MAMDANI"
rule <- matrix(
  c("low", "and", "yes", "and", "low", "and", "low", "->", "high",
    "medium", "and", "no", "and", "low", "and", "low", "->", "medium",
    "high", "and", "no", "and", "medium", "and", "high", "->", "low"),
  nrow = 3, byrow = TRUE)
object <- frbs.gen(range.data, num.fvalinput, names.varinput,
                   num.fvaloutput, varout.mf, names.varoutput, rule,
                   varinp.mf, type.model, type.defuz, type.tnorm,
                   type.snorm, func.tsk = NULL, colnames.var, type.implication.func, name)
plotMF(object)
summary(object)
res <- predict(object, newdata)$predicted.val
```

Figure 3.10 illustrates the rules between the variables in the R output. For instance, the first rule is as follows:

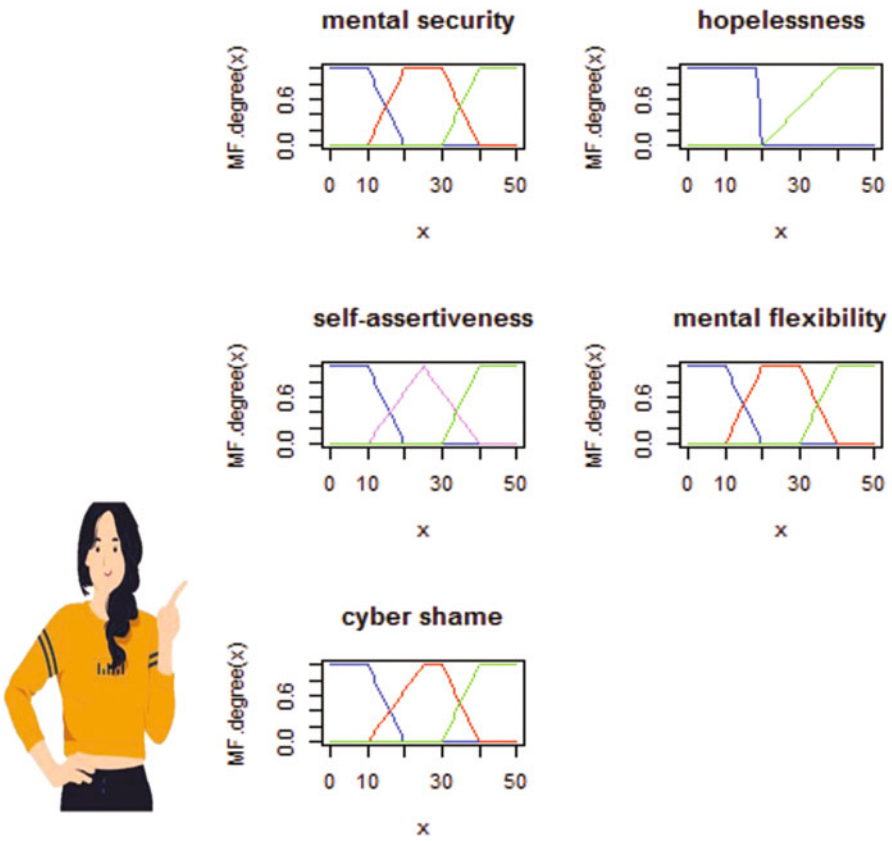


Fig. 3.10 R plot for the Mamdani fuzzy inference system

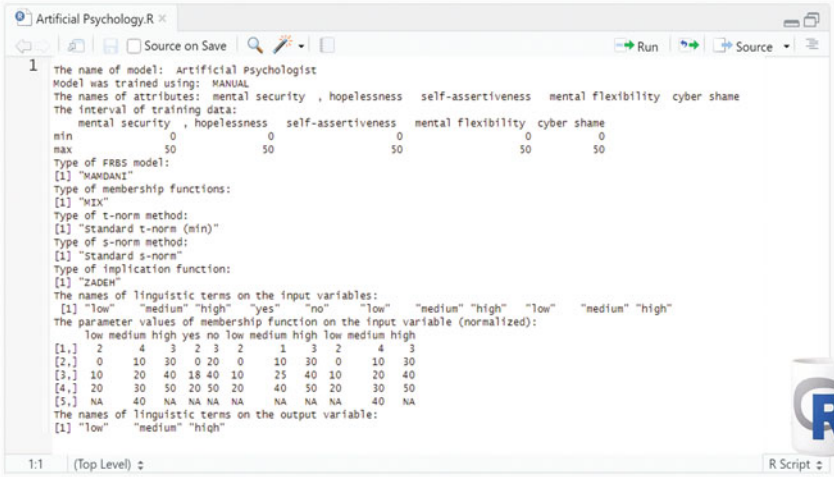


Fig. 3.11 R results of the Mamdani fuzzy inference system

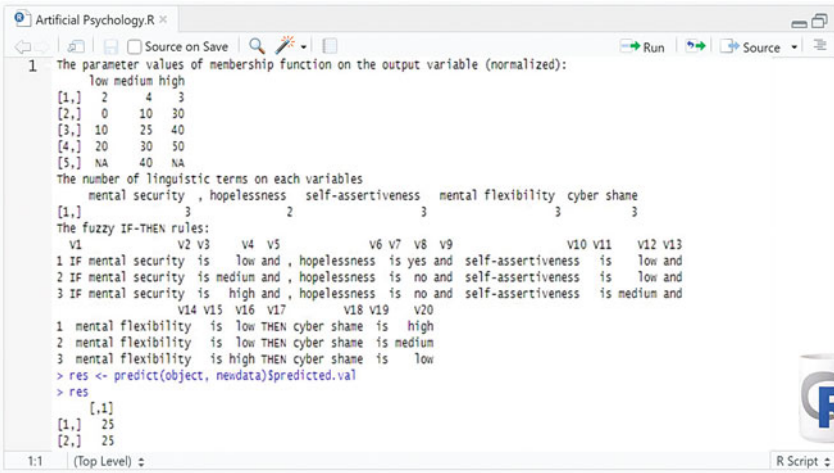


Fig. 3.12 R results of the Mamdani fuzzy inference system (continued)

"If mental security is low and hopelessness is yes and self-assertiveness is low and mental flexibility is low, then cyber shame is high."

The Figures 3.11 and 3.12 shows the form of the membership functions.

Example 3.2 An Artificial psychologist aims to model pain feeling (PF) based on childhood trauma (CT) experience and alexithymia (ALX) using the fuzzy inference system (FIS) (Fig. 3.13).

After a careful review of the literature, different variables are evaluated and CT and ALX are selected. The artificial psychologist interviewed PF experts and added other variables. In fact, a careful literature review and interviews with experts and the

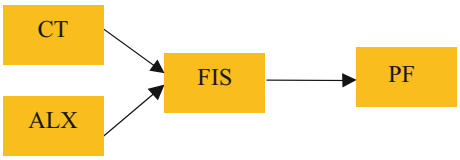


Fig. 3.13 Hypothetical model

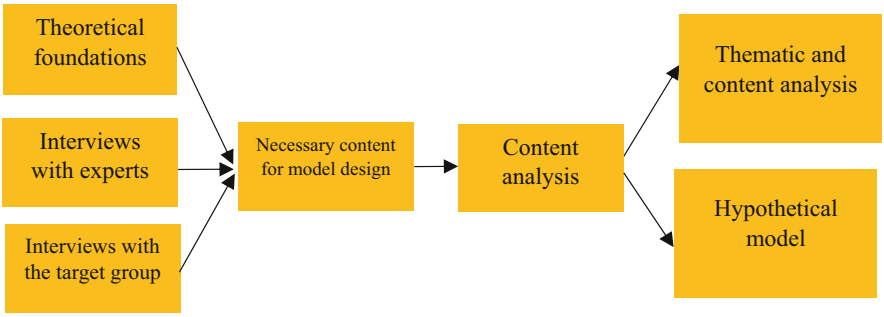


Fig. 3.14 Flow work for developing a hypothetical model

target population (people with physical pain) are the three primary sources for designing and developing the hypothetical model (Fig. 3.14).

Here, the researcher can use sample data from the target or expert groups to obtain the values of variables. In this study, artificial psychologists prepare the CT and ALX questionnaires with sufficient validity and reliability.

PF is measured within 0–30. Assuming that the range of the CT and ALX scales is within 0–10 and PF is within 0–30, the range for the first two cases is determined by the scores, and skeletal PF is determined by the opinions of relevant researchers. Having determined these ranges, an artificial psychologist determines fuzzy sets based on researcher and expert opinion (Table 3.3).

Artificial psychologists then determine the suitable fuzzy membership function for the fuzzy set. Selecting and using one of the numerous fuzzy membership functions requires practical and theoretical insight. The triangular fuzzy membership function is typically appropriate for psychology studies employing 1-, 3-, and 5-point Likert scales. As mentioned earlier, the triangular function has three parameters, namely a, b, c; and the trapezoidal function is also used due to its similarity to the triangular function, but instead of only one value from the x axis having a membership degree of 1, a range of numbers has the membership function of 1, giving the trapezoidal function four parameters, namely a, b, c, and d. Table 3.3 shows the variables, range of scores, fuzzy set, and membership function parameters.

Figures 3.16, 3.17, and 3.18 demonstrate the membership function after running the R codes. After fuzzy sets, membership functions, and their parameters, the next step is the fuzzy rule-based inference system (FRBIS). This example uses a Mamdani-type fuzzy inference system. Since the number of rules could be large,

Table 3.3 Variables, range of scores, fuzzy set, and membership function parameters

Variable	Range of scores	Fuzzy sets	Membership function	Parameters
CT	0–10	Low/middle/high	Triangular/triangular/triangular	(0,1,2)
				(2,3,7)
				(4,6,10)
ALX	0–10	Mild/moderate/ severe	Trapezoidal/trapezoidal/ trapezoidal/	(0,1,2,2)
				(2,3,3,4)
				(5,8,10,11)
Skeletal PF		Little/tolerate/much	Triangular/triangular/triangular/	(0,3,8)
				(8,13,18)
				(18,23,30)

Table 3.4 Fuzzy inference rules

Rule	Average importance	Weight
1	4	0.75
2	5	1
3	2.5	0.4

Rule 1. If CT is low and ALX is mild, then PF is little (0.75)
Rule 2. If CT is middle, then PF is tolerable (1)
Rule 3. If CT is high or ALX is moderate, then PF is much (0.4)

Table 3.5 Survey of experts to determine the relative importance of each rule

No.	Rules	Importance				
		Very low (1)	Low (2)	Moderate (3)	High (4)	Very high (5)

the fuzzy inference system can be implemented in several stages, which is also known as the hierarchical fuzzy model.

In this example, an artificial psychologist designs the rules using their own logic and assembling a panel of experts. These rules could be considered composite hypotheses and are presented in the following example (Table 3.4). Table 3.4 indicates the rules of the Mamdani-type fuzzy inference system.

Importance weights can also be determined for rules. With a group of five experts on a 5-point Likert scale ((very low: 1), (low: 2), (average: 3), (high: 4), (very high: 5)), the artificial psychologist aiming to determine the importance of each rule takes their average (Table 3.5) and normalizes them into a standard weight.

There are different ways to normalize the resulting averages, and the following equation is a simple solution:

$$x = \frac{X - \min}{\max - \min}$$

The expert survey to determine the importance and ultimately the weight of each rule is as follows:

After determining the rules, the artificial psychologist is required to specify the defuzzification method of the study. As mentioned earlier, there are various defuzzification methods. Here, the artificial psychologist employs the centroid method to convert the obtained fuzzy number into a crisp number. Listing 3.6 depicts the R codes of all stages.

Figure 3.15 shows the output, including the fuzzy system's specifications. To evaluate the model, suppose a person who has experienced skeletal pain responds to emotional dyslexia and childhood experience scales with scores of 4 and 5, respectively. According to the Mamdani-type fuzzy inference model, the muscle pain score will be 12.997. For a depressed individual with scores of 5 and 7 in ALX and CT, muscle pain will be 16.979 (Figs. 3.16 and 3.17).

The surface plot suggests that higher CT and ALX increase the probability of skeletal PF. This graph shows that the rules are designed properly (Listing 3.7).

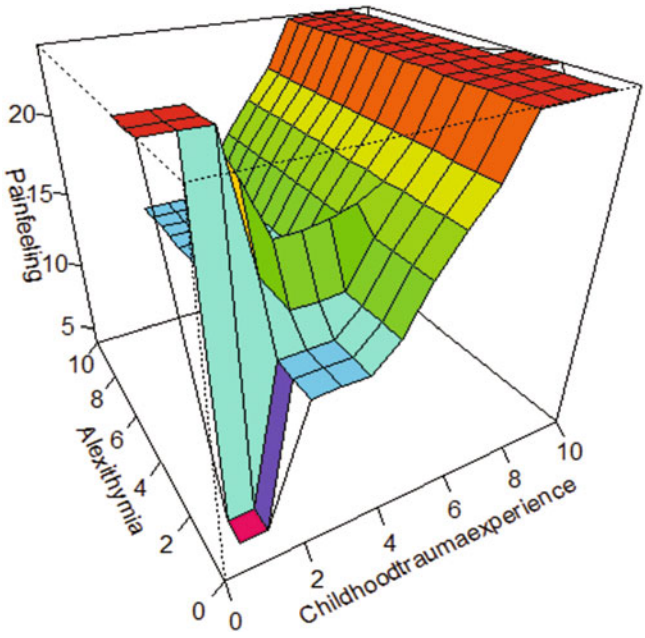


Fig. 3.15 R surface plot for the Mamdani fuzzy inference system

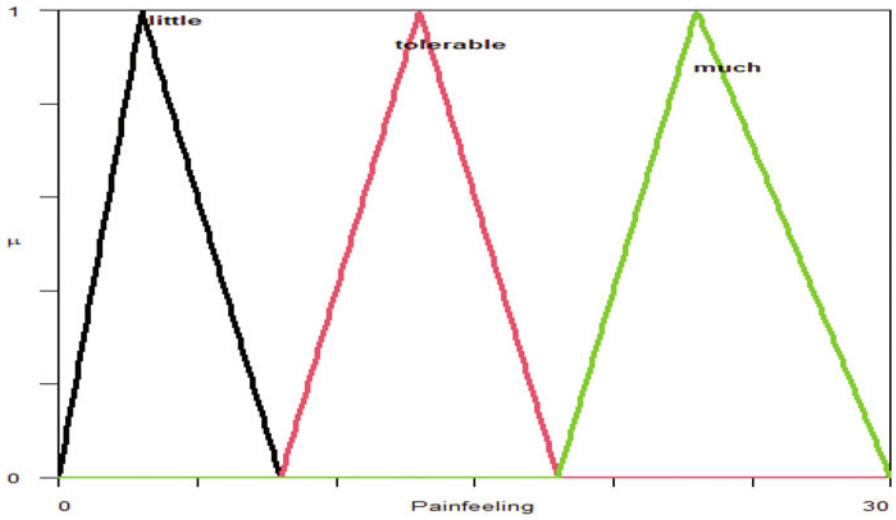


Fig. 3.16 Triangular fuzzy membership of pain feeling

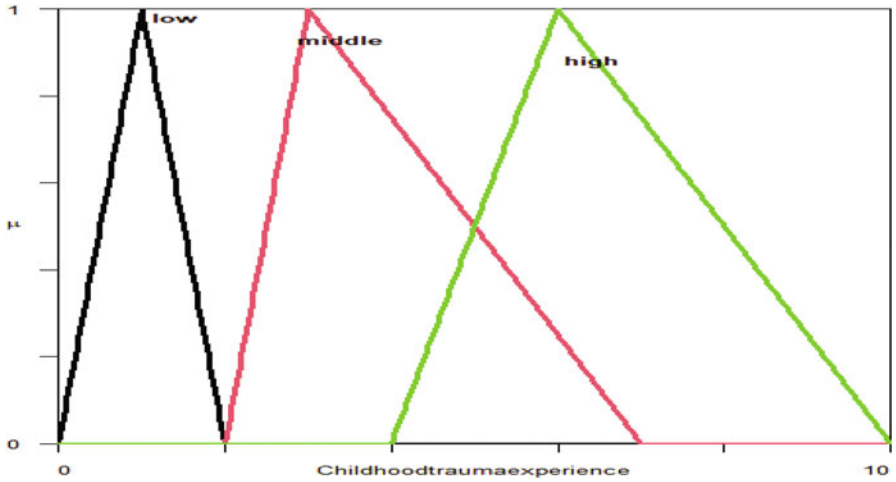


Fig. 3.17 Triangular fuzzy membership of childhood trauma experience

Listing 3.7 R codes for the Mamdani fuzzy inference system (continued)

```
library(FuzzyNumbers)
library(sets)
library(FuzzyR)
fis= newfis('Artificial Psychologist')
fis= addvar(fis, 'input', 'Childhoodtraumaexperience', c(0,10))
fis= addvar(fis, 'input', 'Alexithymia', c(0,10))
fis= addvar(fis, 'output', 'Painfeeling', c(0,30))
fis= addmf(fis, 'input', 1, 'low', 'trimf', c(0,1,2))
fis= addmf(fis, 'input', 1, 'middle', 'trimf', c(2,3,7))
fis= addmf(fis, 'input', 1, 'high', 'trimf', c(4,6,10))
fis= addmf(fis, 'input', 2, 'mild', 'trapmf', c(0,1,2,2))
fis= addmf(fis, 'input', 2, 'moderate','trapmf', c(2,3,3,4))
fis= addmf(fis, 'input', 2, 'severe', 'trapmf', c(5,8,10,11))
fis= addmf(fis, 'output', 1, 'little', 'trimf', c(0,3,8))
fis= addmf(fis, 'output', 1, 'tolerable', 'trimf', c(8,13,18))
fis= addmf(fis, 'output', 1, 'much', 'trimf', c(18,23,30))
rules= rbind(c(1,1,1,0.75,1), c(2,0,2,1,2), c(3,2,3,0.4,2))
fis= addrule(fis, rules)
showfis(fis)
showrule(fis)
evalfis(c(4,5), fis)
evalfis(rbind(c(4,5), c(5,7)), fis)
gensurf(fis)
plotmf(fis, "output", 1)
plotmf(fis, "input", 2)
plotmf(fis, "input", 1)
```

```
> evalfis(rbind(c(4,5), c(5,7)), fis)
         [,1]
[1,] 12.99757
[2,] 16.97906
```

```
1. If (Childhoodtraumaexperience is low) and (Alexithymia is mild) then (Painfeeling is little) (0.75)
2. If (Childhoodtraumaexperience is middle) then (Painfeeling is tolerable) (1)
3. If (Childhoodtraumaexperience is high) or (Alexithymia is moderate) then (Painfeeling is much) (0.4)
```

3.12 Toward Fuzzy Rule Mining

Basically, there are two different methods to construct fuzzy rule-based systems (FRBSs), including classification (FRBCSs) and regression systems (FRBRSs), depending on the information. One method or strategy is to obtain information from human experts in the field. In this strategy, knowledge is defined by artificial psychologists who interview experts in the field to extract and represent their knowledge. Although this method is most commonly used, sometimes it is not feasible because of the lack of knowledge. The second strategy is to apply learning methods for extracting knowledge from data in FRBSs. Some of the strategies are used for FRBRSs, and some others are used in FRBCSs. In this section, we try to discuss some of both (Figs. 3.18 and 3.19).

A fuzzy rule-based classification system (FRBCS) or fuzzy rule-based regression system consists of two main conceptual elements: (a) the fuzzy rule base (FRB) and (b) the fuzzy reasoning method (FRM). The first element provides an association

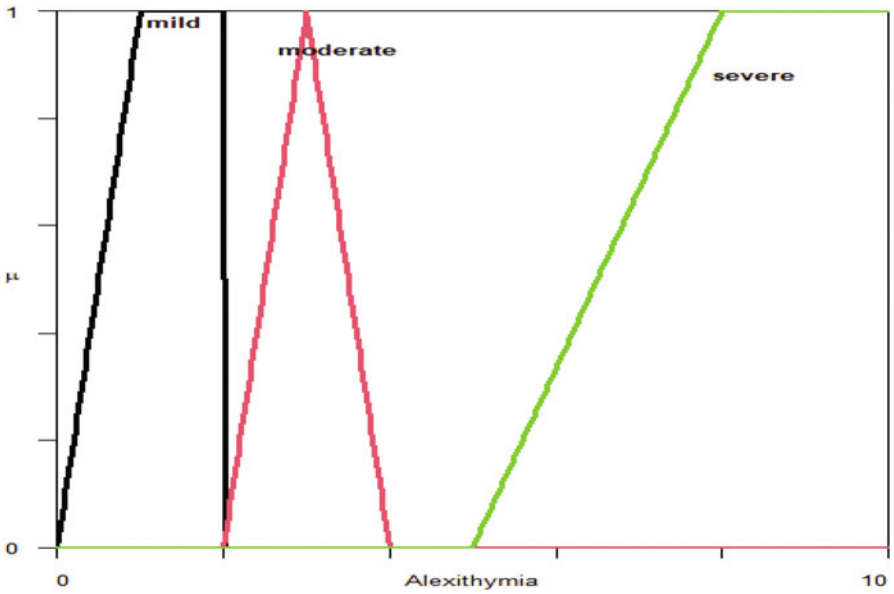


Fig. 3.18 Triangular fuzzy membership of alexithymia

```
> evalfis(rbind(c(4,5), c(5,7)), fis)
         [,1]
[1,] 12.99757
[2,] 16.97906

1. If (Childhoodtraumaexperience is low) and (Alexithymia is mild) then (Painfeeling is little) (0.75)
2. If (Childhoodtraumaexperience is middle) then (Painfeeling is tolerable) (1)
3. If (Childhoodtraumaexperience is high) or (Alexithymia is moderate) then (Painfeeling is much) (0.4)
```

```
1.  Name            Artificial Psychologist
2a. Type            mamdani
2b. mfType          t1
3.  Inputs/Outputs  [ 2 1 ]
4.  NumInputMFs     [ 3 3 ]
5.  NumOutputMFs    [ 3 ]
6.  NumRules        3
7.  AndMethod       min
8.  OrMethod        max
9.  ImpMethod       min
10. AggMethod       max
11. DefuzzMethod    centroid
```

Fig. 3.19 R outputs of the Mamdani fuzzy inference system

between the space of pattern features and the space of consequent classes or target. The second element provides a mechanism to classify or predict a given pattern or values based on the first part (Jiao et al., 2015).

In classification, the outputs are categorized. As a result, in this model the antecedent part is linguistic variables, and the consequent part is a class *Cj* from a prespecified class set $C = \{C1, \ldots, CM\}$, in regression task the target outputs are quantitative as values (Riza et al., 2015).

3.12.1 Adaptive Network-Based Fuzzy Inference System (ANFIS)

Neuro-fuzzy inference systems were developed in 1993 by J.S. Roger Jang. This method can be considered as an exploratory fuzzy inference system if we can name Mamdani and Takagi–Sugeno–Kang or Sugeno-type (TSK) fuzzy inference systems as a confirmatory one. This system will be discussed below. An adaptive network-based fuzzy inference system (ANFIS) consists of a Takagi–Sugeno–Kang or TSK–FRBS model built out of a five-layered network architecture. Both artificial neural network and fuzzy logic are used in ANFIS' architecture (Avcı, 2008; Avcı & Akpolat, 2006; Avcı et al., 2005). The "ANFIS" is a learning algorithm that consists of two forward and backward processes. The forward stage includes the layers as follows (Fig. 3.20).

Layer 1 The fuzzification process in which crisp values are transformed into linguistic values using the Gaussian function (or the other fuzzy membership functions) as the shape of the membership function.
Layer 2 The inference stage using the t-norm operator (the AND operator).
Layer 3 Calculating the ratio of the strengths of the rules.
Layer 4 Calculating the parameters for the consequent parts.
Layer 5 Computing the overall output as the sum of all incoming inputs.

Layer 1 receives the inputs and transforms them into the fuzzy value using membership functions. Layer 2 multiplies the fuzzy signals obtained from Layer 1 and provides the firing strength of the rule. Layer 3 is the rules layer where all outputs from Layer 2 are normalized. Layer 4 provides the inference of rules, and all signals are transformed to crisp values. The final layer summarizes all the signals and provides the outputted crisp value (Cvetković et al., 2020).

The backward stage is a process in which the database is estimated. This database includes the parameters of the membership functions in the antecedent part and the coefficients of the linear equations in the consequent part (the output is not a value

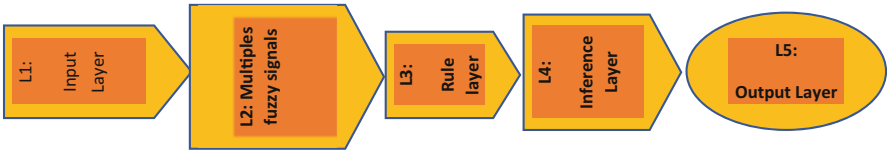


Fig. 3.20 ANFIS layer

like the Mamdani Inference system). Since the Gaussian function is a membership function in this method, therefore, it is expected that mean and variance are optimized as two parameters of this function. In this step, the least squares method is used to perform the parameter learning. For the prediction phase, the method performs normal fuzzy reasoning of the TSK model (Lala Septem Riza et al., 2015).

The Takagi–Sugeno fuzzy model is a type 3 fuzzy inference system, where the rule outputs are a linear combination of inputs along with a constant like a regression, and the final output is the weighted average of every rule's output. For a first order of TSK fuzzy model, a typical rule set with base fuzzy if-then rules can be expressed as if x is $A1$ and y is $B1$, then $f1 = p1x + q1y + c1$ (Avcı, 2008; Avcı et al., 2005, 2007).

IF-THEN rules for a two-input Takagi–Sugeno (TSK) system are described as follows:

$$\text{Rule 1: IF } x \text{ is } A_1, y \text{ is } B_1, \text{THEN } f_1 = p_1 x + q_1 y + c_1.$$

$$\text{Rule 2: IF } x \text{ is } A_2, y \text{ is } B_2, \text{THEN } f_2 = p_2 x + q_2 y + c_2$$

$$\text{Rule 3: IF } x \text{ is } A_3, y \text{ is } B_3, \text{THEN } f_3 = p_3 x + q_3 y + c_3$$

where x, y are the inputs in the crisp values set, A_i, B_i are the linguistic terms, p_i, q_i are the consequent parameters, and f_1, f_2, f_3 are the linear combination of inputs along with a constant (c_i).

In summary, ANFIS consists of IF-THEN rules and pairs of input–output and learning algorithms from a neural network (Avcı, 2008; Avcı et al., 2005, 2007; Jang, 1993; Turkoglu & Avcı, 2008). During the forward stage, when the inputs are provided to the model, the consequent parameters are updated, and the initial parameters are kept fixed; using least squares estimation, the consequent parameters are updated in Layer 4, and the final output is calculated accordingly. The backward stage starts immediately after calculating the final output. In this stage, the error is propagated back to Layer 1, and the initial parameters are updated. The consequent parameters are kept fixed (Chopra et al., 2021).

A standard ANFIS has some assumptions. The system is zero-order or a first order Sugeno-type inference system. Membership functions of output are the same, whether they be constant or linear. Each rule is of a specific membership function for each output variable, and all rule weights are 1 (Cvetković et al., 2020) (Fig. 3.21).

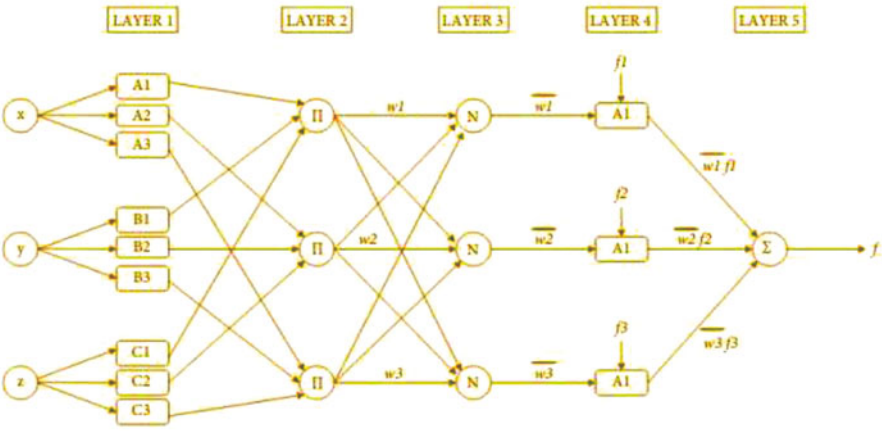


Fig. 3.21 Standard structure of ANFIS. w_1, w_2, and w_3 are the weights of the neurons, and w_1, w_2, and w_3 are the normalized weights of the neuron. (Chopra et al., 2021)

As mentioned, ANFIS is based on a TSK system. In this section, we further discuss the Sugeno method. A Sugeno-type method (or Takagi–Sugeno–Kang) consists of fuzzy inputs and a crisp output (linear combination of the inputs). It is computationally efficient and suitable for optimization and adaptive techniques (Sugeno & Kang 1988). The Sugeno method provides fuzzy rules from a given input–output database. It changes the consequent (then part) of the Mamdani rule with a function (Equation) of the input variables. The T-S style fuzzy rule is: IF x is A AND y is B, THEN z is $f(x, y)$, where x, y, and z are linguistic variables, A and B are fuzzy sets on a universe of discourses X and Y, and $f(x, y)$ is a mathematical function (Du & Zhang, 2008). Sugeno-type FIS uses a weighted average to compute the crisp output, while Mamdani-type FIS uses the technique of defuzzification of a fuzzy output. The first two parts of the fuzzy inference process, fuzzifying the inputs and applying the fuzzy operator, are the same (Du & Zhang, 2008). The main difference is that the Sugeno output membership functions are either linear or constant (Du & Zhang, 2008). In Fig. 3.22, different types of fuzzy systems are shown. Type two is Mamdani FIS with output function based on overall fuzzy output, while type three is the Takagi–Sugeno fuzzy inference.

Practical Example Using R

An artificial social psychologist aims to examine anxiety based on conscientiousness and extroversion with fuzzy modeling. There are two goals: (1) to measure the predictive power of the fuzzy model and (2) to extract the fuzzy rules. They wonder about the possible fuzzy rules between anxiety, conscientiousness, and extroversion. In a heuristic fuzzy model, the psychologist actually seeks to extract the rules based on the data from a large sample of high school students. To this end, the Beck

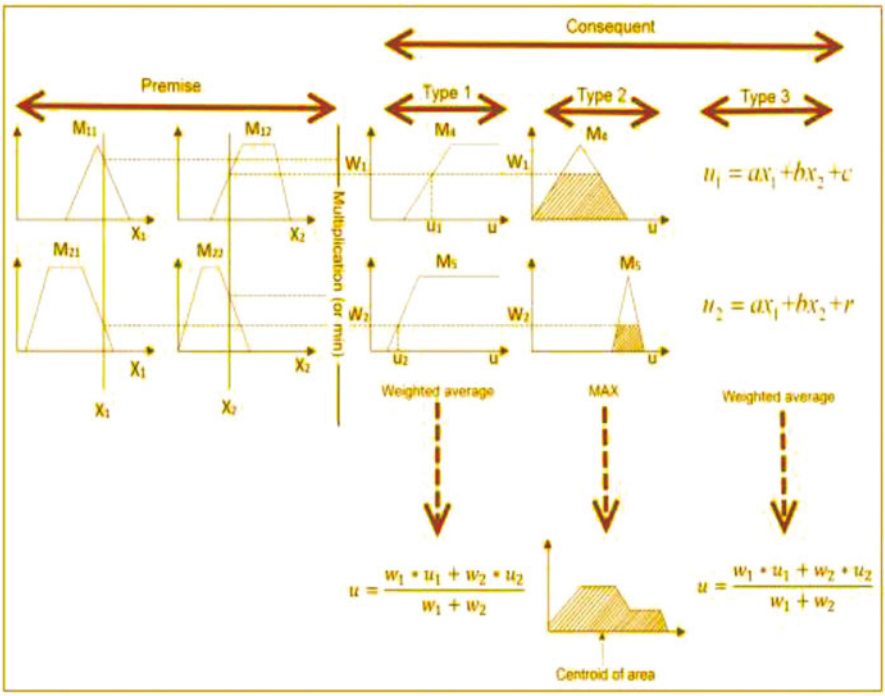


Fig. 3.22 Different types of fuzzy inference system. (Mehran, 2008)

Anxiety Inventory and NEO_AC are calculated for a sample of 861 students (ANFF file), and the model's accuracy is measured using ANFIS while also extracting the fuzzy rules of the three variables.

Listing 3.7 depicts the R codes for ANFIS implementation. In this study, the artificial social psychologist trains the model with 500 people and 361 people as the test sample. The fuzzy set has 5 degrees (very small, small, medium, large, and very large) in 5 iterations, the fuzzy membership is Gaussian, and the implication function is ZADGH.

According to Fig. 3.23, the membership function reveals the two variables of conscientiousness and extroversion.

Figure 3.24 illustrates the evaluation of the trained model versus the test data, and the test model with real data covering a sample of 361 people. Since the MSE, RMSE, and SMAPE values of 202.4, 14.22, and 1.91, respectively, may not satisfy the researcher, the number of fuzzy sets and the sample size can be increased, or even the fuzzy membership function can be altered. Figure 3.25 indicates R outputs of ANFIS implementation. Figure 3.26 demonstrates the extracted fuzzy rules (Listings 3.8 and 3.9).

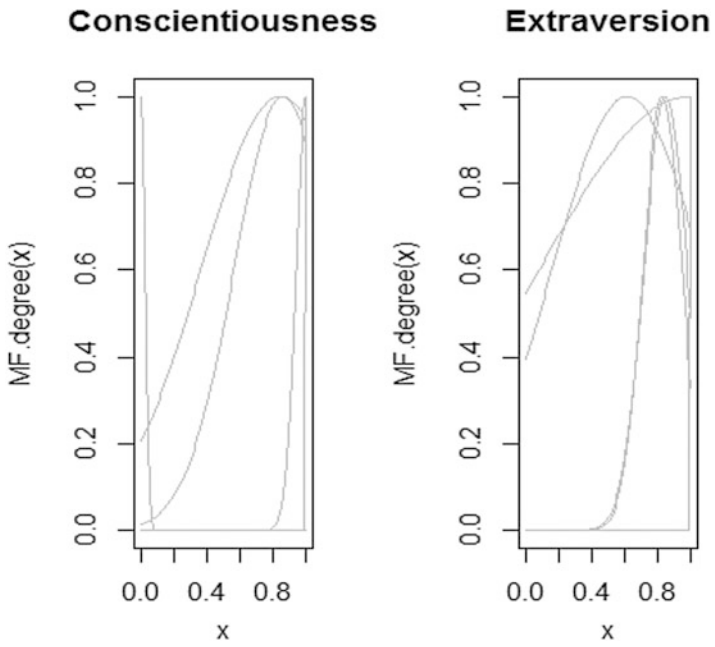


Fig. 3.23 The membership functions of conscientiousness and extroversion

Fig. 3.24 The result of the evaluation of the trained model

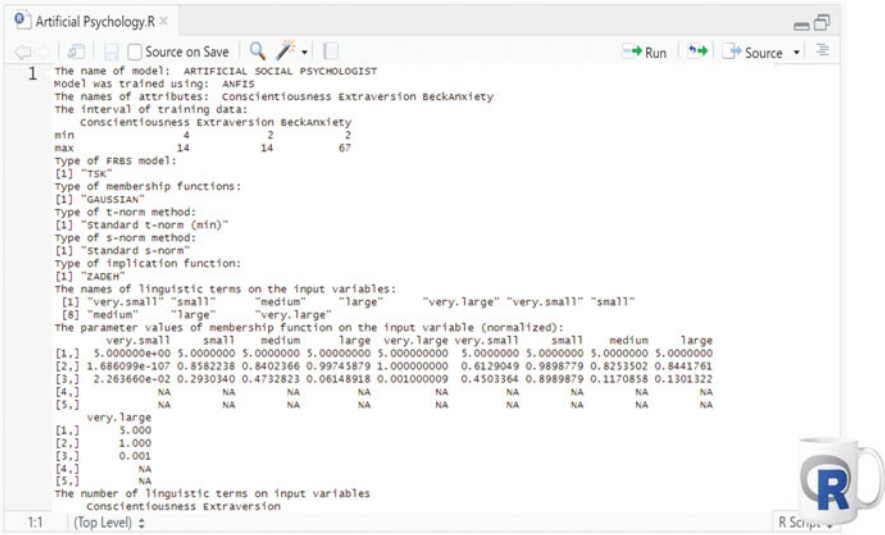


Fig. 3.25 outputs of ANFIS implementation

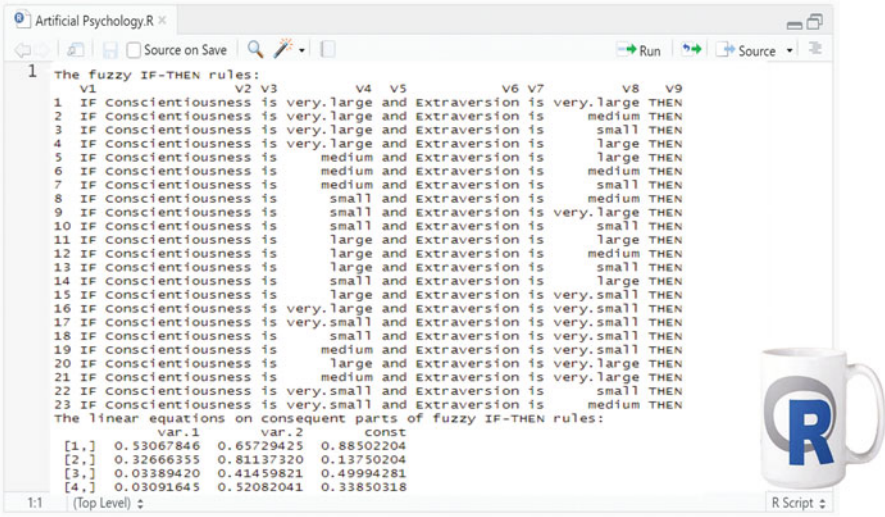
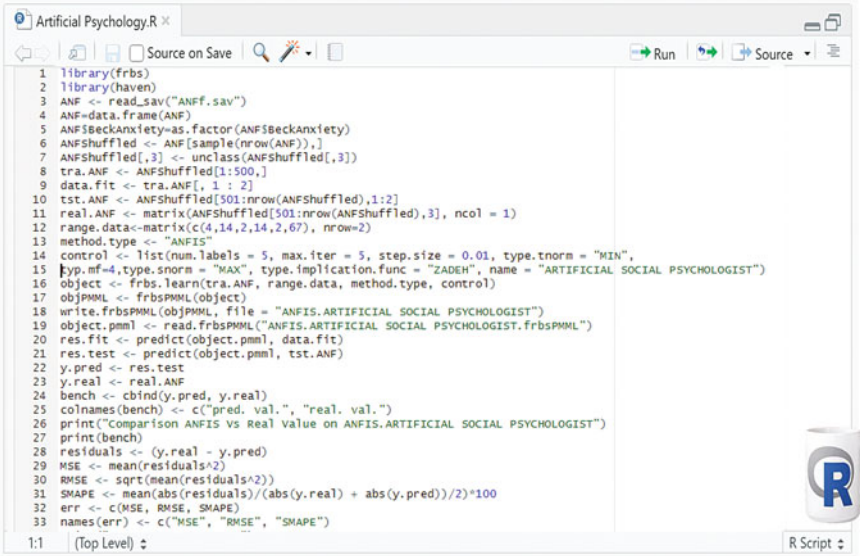


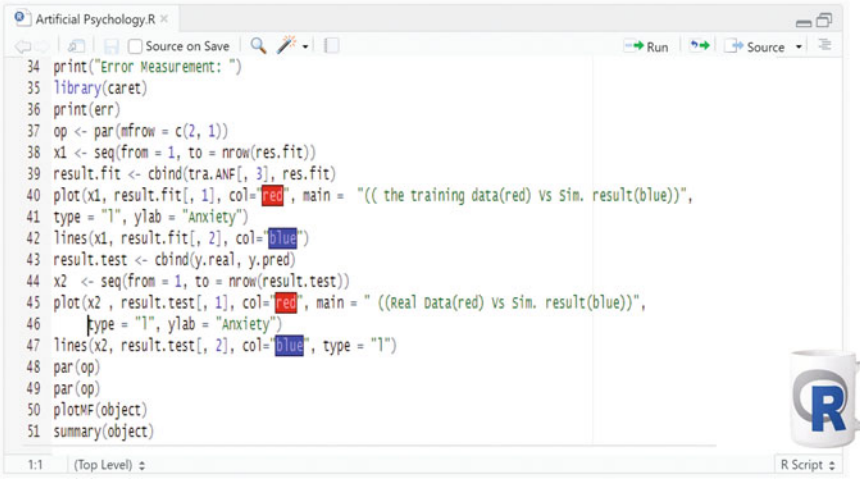
Fig. 3.26 The extracted fuzzy rules

Listing 3.8 R codes for ANFIS implementation



```
library(frbs)
library(haven)
ANF <- read_sav("ANFf.sav")
ANF=data.frame(ANF)
ANF$BeckAnxiety=as.factor(ANF$BeckAnxiety)
ANFShuffled <- ANF[sample(nrow(ANF)),]
ANFShuffled[,3] <- unclass(ANFShuffled[,3])
tra.ANF <- ANFShuffled[1:500,]
data.fit <- tra.ANF[, 1 : 2]
tst.ANF <- ANFShuffled[501:nrow(ANFShuffled),1:2]
real.ANF <- matrix(ANFShuffled[501:nrow(ANFShuffled),3], ncol = 1)
range.data<-matrix(c(4,14,2,14,2,67), nrow=2)
method.type <- "ANFIS"
control <- list(num.labels = 5, max.iter = 5, step.size = 0.01, type.tnorm = "MIN",
type.mf=4,type.snorm = "MAX", type.implication.func = "ZADEH", name = "ARTIFICIAL SOCIAL PSYCHOLOGIST")
object <- frbs.learn(tra.ANF, range.data, method.type, control)
objPMML <- frbsPMML(object)
write.frbsPMML(objPMML, file = "ANFIS.ARTIFICIAL SOCIAL PSYCHOLOGIST")
object.pmml <- read.frbsPMML("ANFIS.ARTIFICIAL SOCIAL PSYCHOLOGIST.frbsPMML")
res.fit <- predict(object.pmml, data.fit)
res.test <- predict(object.pmml, tst.ANF)
y.pred <- res.test
y.real <- real.ANF
bench <- cbind(y.pred, y.real)
colnames(bench) <- c("pred. val.", "real. val.")
print("Comparison ANFIS Vs Real Value on ANFIS.ARTIFICIAL SOCIAL PSYCHOLOGIST")
print(bench)
residuals <- (y.real - y.pred)
MSE <- mean(residuals^2)
RMSE <- sqrt(mean(residuals^2))
SMAPE <- mean(abs(residuals)/(abs(y.real) + abs(y.pred))/2)*100
err <- c(MSE, RMSE, SMAPE)
names(err) <- c("MSE", "RMSE", "SMAPE")
```

Listing 3.9 R codes for ANFIS implementation (continued)



```
print("Error Measurement: ")
library(caret)
print(err)
op <- par(mfrow = c(2, 1))
x1 <- seq(from = 1, to = nrow(res.fit))
result.fit <- cbind(tra.ANF[, 3], res.fit)
plot(x1, result.fit[, 1], col="red", main =  "(( the training data(red) Vs Sim. result(blue))",
type = "l", ylab = "Anxiety")
lines(x1, result.fit[, 2], col="blue")
result.test <- cbind(y.real, y.pred)
x2  <- seq(from = 1, to = nrow(result.test))
plot(x2 , result.test[, 1], col="red", main = " ((Real Data(red) Vs Sim. result(blue))",
     type = "l", ylab = "Anxiety")
lines(x2, result.test[, 2], col="blue", type = "l")
par(op)
par(op)
plotMF(object)
summary(object)
```

3.12.2 Genetic Cooperative-Competitive Learning (GCCL)

This method is based on Ishibuchi et al. (1999) using genetic cooperative-competitive learning (GCCL) to handle classification problems. In this method, a chromosome describes each linguistic IF-THEN rule using integers as its representation of the antecedent part. In the consequent part of the fuzzy rules, the heuristic method is carried out to automatically generate the class. The evaluation is calculated for each rule, which means that the performance is not based on the entire rule set. The method works as follows: Step 1: Generate an initial population of fuzzy rules. Step 2: Evaluate each fuzzy rule in the current population. Step 3: Generate new fuzzy rules by genetic operators. Step 4: Replace a part of the current population with the newly generated rules. Step 5: Terminate the algorithm if the stopping condition is satisfied; otherwise, return to Step 2 (Lala Septem Riza et al., 2015).

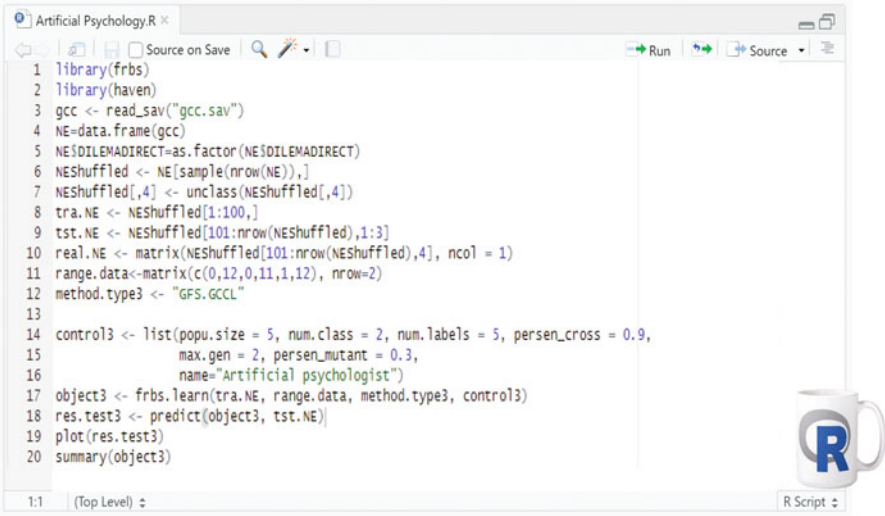
The genetic cooperative-competitive learning (GCCL) algorithm (Ishibuchi et al., 1999) employs a GA to optimize the rule base while the database is fixed. Thus, a computationally effective classifier with an interpretable rule base can be obtained using genetic fuzzy systems for rule induction processes. Fuzzy systems have shown their usefulness in solving a wide range of problems in different application domains. The use of evolutionary algorithms (EAs), and particularly genetic algorithms (GAs), in the design of fuzzy systems allows us to equip them with the learning and adaptation capabilities. The result of this hybridization between fuzzy logic and GAs leads to genetic fuzzy systems (GFSs) (Throckmorton et al., 2015). The most influential aspect of any GFS is the genetic representation of the solutions. In this sense, the proposals in the specialized literature follow two approaches in order to encode rules within a population of individuals (Throckmorton et al. 2015). The "Chromosome = Set of rules" is also called the Pittsburgh approach, in which each individual represents a whole rule set. Thrift proposes in Chen et al. (2017) a method that follows this approach. In turn, within the "Chromosome = Rule" approach, there are three generic proposals: the Michigan approach, the IRL (iterative rule learning), and the GCCL (genetic cooperative-competitive learning) approach, in which the complete population or a subset of it codifies the rule base. This approach makes it necessary to introduce a mechanism to maintain the diversity of the population in order to avoid all individuals in the population converging to the same area of search space.

Practical Example Using R

The artificial cognitive psychologist seeks to explore fuzzy rules to predict a class (duty-oriented, utility-oriented) based on three variables: neuroticism, extroversion, and lie. Rules were obtained using GFS-GCC and SLAVE algorithms.

The different forms of this fuzzy rule-mining are shown in the following figures. Listing 3.10 depicts the r codes for running the two algorithms.

Listing 3.10 R codes for GCCL implementation



As the output from R software reveals, the artificial cognitive psychologist divides people into two categories based on a task: people in direct dilemma preferring duty-orientation and the group preferring utility-orientation (utility-oriented: class 1, duty-oriented: class 2). Then, they are all measured in terms of extroversion, lie, and neuroticism with valid and reliable scales.

Then, GFS_GCCL is used for the soft triangular fuzzy membership function ± (product), standard s-norm, and implication function equal to ZADEH and a fuzzy set of 5 (very small, small, medium, large, and very large). The population created in each generation has 5 members with a crossover probability of 0.9 and a mutation probability of 0.3. In this study, the maximum number of generations for the genetic algorithm is 2 for ease of running. According to the procedure, 80% of the sample size (100 out of 125 people) is the training sample, and 20% (245 people) is the test sample.

Figure 3.27 shows the fuzzy classification rules. For example, in rule 5:

If neuroticism is very large and extroversion is small and the lie is large, then the direct dilemma is utility-oriented (class 1).

According to Fig. 3.27, the certainty factor for the rules indicates that certainty is 85% for rule 1 and 67% for rule 5.

The certainty factor indicates trust in rules in rule-based systems.

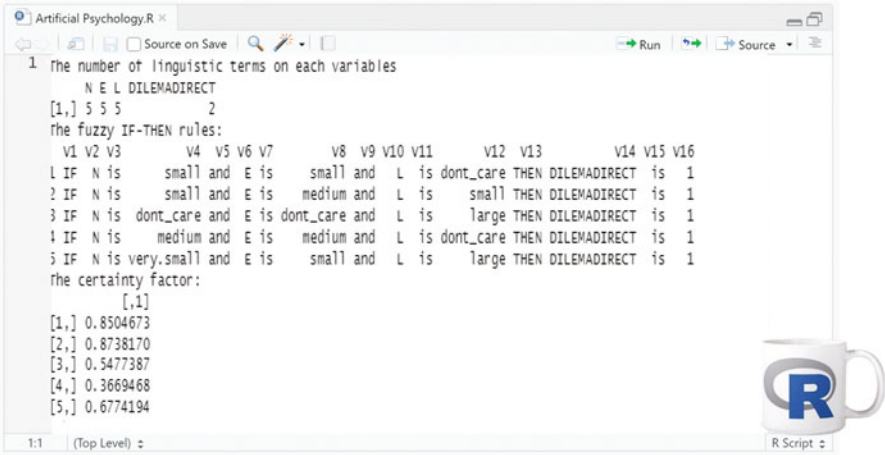


Fig. 3.27 The fuzzy classification rules and the certainty factor

3.12.3 Structural Learning Algorithm on Indefinite Environment (SLAVE)

SLAVE (Structural Learning Algorithms in Vague Environments) is an inductive learning algorithm, which was initially proposed in González et al. (1994) and later developed in Gonzalez and Perez (1999). The basic element of the SLAVE learning algorithm is its rule. IF $X1$ is A and X_n is A_n, THEN Y is B, where each variable X_i has a referential set U_i and takes values in a finite domain (term set) D_i, $i = 1, \ldots, n$. The referential set for Y is V and its domain is F. The value of the variable Y is B, where $B \in F$ and the value of the variable X_i is A_i, where Afi $\in P(D_i)$ and $P(D_i)$ denotes the set of subsets of D_i.

Slave algorithms have used iterative approaches to discover fuzzy association rules. The SLAVE algorithm is one of the fuzzy rules learning algorithms developed after 1994. It is used as a benchmark new algorithm (Ishibuchi et al., 1999). Since its implementation in 1996, this algorithm has been altered many times. When the SLAVE algorithm was first executed in 1994, there were very few algorithms for fuzzy rule learning. Significant recommendations were proposed by Wang et al. Both algorithms were centered on control rule learning (a regression problem). For working with a noise-influenced framework, the SLAVE approach has been discovered where the existing learning approach does not distribute expected outputs in some situations (Tsai & Chen, 2010). The algorithm can define the rules which describe the system from all the variables proposed (feature selection). The SLAVE algorithm mainly uses iterative approaches. Gonzalez and Perez proposed a modified initial iterative approach used in SLAVE (Blake, 1998). In this process, their thought was to incorporate more data to learn one single rule. This data is merged into the iterative methodology over an alternate proposition of calculus to determine the positive and negative guide to a rule. A new function and extra genetic operators are

also proposed that can decrease the time required for learning and develop the understanding of the rules that are obtained. Gonzalez and Perez further worked on the SLAVE algorithm. Later they reduced the time needed for learning processing to obtain a complete rule in each iteration (Wang et al., 2015).

Practical Example Using R

The artificial cognitive psychologist repeats the study using the SLAVE algorithm. This algorithm, which is very similar to GFS_GCCL, has a probability of crossover of 0.9, 5 fuzzy sets, 5 iterations, a maximum of 3 generations, a mutation probability of 0.3, an interval of 0.25–0.75 for the threshold of noise, and an epsilon of 0.1. Epsilon is a number between 0 and 1 that indicates the covering factor.

Listing 3.11 illustrates the R codes for implementing this method. Figure 3.28 shows the R output for this method, which was used to extract 10 rules. Rule 10 states that

> If neuroticism is very large, extroversion is not care, and the lie is small, then the direct dilemma is duty-oriented (class 2).

Therefore, these two methods can be used for classification-fuzzy rule mining.

Listing 3.11 R codes of SLAVE implementation

Artificial Psychology.R

Source on Save | Run | Source

```
library(frbs)
library(haven)
gcc <- read_sav("gcc.sav")
NE=data.frame(gcc)
NE$DILEMADIRECT=as.factor(NE$DILEMADIRECT)
NEShuffled <- NE[sample(nrow(NE)),]
NEShuffled[,4] <- unclass(NEShuffled[,4])
tra.NE <- NEShuffled[1:100,]
tst.NE <- NEShuffled[101:nrow(NEShuffled),1:3]
real.NE <- matrix(NEShuffled[101:nrow(NEShuffled),4], ncol = 1)
range.data<-matrix(c(0,12,0,11,1,12), nrow=2)
method.type2 <- "SLAVE"
control2 <- list(num.class = 2, num.labels = 5,
                 persen_cross = 0.9, max.iter = 5, max.gen = 3, persen_mutant = 0.3,
                 k.lower = 0.25, k.upper = 0.75, epsilon = 0.1, name="Artificial Psychologist")
object2 <- frbs.learn(tra.NE, range.data, method.type2, control2)
res.test2 <- predict(object2, tst.NE)
plot(res.test2)
summary(object2)
```

1:1 (Top Level) R Script

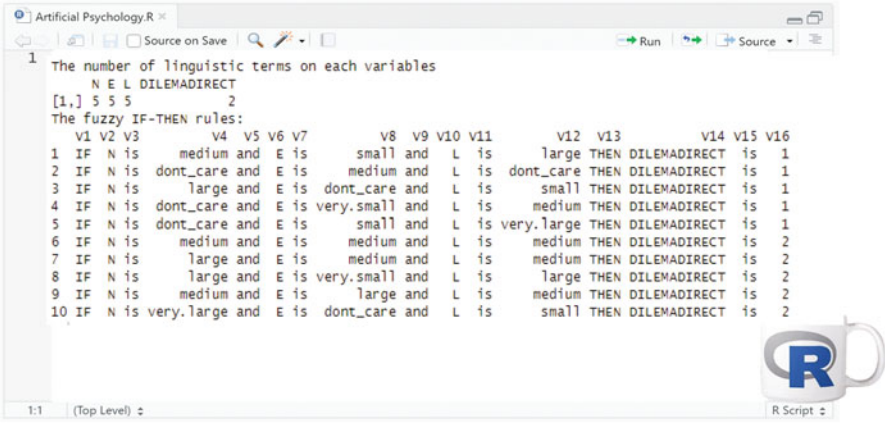
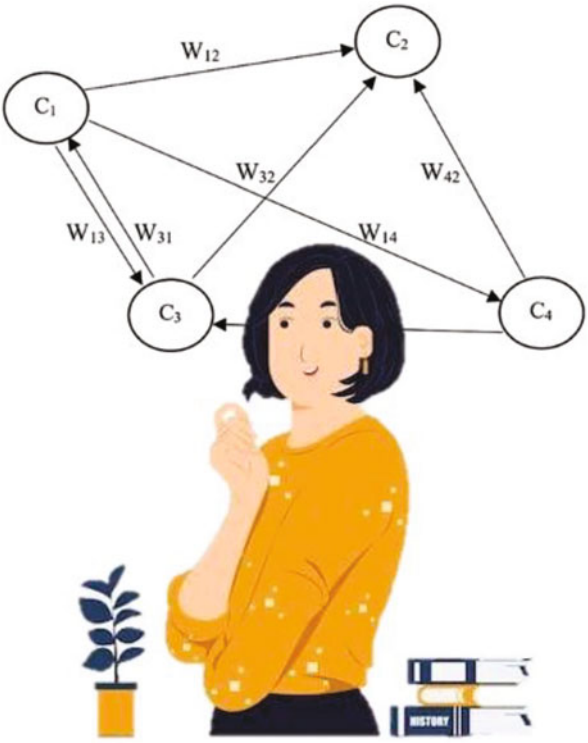


Fig. 3.28 The R output of SLAVE and the extracted rule

Chapter 4
Fuzzy Cognitive Maps



H. Farahani et al., *An Introduction to Artificial Psychology*,
https://doi.org/10.1007/978-3-031-31172-7_4

4.1 Fuzzy Modeling of Human Knowledge: Toward Fuzzy Cognitive Maps in Psychology

Psychology is the study of humans and, in a deeper sense, the study of the mind and mental processes filled with many ideas to achieve goals and desires. Individuals gain valuable experiences in various situations that contribute greatly to the identification of relevant determinants. Those who have failed in relationships, individuals who have participated in weight loss classes, mothers of autistic children, the LGBTQ community, and those who have lived with MS for years are familiar with the basic mechanisms of such experiences. Quantitative studies often focus on factors or predictors or on the statistical inference of relationships between variables (Giabbanelli & Crutzen, 2014); rarely are individuals asked to estimate the effects of variables. For example, studies attempt to identify factors that affect the potential for addiction and measure the significance of such effects using inferential statistics, such as regression. However, these effects are not quantified. Perhaps a psychologist who is a proponent of Sigmund Freud believes it is inappropriate to ask individuals to estimate the effect of variables. They believe that individuals cannot make accurate estimates because people are strongly influenced by the unconscious. Other theories, however, take the opposite view. Research has shown that individuals can accurately estimate how variables affect each other based on an internal schema of a complex system (Hayes & Andrews, 2020).

In order to gain knowledge of the human experience, people are asked to share their knowledge; they are asked to share their experiences rather than focusing on a single experience.

It can be said that individuals, consciously or unconsciously, reduce complexity when they are asked to create models of their experiences; this process can be done to avoid information overload and thus reduce mental effort (Vennix, 1996). Several studies have shown that this simplification is independent of expertise (Axelord, 1974), and solutions have been developed to address this challenge (Papageorgiou et al., 2020; Apostolopoulos et al., 2021).

Psychologists tend to come up with simple solutions! The simplest solution is to collect a data set rather than a person's experience. Research has shown that data sets are also prone to bias and would not necessarily lead to more accurate results (Vennix, 1996).

The Delphi method was introduced in the 1960s to collect group responses. Although it has resulted in accurate responses to a large extent, three major challenges still need to be addressed:

1. Conflicts among participants due to the use of majority rule affect the Delphi method.
2. Relationships have different weights, and the process focuses on the most influential relationships in a system; however, expressing causal relationships (e.g., very high or high) in linguistic terms contains vagueness. Linguistic terms are not unambiguous! This is an inherent property of words.

3. Human knowledge is subject to uncertainty. Uncertainty is the price of living in this world. There is no free lunch!! Everything has its price, and uncertainty is the price we pay for the understanding of the real world; a not quite unaffordable price!

Thus, individuals' answers contain uncertainty and they try to express relationships based on their knowledge to a reliable extent. Fuzzy cognitive maps (FCMs) provide a technique to deal with such challenges. An FCM is a mathematical model that targets such problems, using fuzzy set theory to gather participants' experiences. FCMs were developed to represent uncertainty and vagueness mathematically to provide formulated tools for dealing with impressionism in real-world problems (Li et al., 2006). An FCM includes linguistic terms and helps participants share their experiences. For example, participants may say, "My relationship with my mother strongly influences my mentalized affectivity." The term "strongly affect" has different meanings for different people. It is defined by fuzzy set theory based on overlapping membership functions.

Linguistic terms are integrated into an FCM by rules. For example, if 35 out of 50 alpha women say, "my relationship with my mother has a very strong influence on my mentalized affectivity," and the remaining 15 say, "my relationship with my mother has a strong influence on my mentalized affectivity," the confidence factors of the linguistic terms "very strong" and "strong" are calculated as 0.7 and 0.3, respectively. This is discussed as a fuzzy implication rule in the development of FCMs. The confidence factor represents uncertainty and plays a key role in quantifying participants' general experiences.

To model relationships between variables, several approaches have been introduced. Structural equation modeling (SEM) cannot use linguistic variables and scenario and feedback analyses. The system dynamics approach requires a mediator to resolve vagueness, uncertainty, and conflict among participants. Bayesian networks allow inference under uncertainty but classically fail in the study of feedback. These shortcomings need to be addressed in FCMs.

4.2 Modeling Based on Psychological Knowledge

Where does the psychological knowledge of researchers in the psychology literature, which is the modeling engine of artificial intelligence (AI) psychologists, come from?

The knowledge needed in the psychological literature is obtained through qualitative and quantitative methods. Psychological researchers attempt to explore the mindset of the target group using qualitative schemas, often using content analysis, thematic analysis, grounded theory, and phenomenology (phenomenography) in the form of nondirective in-depth interviews. For example, a target group may include experts, patients, mothers, and couples who can provide the knowledge needed to construct the model.

A researcher is like a bee sitting purposefully on a flower, feeding on nectar and producing honey. Researchers seek in-depth, qualitative assessments of phenomena that are not precise, but rich in concepts. They search for the gold of facts in the minds of informants. After interviewing the target group, the interviews are coded using transcription to extract themes and categories. Once the extraction of themes and categories is complete and the trustworthiness of the results is assessed, an FCM can be developed based on the resulting conceptual model using the principles of qualitative methodology.

In addition, a significant amount of psychological knowledge is obtained through quantitative methods based on questionnaires, scales, standard tests, observation checklists, smart devices, and archived patient records. In such methods, concepts and variables can be extracted by reviewing the literature before implementing a statistical model. The resulting coefficients can then be used to create weighted matrices for an FCM. Therefore, two FCM models can be developed: (1) the thematic tree-based FCM (TTBFCM) and (2) the SEM-based FCM (SEMBFCM).

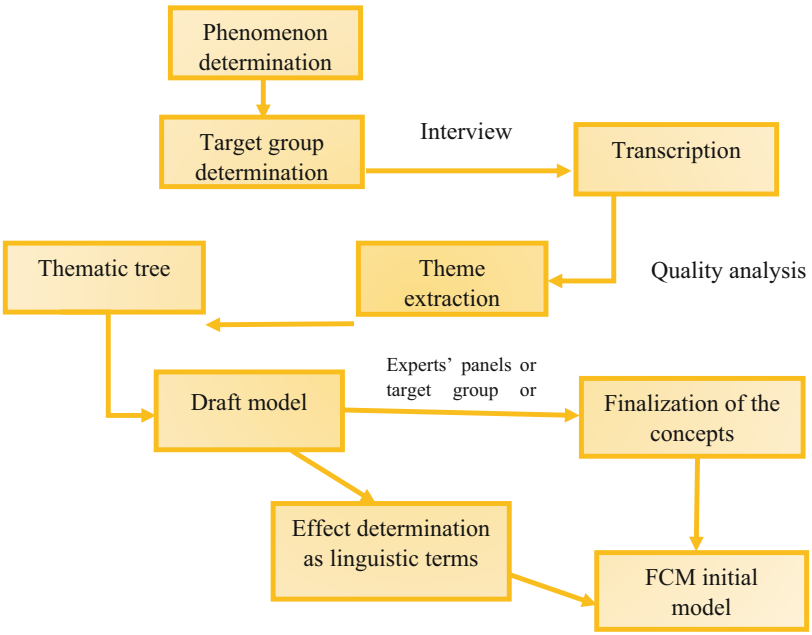


Fig. 4.1 The procedure of constructing TTBFCM

TTBFCMs can be referred to as qualitative or expert-based models (thematic tree-based models in a more general sense). Themes are extracted from nondirective in-depth interviews with the target group and coded using qualitative methods. However, psychologists are aware that a qualitative schema must be aligned with psychology to extract themes.

Once theme extraction and the evaluation of the trustworthiness of the themes are completed, a tree of themes is created. This is discussed below as the basis of the primary FCM models.

The thematic tree is created in draft form and presented to the expert panel (at least ten experts) to be finalized based on the concepts (nodes) or themes, their causal relationships, feedbacks, and loops. The experts are also asked to express their opinion on the causal relationships linguistically (e.g., very high or high). The arrows of the model are labeled. Figure 4.1 illustrates the thematic tree created to assess adjustment after self-injury in female adolescents as the primary model of an FCM.

Domain experts sometimes have different views on model concepts. In such a case, each causal matrix must be completed by adding new columns and rows with zeros for additional concepts. In this way, all causal matrices have the same dimension, and the integration of FCMs can yield a comprehensive description of a single FCM. For example, experts 1 and 2 propose FCM1 and FCM2 for working memory. Then, an integrated FCM can be obtained (Fig. 4.2).

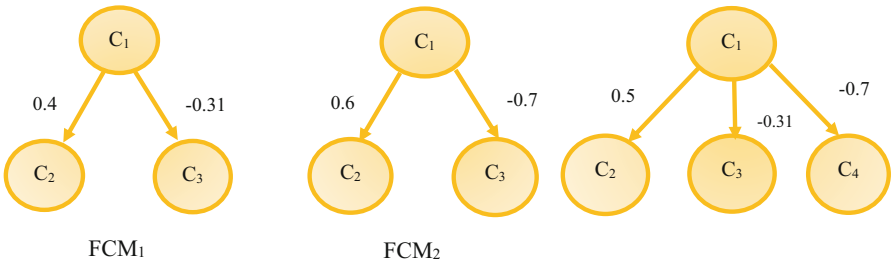


Fig. 4.2 An integrated FCM using two experts' FCMs

The essentiality of FCMs could be determined, as shown in the below table.

Group Name	Essential	Not essential but useful	Non-essential	Comments

Then, the CVR was calculated as

$$\mathrm{CVR} = (N_e - N/2)/(N/2) \tag{4.1}$$

where N_e is the number of experts who consider the node necessary, while N is the total number of experts (10 in this case). The CVR of each node was compared with the Lawshe table, and the nodes whose CVRs were above the Lawshe table remained in the model (Ayre & Scally, 2013).

SEMbFCMs are also a technique for extracting knowledge from questionnaires, scales, tests, smart tools, and observational checklists or brain data, e.g., electroencephalography (EEG). A model can be developed by reviewing the literature, expertise in the field, and expert opinion on SEMbFCMs. Quantitative techniques are used to quantitatively measure the model on a larger scale; the model can then be fitted using SEM. Partial least squares (PLS) will be used for small samples (Hair et al., 2022). The standardized path coefficients are assumed to be the weights of the FCM matrix.

However, statistical feedback does not exist in SEM and can be obtained from the expert perspective (the structural and relational modification of the model). Figure 4.3 shows the procedure for designing a SEMbFCM model. The FCMs are explained in more detail below.

Axelord (1976) proposed cognitive mapping theory (CMT) to model social science knowledge. Kosko (1986) then introduced FCMs as an extension of CMT to model and analyze complex systems. FCMs can be considered a knowledge-based methodology that simulates dynamic systems. It integrates fuzzy logic, neural networks (which will be explained in the next chapters), and cognitive maps to

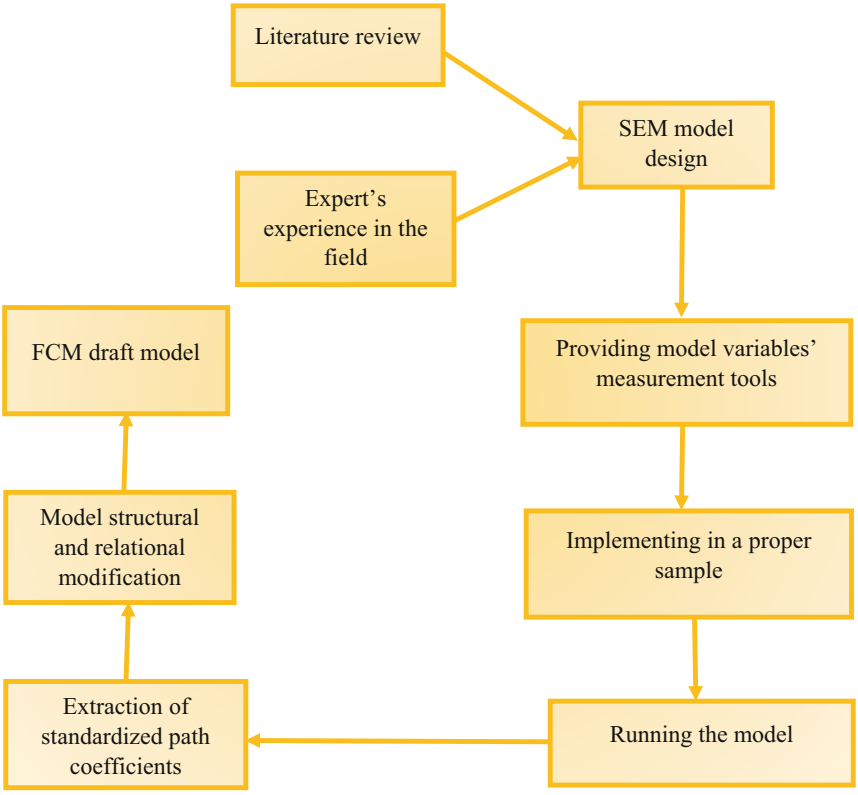


Fig. 4.3 The procedure for designing a SEMbFCM model

represent knowledge about systems whose characteristics are uncertainty, causality, and complex processes (Papageorgiou, 2013).

The FCM is structurally a fuzzy directed graph with feedbacks consisting of components known as nodes and causal relationships between components (nodes) known as weighted directed edges. The dynamics of such systems are tested by simulating their behavior in discrete simulation steps. In general, FCMs can be designed based on inputs from domain experts, system data, or a mixture of experts and data (i.e., expert-based, data-driven, and hybrid FCMs).

The node concept is determined based on the activation degree. A degree of activation refers to the degree to which the corresponding concept (node) influences other nodes in the FCM. Relationships between concepts are defined as positive, negative, or zero. Here, W_{ij} represents the relationship between concepts (nodes) C_i and C_j. It can be positive, zero, or negative. Its sign (whether positive or negative) represents the direction of the effect, while its absolute value represents the intensity of the relationship.

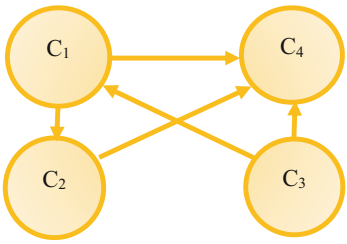


Fig. 4.4 A simple FCM including edges and nodes

Recently, especially in the last 30 years, FCMs have been of great interest to researchers in a variety of fields. They have unique advantages over classical models, including the ability to clearly represent relationships between concepts (nodes) and the study of latent patterns (relationships) between nodes. In addition, their causal semantics can be predicted to analyze, simulate, and test the effects of parameters and estimate the behavior of the system (Fig. 4.4).

As mentioned earlier, there are three approaches in FCM design, namely expert-based, data-driven, and hybrid approaches. However, since explainability and interpretability are at the core of AI psychology (the title of this book), the expert-based approach is said to be the dominant FCM design approach.

The model was designed based on the expert opinions and then delivered to the experts with causal relationships to determine the relationships using linguistic terms (e.g., very strong, strong, moderate, weak, very weak). Then, the linguistic terms or ratings were converted into numerical weights using fuzzy logic (Mkhitaryan et al., 2022). These steps are simple and can be performed by AI psychologists.

Linguistic evaluations are converted into numerical weights in four steps:

Step 1: Defining Fuzzy Membership Functions

Fuzzy membership functions or numbers (e.g., triangular, trapezoidal, and Gaussian) are used to transfer linguistic terms into a specific numerical interval or universe of discourse. The universe of discourse is defined in FCMS in the range $(-1, 1)$ when negative causality is possible; otherwise, the range is $(0, Y)$.

To define a membership function, its form must be determined. As mentioned earlier, there are a variety of membership functions; however, triangular membership functions are the most commonly used. A triangular membership function has an upper bound, a midpoint, and a lower bound (Fig. 4.5).

Step 2: Applying the Fuzzy Implication Rule

In order to determine the degree of activation of the linguistic variables representing the relationships between certain nodes, it is necessary to determine the extent to which the linguistic terms have been verified by the participants. The degree of the causal effect of an antecedent on a consequent in terms of linguistic terms is determined. Then, the fuzzy implication rule is applied to activate the membership functions. In this context, Larsen's product and Mamdani's minimum are often used. Based on Mamdani's rule, the membership function is truncated at the verification level (Eq. (4.1) and Fig. 4.1). Based on Larsen's rule, the membership function is rescaled (Eq. (4.2) and Fig. 4.2).

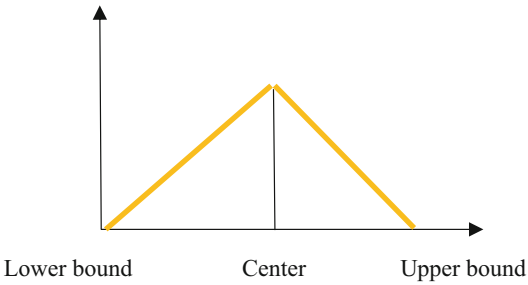


Fig. 4.5 A triangular membership function

Table 4.1 Techniques for integrating membership functions

Argument	Option	Description	Equation
Method	"Algsum"	Family algebraic sum	$x + y - xy$
	"Esum"	Family Einstein sum	$\frac{(x+y)}{1+xxy}$
	"Hsum"	Family Hamacher sum	$\frac{x+y-z*y}{(1-x*y)}$

Mamdani's Minimum:

$$H_R(x, y) = \min[H_A(x), H_B(y)] \tag{4.2}$$

Larsen's Product:

$$H_R(x, y) = H_A(x).H_B(y) \tag{4.3}$$

The next step after applying the percentage importance in the rule is aggregation.

Step 3: Integrating the Fuzzy Membership Functions
The activated fuzzy membership functions are integrated. Several techniques have been proposed for integrating membership functions (Table 4.1).where x and y are the membership values of the linguistic terms activated in the previous step.

The last step of FCM is routine defuzzification of the integrated membership functions.

Step 4: Defuzzification
A number of defuzzification methods have been introduced, including the Center of Gravity method. Defuzzification is used to convert the fuzzy FCM values into crisp numbers. Several defuzzification methods are available in Python, as shown in Table 4.2.

Simulation is an important aspect of FCMs. The dynamics of an FCM are determined by simulating its behavior in discrete simulation steps. The concept values are updated in the simulation using inference methods. The FCM expert must use the following inference rules:

Table 4.2 Defuzzification methods

Argument	Option	Description
Method	"Centroid"	Centroid
	"Bisector"	Bisector
	"Mom"	Mean of maximum
	"Som"	Min of maximum
	"Lom"	Max of maximum

Kosko
Kosko's activation rule with self-memory (modified Kosko)
Rescaled activation rule

A_j^t is the value of node j in step t, W_{ij} is the causal effect of concept (node) j on concept i, and $f(x)$ is a transfer function. The transfer function is used so that the values fall within a certain range – for example, (0, 1) and (−1, 1) for the sigmoid and hyperbolic tangent transfer functions, respectively. There are five transfer functions:

- Sigmoid → X is defuzzified value λ is a steepness parameter for the sigmoid function
- Saturation function
- Bivalent
- Trivalent
- Hyperbolic

The simulation runs until one of two discontinuity criteria is met: (1) the concepts have a difference in two consecutive steps that is smaller than the threshold difference so that the network converges at a fixed point, or (2) the predefined maximum number of iterations has been completed.

$$\exists(E1, 2, \ldots, t-1) : \left|A^{t+1} - A\right| < \text{threshold} \tag{4.4}$$

In FCM inference and simulation, the FCM expert defines an initial state vector in which the values of the concepts are determined by the FCM expert. An FCM weighting matrix, i.e., a connection or adjacency matrix, is also required. Once the initial state vector and the FCM weight matrix are multiplied, the system is updated until one of the two termination criteria is met.

4.3 Optimization in FCMs

An expert-based FCM can converge to undesirable positions, which is a drawback. Therefore, such models predict only extreme values. Researchers have proposed a number of techniques to circumvent this drawback (Lavin et al., 2018).

The active Hebbian learning (AHL) and nonlinear Hebbian learning (NHL) approaches have been proposed. They are based on the Hebbian learning rule. Here, machine learning (ML) and the FCM approach are integrated. These algorithms aim to modify the initial weight matrix so that the selected nodes converge in the predefined region. These nodes are referred to as DOCs. Although human knowledge is a powerful tool in the design of FCMs, expert opinions may be lacking or too subjective and inaccurate in some cases (Papageorgiou, 2011). In addition, there may be too large a number of variables/components. Learning algorithms are a mechanism to address such challenges. These algorithms increase the accuracy of the weights and reduce the dependence of the weights on expert opinions, improving the performance of FCM by creating a learned weight matrix (Papakostas et al., 2011; Papageorgiou, 2011). Overall, learning algorithms cope with the convergence problem of FCMs and enable reliable decisions (Papageorgiou et al., 2004).

Learning algorithms are similar to simulations. However, learning algorithms update the values of the concepts (nodes) at each time step, whereas simulations only change the concepts (nodes) (Papageorgiou et al., 2004). AHL updates the nodes and weights asynchronously based on the activated pattern sequence. It not only optimizes the existing wedges but also creates new weights between nodes that may not be desirable.

Based on such optimization algorithms, the learning process continues until one of two termination criteria is met: (1) the fitness function F_1 of each DOC decreases at each time step, or (2) the DOCs fall within the predefined range.

The fitness functions of the DOCs are calculated (Eq. (4.5)). For the second criterion, it is important to identify whether the DOC is fixed (F_2). A threshold is defined, which should be (0.001, 0.005). As soon as it is reached, the process is terminated. The learning process is terminated as soon as the two termination criteria are met. Otherwise, the predefined number of iterations is performed.

$$F_1 = \sqrt{\left| \mathrm{DOC}_j^k - \frac{\mathrm{DOC}_j^{\min} - \mathrm{DOC}_j^{\max}}{2} \right|^2} \tag{4.5}$$

$$F_2 = \left| \mathrm{DOC}_j^{k+1} - \mathrm{DOC}_j^k \right| < e \tag{4.6}$$

To implement the AHL and NHL algorithms, the initial weight matrix, initial concept values, and DOCs must be determined. In addition, the learning rate, the decay coefficient, and the coefficient of the sigmoid function should be determined as hyperparameters. The slope and decay rate of NHL and AHD as well as the learning rate are usually set to [0.9, 1001], (0.99, 1), (0.01, 0.1), and (0.00, 0.1), respectively. It should be noted that population-based search methods exist for learning in FCMs used to optimize weights and slopes (Mkhitaryan et al., 2022).

4.4 Scenario Analysis in FCMs

The analysis of scenarios and their effects on the model is an advantage of FCMs. A scenario can be assumed to be a new FCM factor whose causal effect on the target concepts (nodes) is measured (continuous interventions) or that changes the base values of the concepts (nodes) to the values of the target concept (nodes) values (one-time interventions).

To evaluate continuous interventions in Python, the name of the intervention, the concepts (nodes), the targets, and the effects of the nodes on each other must be determined. For one-time short interventions, on the other hand, the name of the intervention and new initial states must be determined. It is also proposed to evaluate the effectiveness of a given intervention in the range of (0, 1), where 1 is the maximum effectiveness. The strength (intensity) of the intervention is intended, and the actual value of the change (increase or decrease) after the intervention is simply determined (Mkhitaryan et al., 2022).

Actual effectiveness due to intervention = expected effectiveness * amount of effectiveness after intervention.

To assess explainability, which is the main goal of explainable AI psychology, charts are used to measure the changes after each intervention.

4.4.1 Practical Example Using R

An artificial psychologist tries to simulate a model based on FCM to explain cyberloafing. He derives an initial conceptual model based on an in-depth review of the related research background. In this model, he specifies two important variables $C1$ and $C2$ in relation to the target variable (cyberloafing): (smartphone addiction = $C1$) and (loneliness = $C2$). He then consults with a research team who have worked in the field of Cyber Psychology. They also suggest the stress variable and the final model is conceptualized as follows, where $C3$, stress, and $C4$ are the target variable, i.e., cyberloafing (Fig. 4.6).

The assumed model is presented to two experts who work in the field of mental health and health psychology, and they are asked to specify the label of each path (edge) in an 11-category scale.

Table 4.3 shows the fuzzy set, the type of fuzzy membership, the parameters, the direction (sign), and the size (the degree of influence of each variable).

Then the experts were asked to specify in the figure if they consider each of the fuzzy sets suitable for each edge in the assumed model. The figure shows the opinion of expert 1 about the routes.

Then, using the necessary codes in the software of these comments in the form of Fuzzy Logic, it is aggregated and the matrix of weights is formed, or adjacency matrix (Table 4.4).

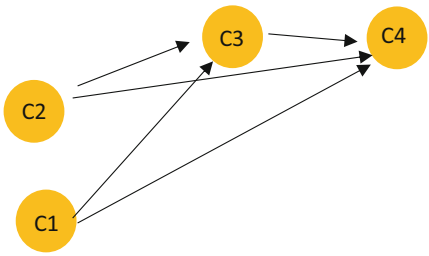


Fig. 4.6 Hypothetical model of FCM

Table 4.3 The fuzzy set, type of fuzzy membership, and the parameters

Parameter	The type of fuzzy membership functions	Fuzzy set
(−1, −0.75, −0.5)	Triangular	Very high negative (−VH)
(−1, −0.75, −0.70)	Triangular	High negative (−H)
(−0.75, −0.5, −0.25)	Triangular	Negative average (−V)
(−0.5, −0.25, 0)	Triangular	Negative bottom (−L)
(−0.25, 0, 0)	Triangular	Very low negative (−VL)
(−0.001, 0, 0.001)	Triangular	Don't know (NA)
(0, 0.25, 0.5)	Triangular	Very low positive (+VL)
(0.25, 0.75, 0.8)	Triangular	Positive low (+L)
(0.5, 0.75, 1)	Triangular	Medium positive (+M)
(0.57, 1, 1)	Triangular	High positive (+H)
(0.75, 1, 1)	Triangular	Very high positive (+VH)

Table 4.4 The adjacency matrix of the FCM

	*C*1	*C*2	*C*3	*C*4
C1	0	0.75	0.28	0.92
C2	0	0	0.79	0.89
C3	0	0	0	0.90
C4	0	0	0	0

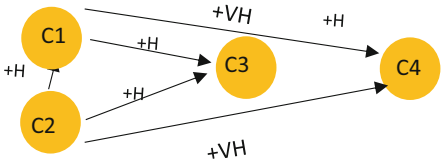


Fig. 4.7 CM model based on experts' opinion

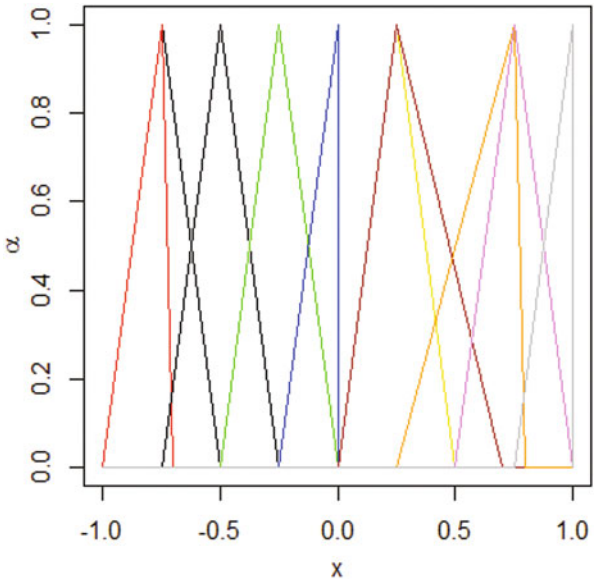


Fig. 4.8 The fuzzy membership function showed in Table 4.4

After 5 iterations, $C3$ = stress = 0.71 and target variable = cyberloafing = 0.84. The diagram is given as Fig. 4.7. In FCM, the initial value determined by the artificial psychologist for variables $C1$ and $C2$ is called the initial vector or activation vector. This vector contains values related to concepts or primary nodes, each of which is activated by assigning the number 1 or a number between 0 and 1. The weight matrix or adjacency matrix actually shows the normalized relationships between concepts or variables, which are usually placed in the distance (0 and 1) or (-1 and +1). The number of iterations is the number of times after which the values obtained from the simulation do not change more than a certain limit (threshold). For fuzzy inference based on FCM, it is necessary to determine the rules. In general, Dikopoulou et al., (2018) introduce six different rules for fuzzy inference: (1) Kosko, "k," (2) modified Kosko, "mk," (3) rescale, "r," (4) Kosko clamped, "ks," (5) modified Kosko clamped, "mkc," and (6) rescale clamped, "rc." As a pre-selected default, the inference rule is usually set to Kosko. In this example, as specified in the R codes, the artificial psychologist used Kosko's rule for inference.

Since the output of the vector product in the weights matrix may be a non-fuzzy number {out of 0 and 1 or -1 to +1} to convert that number to a number that is in the range of 0 to 1 or -1 to +1 The transformation function is used.

These functions are bivalent, "b," trivalent, "tr," sigmoid, "s," or hyperbolic tangent, "t.". From the function sigmoid, "s" is used.

To stop the iteration, the value of epsilon (e) is used, which indicates the minimum difference between the concepts resulting from the iterations. By default, these values are 0.001. In this example, $e = 0.001$ was considered (Figs. 4.8 and 4.9).

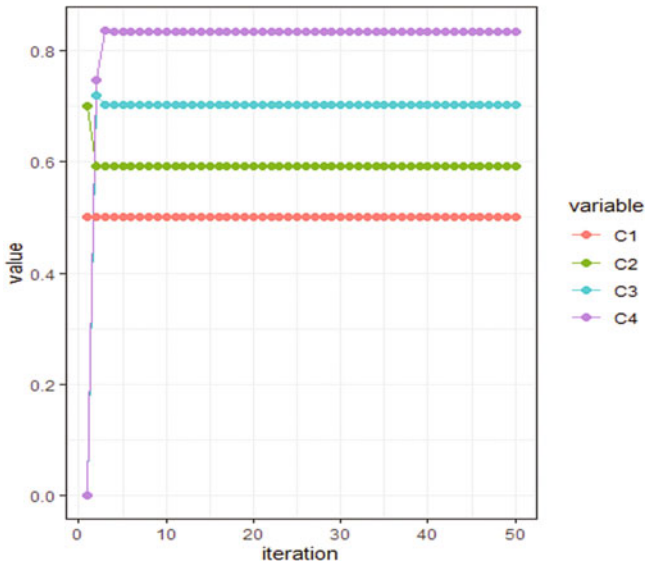


Fig. 4.9 The plot of the iteration step

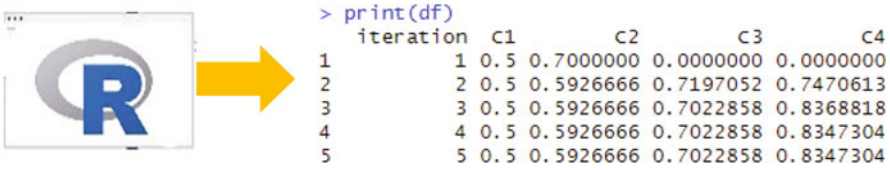


Fig. 4.10 Converging the FCM at iter5

As explained in the previous section, since the weight matrix or adjacency matrix may falsely converge, the AHL and NHL approaches are used for this reason. The purpose of these two algorithms is to modify the initial weights matrix so that the nodes converge in the preferred area and do not reach an incorrect and premature convergence. It is possible to obtain adjusted matrices by using Python software and running AHL and NHL and then import them into R for calculation.

Figure 4.10 shows the output of AHL. In Python software, there is a very good library for FCM implementation, which is out of the scope of this book. Based on that, more diverse and interesting findings can be achieved with this method.

Fuzzy sets and their parameters were approved by experts; see the R codes to draw the shape of this table and its output in Listings 4.1 and 4.2.

Listing 4.1 R codes for FCM

```
library(fcm)
library (reshape2)
library (ggplot2)
act.vec <- data.frame(.5,.7,0,0)
colnames(act.vec) <- c("C1","C2","C3","C4")
C1=c (0,0.75,0.78,0.92)
C2=c (0,0,0.79,0.89)
C3=c (0,0,0,0.90)
C4=c (0,0,0,0)
w.mat <- matrix (c (C1,C2,C3,C4), nrow = 4, ncol=4, byrow = TRUE)
colnames(w.mat) <- c("C1","C2","C3","C4")
w.mat <- as.data.frame(w.mat)
w.mat
output <- fcm.infer(act.vec,w.mat,50,"k","s")
iteration <- as.numeric(rownames(output$values))
df <- data.frame(iteration,output$values)
df2 <- melt(df,id="iteration")
ggplot(data=df2, aes(x=iteration,y=value,group=variable,
colour=variable))+theme_bw()+geom_line(size=0.7)+geom_point(size=2)
print(df)
```

Listing 4.2 R codes for FCM (continued)

```
library(FuzzyNumbers)
A <- TriangularFuzzyNumber(-1, -0.75, -0.50)
B <- TriangularFuzzyNumber(-1, -0.75, -0.70)
plot(A, xlim=c(-1,1))
plot(B, add=TRUE, col="red", lty=1)
C <- TriangularFuzzyNumber(-0.75, -0.5, -0.25)
plot(C, add=TRUE, col="black", lty=1)
D <- TriangularFuzzyNumber(-0.5, -0.25, 0)
plot(D, add=TRUE, col="green", lty=1)
E <- TriangularFuzzyNumber(-0.25, 0, 0)
plot(E, add=TRUE, col="blue", lty=1)
F <- TriangularFuzzyNumber(0, 0.25, 0.50)
plot(F, add=TRUE, col="yellow", lty=1)
G <- TriangularFuzzyNumber(0, 0.25, 0.70)
plot(G, add=TRUE, col="brown", lty=1)
K<- TriangularFuzzyNumber(0.25, 0.75, 0.80)
plot(K, add=TRUE, col="orange", lty=1)
L<- TriangularFuzzyNumber(0.5, 0.75, 1)
plot(L, add=TRUE, col="violet", lty=1)
M<- TriangularFuzzyNumber(0.75, 1, 1)
plot(M, add=TRUE, col="gray", lty=1)
```

	C1	C2	C3	C4
C1	0.000000	0.603612	0.638776	0.728636
C2	0.036480	0.000000	0.638334	0.041642
C3	0.040496	0.047455	0.000000	0.046261
C4	0.037046	0.043394	0.048412	0.000000

Fig. 4.11 The output for AHL

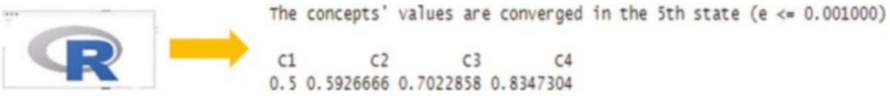


Fig. 4.12 The results of the simulation of FCM

Figure 4.11 shows the output of AHL. In Python software, there is a very good library for FCM implementation, which is out of the scope of this book. Based on that, more diverse and interesting findings can be achieved with this method.

The final concept vector is to be $C\,(5) = [0.50, 0.59, \mathbf{0.71,0.84}]$. The final values of decision concepts are stress $= 0.71$ and cyberloafing $= 0.84$ of the decision-making problem (Fig. 4.12).

Farahani et al. (2021a, b, c) used FCM for studying psychological well-being. In this paper, you can find some computational detail.

Chapter 5
Network Analysis in AP

5.1 Network Analysis in AP

AP is introduced to improve the quality of analyzing psychological models. Psychological phenomena are not independent from each other as consciousness is a dynamic and continuous phenomenon. Human behaviors, however, might seem to be discrete and singular but are not. Therefore, in order to understand a phenomenon one should study it as a series of complex and interrelated features, viewing them as a psychological network. AP, in this sense, could be called the science of psychological networks, networks of complex and interrelated variables. These networks

H. Farahani et al., *An Introduction to Artificial Psychology*,
https://doi.org/10.1007/978-3-031-31172-7_5

are small or big, simple or complex, visible or invisible and dynamic or static. Psychological features in a network could comprise of different types of factors: factors showing cause and effect, environmental factors, character-related factors, and psychological or physical factors. The relation among the features could be negative or positive, directional, unidirectional or reciprocal.

Borgatti, et al. (2009) believe that networks have influenced all aspects of human psychology. In the last decade, network analysis has been viewed as a significant analytical approach in psychological research (Hevey, 2018). Network analysis has a long history in the field, and in recent years there has been a shift from a latent-based approach to a network approach, in order to explain the correlation among variables. In this approach, the observed pattern in the form of correlation can be explained by a mutualism model, in which features have reinforcing and mutual relations. In the network analysis model, the relationship between variables represents the psychological phenomenon (De Schryver, 2015).

Therefore, conceptually, network is an acceptable model of psychological phenomena, enabling precise analysis, and statistically it is a method to study the relations among variables simultaneously.

Psychological constructs are inherently complex and therefore the researchers have for some years focused on restricting the phenomena studied to that of significant variables. Such a reductionism leads to a limiting of our knowledge of the phenomena as a whole (Barabasi et al., 2011).

Network approaches in AP focus on multivariate data to advance several goals. Borsboom et al. (2021a, b) elaborated on these goals. First, they can be used to explore the structure of high-dimensional data in the absence of strong prior theory on how variables are related. Second, in these analyses, psychometric network analysis complements existing techniques for the exploratory analysis of

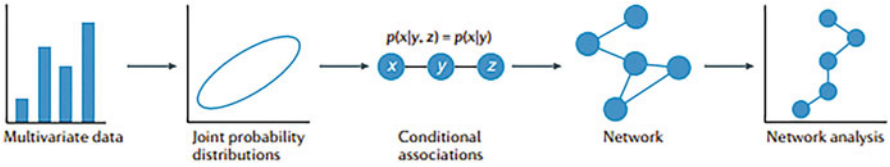


Fig. 5.1 Structure of psychometric network analysis. (Derived from Borsboom et al., 2021a, b, p. 2)

psychological data, such as exploratory factor analysis (which aims to represent shared variance due to a small number of latent variables) and multidimensional scaling (which aims to represent similarity relations between objects in a low-dimensional metric space). The unique focus of psychometric network analysis is on the patterns of pairwise conditional dependencies that are present in the data. Network representations can be used to communicate multivariate patterns of dependency effectively, because they offer powerful visualizations of patterns of statistical association. In other words, it enables the discovery of a communal structure. Third, network models can be used to generate causal hypotheses, as they represent statistical structures that may offer clues to causal dynamics. As stated by Pearl (2000), Spirtes et al. (2000), and Haslbeck et al. (2021), networks that represent conditional independence relationships form a gateway that connects correlations to causal relationships.

Many still believe that causality is derived from the theory and not data. To illustrate network analysis, Fig. 5.1 could be used.

5.2 Structural Analysis of Psychological Network

At an abstract level, a network refers to various structures, consisting of variables. These variables are called nodes and the relationships between these nodes are called edges. Despite social networks, in which associations between people are directly observable (for example friends and enemies), edges in psychological networks depend on statistical analyses and partial correlations among nodes, which represent the power of associations among them. In visual representation of a network, green (or blue) edges represent positive associations and red edges represent negative associations. The thickness of each edge shows the power of association (Jones et al., 2018). Edges can be either weighted or unweighted. Weighted edges reflect the direction and strength among nodes. Alternatively, the edge may be unweighted and simply represent the presence vs. absence of a relationship.

A node can represent a single item from a scale, a sub-scale, or a composite scale. The choice of node depends upon the type of data that provide the most appropriate and useful understanding of the questions to be addressed. This is along with explainability and interpretability, which are the bases of AP. Edges can represent different types of relationships, for example, comorbidity of psychological

symptoms or correlations between attitudes (Hevey, 2018). Farahani et al. (2021a, b, c) investigated the application of network analysis for capturing comorbidity structure in mental disorders.

Generally, two types of edges can be present in a network: (1) a directed/directional edge: the nodes are connected in a one-way effect, or (2) an undirected/undirectional edge: the nodes have a mutual relationship.

A directed network can be cyclic. That is, we can follow the directed edges from a given node to end up back at that node or acyclic; i.e., you cannot start at a node and end up back at that node again by following the directed edges.

Pearl (2000) believes that directed networks can represent causal structures and therefore cannot be cyclic and are surely acyclic (Epskamp et al., 2018). However, in the real world, the cyclic assumption is untenable. In addition, directed networks

suffer from the problem that many equivalent models can account for the pattern of relationships found in the data (Bentler & Satorra, 2010).

Despite the plausibility of many causal psychopathological symptom pathways in networks, there is a need to build a stronger rationale to motivate the causal nature of these relationships (Fried and Cramer, 2017).

Regarding time, networks can be cross-sectional data or time-series data.

5.3 Steps in Network Analysis

Estimating a Network

For a network to be formed, first the edges should be determined. In classical statistics, there are several methods to study the relationships between variables such as correlations, covariance, partial correlations, regression coefficients, odds ratios, and factor loadings. However, in network analysis, the estimation method depends on the type of network. Undirected networks occur frequently in psychology and a frequently used model in estimating such undirected networks is the pairwise Markov Random Field (PMRF), which is a broad class of statistical models. A PMRF model is characterized by undirected edges between nodes that indicate conditional dependence relations between nodes. An absent edge means that two nodes are conditionally independent, given all other nodes in the network. An edge indicates conditional dependence given all other nodes in the network.

Edges are estimated depending upon the type of data. If continuous data are multivariate and normally distributed, analyzing the partial correlations using the Gaussian graphical model (GGM) is appropriate (Costantini & Perugini, 2016). If the continuous data are not normally distributed, then a transformation can be applied prior to analysis (Liu et al., 2009). For ordinal data, polychoric correlations are used (Epskamp et al., 2018).

If all the data are binary, the Ising model can be used (van Borkulo et al., 2014). All of these models can be coded in R. When the data comprise a mixture of categorical and continuous variables, the mixed graphical model can be used to estimate the PMRF (Haslbeck & Waldorp, 2018).

The network complexity requires consideration. The parsimony principle is of great importance here. The higher the number of nodes being examined, then the higher the number of edges that have to be estimated and so the network is more complex. For example, in a network with 10 nodes, 45 edges are estimated. In addition, in the case of an Ising model, the number of estimations is more. Moreover, due to confounding variables, some of these correlations might be fabricated and an increase in the number of nodes can lead to over-fitting and very unstable estimates (Babyak, 2004).

It should be borne in mind that correlations and partial correlations are the bases of estimation in undirected weighted networks, and these, like all statistical techniques, are influenced by sample variation and therefore exact zeros will be rarely observed. However, weak and spurious correlations may occur. In order to limit the

number of such spurious relationships, along with benefiting from theoretical backgrounds, a statistical regularization technique is frequently used. A "least absolute shrinkage and selection operator" (LASSO) is such a regularization technique, introduced by Friedman, Hastie, and Tibshirani in 2008. The estimation of the partial correlation networks is done with a tuning parameter set by the researcher. Studies (Wu et al., 2013) show that LASSO performs well in the estimation of partial correlation networks. LASSO reduces some small weak edge estimates to exactly zero, resulting in a sparse network (Tibshirani, 1996). In other words, LASSO yields a more parsimonious model by reducing the number of connections between nodes, which reflects only the most important empirical relationships in the data. Therefore, the absence of an edge does not present evidence that the edge is in fact exactly zero (Epskamp et al., 2017). Note that there is also the threat of omitting actual relationships when using LASSO.

Among many variants of the LASSO, the graphical LASSO (Friedman, et al., 2008) is generally used in network analysis, as it is easily implemented in software and also it is flexible in terms of data type (Epskamp & Fried, 2018).

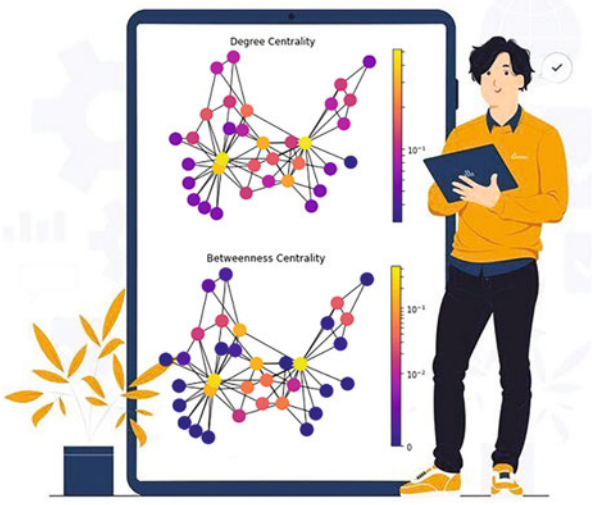
The use of the LASSO requires tuning parameter λ. The higher the λ value, the more edges are removed from the network, directly influencing the network structure. A common method involves estimating a number of networks under different λ. These different networks range from a completely full network to a network with no edges. The LASSO estimates produce a collection of networks and one needs to select the optimal network model. Optimizing is typically achieved by minimizing the extended Bayesian information criterion (EBIC). EBIC works well for both the Ising model and GGM (Foygel & Drton, 2010). EBIC has been used in psychology networks, and it enhances the accuracy of networks (Tibshirani, 1996; Isvoranu et al., 2017).

λ in EBIC is a hyperparameter, which is determined by the researcher and is typically set between 0 and 0.5. Its default value is 0.5 in almost all cases. The explainability and interpretability are of great importance here. The researcher should determine which produced networks are more aligned to theoretical bases. After estimating the network, the important questions to be answered are, Which node is the most important one? Is the global structure of the network dense or sparse? Is the network stable? Do nodes consist of communities or are they singular?

5.4 Descriptive Statistics of Networks

Centrality indices are used to study network descriptions. Centrality indices (CI) represent the relative importance of a node in the context of the other nodes in the network (Borgatti, 2005).

One of the indices of centrality is node strength. How strongly a node is directly connected to other nodes is based on the sum of the weighted number and strength of all connections of a specific node relative to all other nodes.


Degree Centrality
Betweenness Centrality

Another index is closeness, which quantifies the node's relationship to all other nodes in the network by taking into account the indirect connections from that node. A node with high closeness will be affected by changes in any part of the network and can affect changes in other parts of the network quickly (Borgatti, 2005).

In addition, there is an index of betweenness. The betweenness index is defined by the frequency in which a node lies on the shortest path between two other nodes. In other words, this index shows which nodes are bridges.

Clustering is used to interpret psychological networks. The overall network might comprise communities. Each community is a clustering of nodes that are highly interconnected among themselves and poorly connected with nodes outside that cluster (Havey, 2018). Detecting communities helps researchers to interpret the network. Discovering clusters of nodes (communities) helps in interpreting the multiplicity of nodes.

Fried introduces a number of approaches to identify communities. The familiar approaches of latent variable models and "exploratory factor analysis" do this. Communities are in fact factors. There are also more sophisticated approaches, including the Spinglass algorithm. The problem with this algorithm is that it often produces different results every time you run it, and it only allows nodes to be part of one community. Another approach is the Walktrap algorithm. This algorithm provides more consistent results but only allows nodes to be part of one community. Lastly we have the Clique Percolation Method (CPM), which allows nodes to belong to more than one community and fits the needs of psychological research. The last approach is applied in this volume (Epskamp & Fried, 2018).

5.5 Network Accuracy

Determining the accuracy of network analysis is of great importance. In psychological research, the sample size is small, which can limit accuracy. Therefore, looking at the accuracy of node centrality and edge strength is important.

The most common method to estimate the accuracy of edge weights is by calculating confidence intervals (e.g., 95% CI). Epskamp et al. (2018) developed a method that uses bootstrapping (Efron, 1992).

In this method, a model is estimated repeatedly under either sampled or simulated data, and then estimates the required statistic. The more bootstrap samples, the more consistent the results.

Either a parametric bootstrap or nonparametric bootstrap can be applied for edge weights (Bollen & Stine, 1992). Nonparametric bootstrapping could be applied for any type of data and therefore Epskamp et al. (2018) recommended using a method for unbiased estimates with LASSO regularized edges.

5.6 Accuracy of Centrality Indices

The accuracy of the centrality indices can be examined by using different methods. The aim is to determine whether or not the order of centrality indices remains the same after re-estimating the network with less cases or nodes. Therefore, a bootstrap method dropping subsets of cases is applied to assess sensitivity.

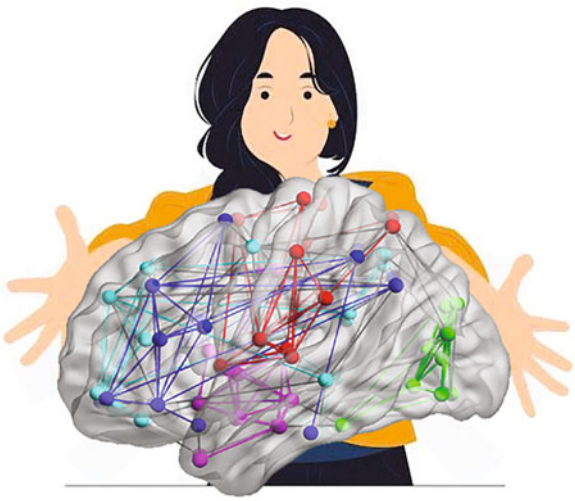
In this method, the correlation stability is calculated. The correlation between the original centrality indices (based on the full data) is compared to the correlation obtained from the subset of data (representing different percentages). If the correlation changes considerably, then the centrality estimate is problematic. A correlation stability coefficient (CS) of at least 0.7 between the original full sample estimate and the subset estimates has been suggested by researchers as being a useful threshold to examine (Epskampet al., 2021).

It should be borne in mind that a CS coefficient shows the maximum proportion of cases that can be dropped, such that with 95% probability the correlation between original centrality indices and subsets is above a threshold.

Borsboom et al. (2021a, b) suggest that to ensure the accuracy of indices, CS should be 0.25 or preferably 0.5.

5.7 Network Science in Psychology

We are living in a connected world, being surrounded by complex networked systems. Blanken et al. (2021) assert that if one wants to summarize all contemporary studies on human processes and behaviors in a sentence, they would say, "It is complicated."



The connection between neurons, interrelation of psychological states and social relations, and the connection of symptoms and psychological disorders can all be considered networks. Network analysis is not a new scientific approach but is considered a new one in psychology. In recent studies, Borsboom has tried to apply this new approach in psychological research. The literature of applying network analysis in psychology is very extensive (for instance, see Fonseca-Pedrero, 2018; Bringmann et al., 2022).

5.8 Network Science in Cognitive Psychology and Neuroscience

5.8.1 Complex System

The features of these systems are neither completely random nor completely regular but instead indicate a rather complex organization. In our surroundings, such systems range from societies, economies, and ecosystems to infrastructure systems, data processing networks, and molecular interactions in biological organisms.

Despite recent breakthroughs in neuroscience, much remains unknown about the brain's complex functions. Network neuroscience is an interdisciplinary branch of science aimed at better understanding these issues. There are two general procedures in this field: (1) using new experimental tools and comprehensive maps and recording dynamic patterns in molecules, neurons, brain regions, and social systems; and (2) using theoretical and computational tools of modern network science.

Nowadays, neuroscience is faced with big neural data. "Big data" typically represent networks that contain relationships or interconnections that link the many elements of large-scale neurobiological systems. These include protein interaction and genetic regulatory networks, synaptic connections and anatomical projections among brain areas, dynamic patterns of neural signaling, and communication associated with spontaneous and task-evoked brain activity and interactions among brain systems, and the environment in the functioning course of behavior. These data have different domains and types (for example, anatomical and functional connectivity, genetic patterns and disease states, and activity in distributed brain regions in relation to behavioral phenotypes) (Bullmore & Sporns, 2009; Medaglia, et al., 2015).

5.8.2 The Brain as a Complex System

The brain is the most complex network known to humans. The human brain is made up of about 100 billion (10^{11}) neurons interconnected by about 100 trillion (10^{14}) synapses, which are anatomically organized at different spatial scales and

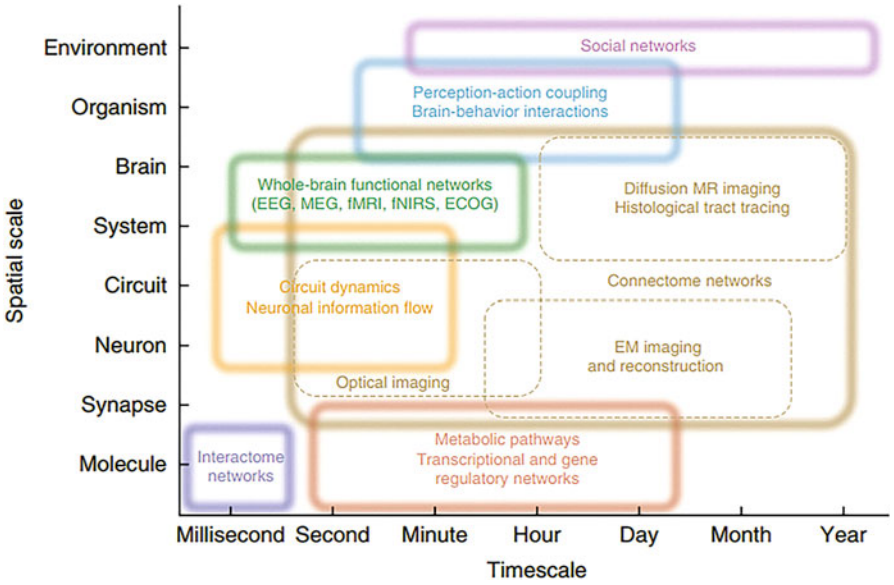


Fig. 5.2 Spatial and temporal scales. (From Bassett & Sporns, 2017)

functionally interact at different time scales. This enormous system is the biological hardware from which all our thoughts, feelings, and behavior emanate (Fornito, et al., 2016).

Network neuroscience entails the analysis of many different networks that are at different temporal and spatial scales. These studies start from the smallest elements for finding the connection between the data encoded in the relationships between genes and biomolecules and continue to higher levels. Network neuroscience is aimed at gaining a deeper insight into how neuron-level processes affect the function of large-scale circuits, neural systems, and the entire brain's structure and function. However, instead of stopping at the brain, network neuroscience asks how these patterns of interconnections in the CNS guide and interact with behavioral patterns: how perception and action are interrelated and how brain–environment interactions affect cognition. The following figure indicates these spatial and temporal scales (Bassett & Sporns, 2017) (Fig. 5.2).

Research suggests that a small number of mutations or risk factors cannot fully explain the biological basics of certain psychiatric diseases. Instead, these psychiatric diseases involve disorders in biological networks.

In neuroscience, descriptive metrics of local and global features of network topology are used in structural and functional data. These analyses rely on non-random topological features such as high clustering and short path length, and network communities (modules) linked by highly connected hub nodes that are in turn densely linked, forming an integrative core or rich club. The descriptive metrics in the graph and brain network will be examined in the following.

5.8.3 *Brain Connectome*

The concept of connectome was first introduced by Olaf Sporns, Giulio Tononi, and Rolf Kotter and independently (Sporns, et al., 2005). Figure. 5.3 depicts the first attempts to obtain the brain's functional matrix. The concept of connectome originally represented knowledge about the brain's cellular wiring diagram. With the

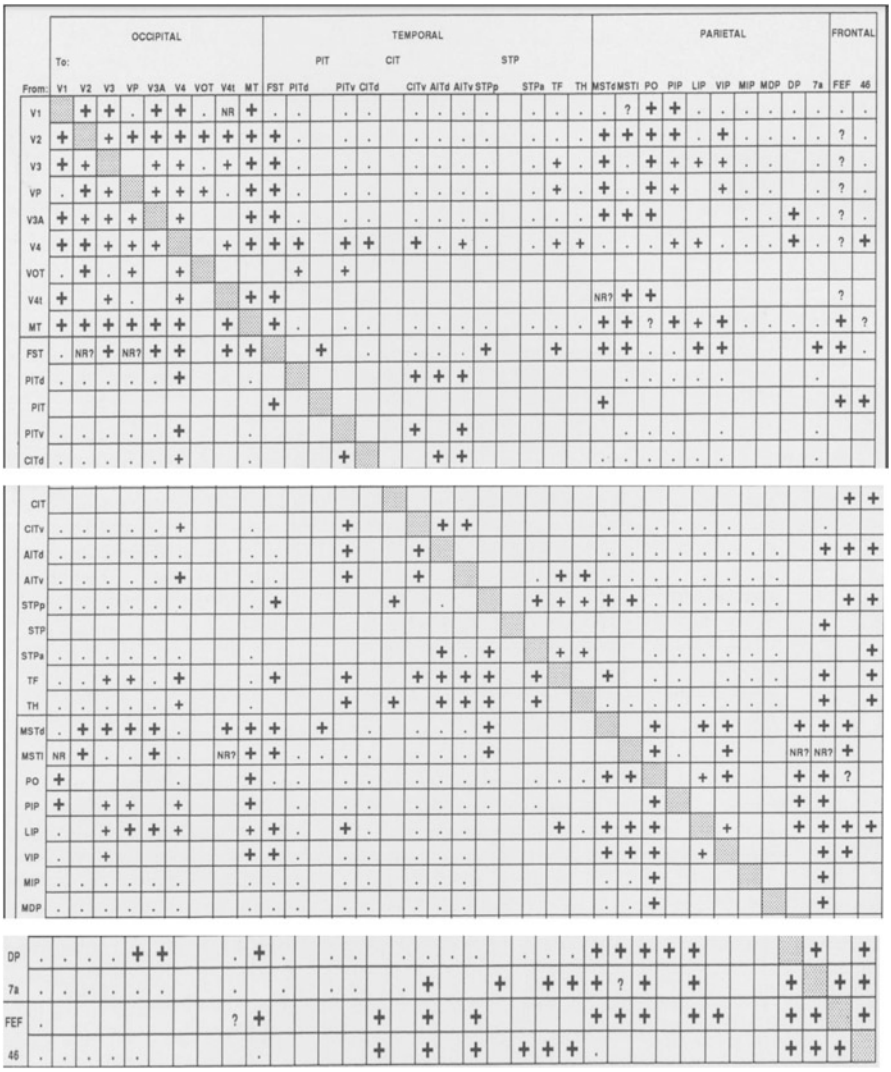


Fig. 5.3 One of the early attempts to systematically create a brain connectivity matrix (Felleman & Van Essen 1991) shows the connection of 32 neocortical regions involved in eyesight function of the macaque

developments of the last ten years, however, this has expanded to cover more general concepts. In fact, the connectome includes the matrix of anatomical connections between large-scale brain regions and between neurons, plus the functional matrix obtained through MRI (low frequency, <0.1 Hz) or EEG (high frequency, >500 Hz).

Further studies have shown that a person's connectome is unique and similar to a fingerprint. Machine learning can use these neuroimaging-based biomarkers to build diagnostic or prognostic instruments. Visualizing and interpreting these models can complement statistical analysis to provide insight into the dysfunction of resting-state patterns in brain disorders.

5.8.4 Various Scales for Network Analysis of the Brain

Unfortunately, there is no single technology to measure brain networks at all biological scales. This means that considering connectomics at multiple scales inevitably requires considering several different measurement methods. However, graph theory does not consider the scale and measurement method and presents a comprehensive language for understanding brain network topology.

Connectomics typically makes a distinction between three spatial scales to determine nodes: namely microscopic, mesoscopic, and macroscopic, and the special techniques used at each scale limit how nodes and edges can be defined.

I. Microscopic: The microscopic scale refers to features that can only be detected using microscopic techniques. In other words, the network consists of neurons and synapses. Neural tissue is incredibly dense – rough estimates suggest that there are 90,000 to 100,000 neurons, about a billion synapses, and kilometers of axons and dendrites embedded within 1 square millimeter of the human cortex. Therefore, it is very difficult to find neural structures and imaging devices should have sufficient accuracy to examine this structure. From a biological perspective, this level is very significant but difficult to examine due to its tremendous complexity.
II. Mesoscopy: The mesoscopic scale bridges the gap between the microscopic and the macroscopic scales and combines both methods to precisely understand the connections of the entire brain or large parts of it.
III. Macroscopic: Unlike the microscopic mode, this scale does not need microscopic techniques. In fact, this scale includes analysis of structural and functional interactions in populations of neurons obtained by EEG, MRI, or MEG. In the following, macroscopic methods, graph theory, and machine learning at this scale will be examined (Fornito, et al., 2016).

5.8.5 *Networks in the Brain*

Building a brain network requires defining nodes and edges. The method commonly used for designing predictive models based on functional connectivity has three main steps: brain segmentation, estimation of interactions between defined segments, and finally, application of these connections between brain segments as features that enable the classifier to predict behavioral features, emotional metrics, and other variables (Fornito, et al., 2016). The defined segmentations act as network nodes and their links are considered edges.

Some segmentation techniques are based on mapping anatomical or functional atlases onto an individual's brain whereas other techniques are more data-driven and try to obtain segments based on common features within the data (Fornito, et al., 2016). The first approach is known as hard segmentation, which includes the use of brain atlases, and the second approach uses statistical methods such as independent component analysis (ICA), which are unsupervised. (Unsupervised methods are discussed in the following sections.)

This method considers nodes from different brain regions and uses time series analysis to create the edges used in functional networks. This method is used widely. Then, a functional matrix is created using the correlations between their time series. There are many methods to obtain correlation or covariance between brain regions. Although the correlation matrix is easy to calculate, it is a fully correlated and dense matrix where all the nodes tend to be interconnected and this creates an overly-dense, clustered, and modular network that has dependencies with no anatomical basis. Although partial correlations can somewhat prevent this problem, ultimately, data interpretation is difficult (Pervaiz, et al., 2020).

Typically, there are three categories of brain connectivity: namely structural connectivity, functional connectivity, and effective connectivity.

Structural Connectivity

Structural connectivity refers to anatomical connections between neurons such as microscale axons and synapses or large-scale or macroscopic connections between different cortical areas, which typically employ unique methods at each level: the electron microscopy method at the micro level, axonal tract-tracing at the meso level, and diffusion MRI at the macro level. Since each axon has a source and a destination, this connection is inherently directed (Bullmore & Sporns, 2009) (Fig. 5.4)

Functional Connectivity

Functional connectivity refers to the statistical dependence between separate neural elements that determines which parts of the brain work with each other and which work independently. Their interpretation depends on the type of recordings analyzed. These connections can be directed or undirected.

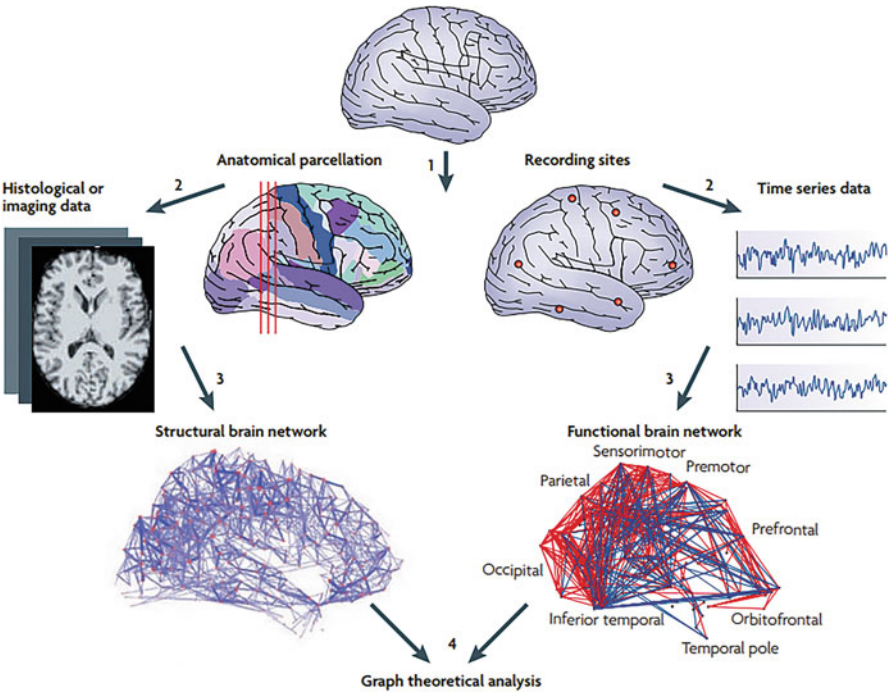


Fig. 5.4 Representation of the stages of creating structural and functional networks of the brain. (Bullmore & Sporns, 2009)

Functional connectivity is measured using statistical dependence metrics between neural time series, which can be discrete or continuous.

At this level, the neuron is considered the basic element of the nervous system. It, therefore, seems reasonable to conclude that the cellular scale is a natural scale for the study of brain connections. This rationale underpinned the first attempt to reconstruct the connectome – the nervous system of the nematode *C. elegans* at the neuron and synapse level. This scale of analysis produces an accurate drawing of nodes and edges, which seems very useful. However, it has disadvantages such as the computational burden, labor-intensive data acquisition and processing, and poor scalability for large neural systems.

In fact, the brain's BOLD signals detected in fMRI indicate spontaneous oscillations without any external stimuli. Biswal et al. observed a high degree of correlation of spontaneous neural activity between bilateral motor areas during the resting state. Since then, cognitive neuroscience research has entered a new era of functional connectivity analysis. Due to its simple experimental design, easy operation, and easy adoption by patients with neuropsychiatric disorders, rs-fMRI has its own unique advantages in the functional study of the human brain. Functional connectivity only shows the synchrony of neural activity of spatially separated brain regions. Researchers have proposed a series of functional connectivity analysis

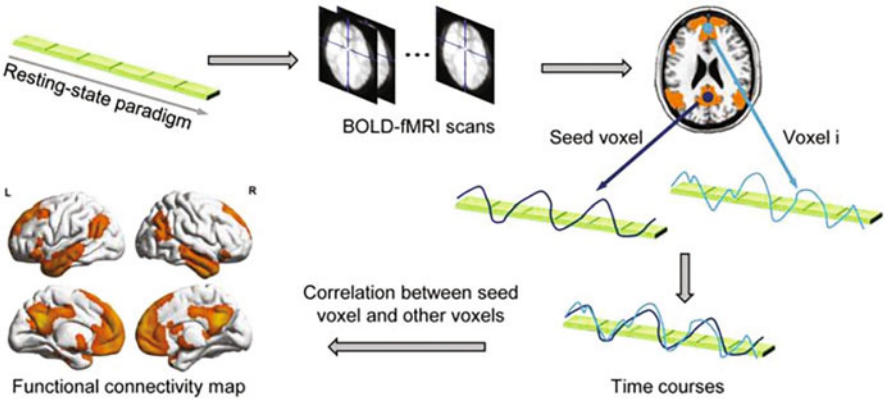


Fig. 5.5 Producing the functional network of the brain. (Hu & Zeng, 2019)

approaches, such as linear correlation analysis, independent component analysis (ICA), principal component analysis (PCA), coherence analysis, and cluster analysis. A node-based analysis is the most common approach to linear correlation analysis. First, regions of interest (ROIs) are selected as nodal regions according to prior knowledge, and then Pearson's correlation coefficient is calculated as a measure of functional connectivity between the period of a given nodal region and voxels in the brain. By dividing the whole brain into several regions (or defining a number of ROIs), region-to-region functional connectivity can be obtained to construct a whole-brain functional connectivity network. There is another approach that does not require considering a single node or ROI and network analysis can be achieved at the whole-brain level (Hu, & Zeng, 2019) (Fig. 5.5).

Effective Connectivity

Effective connectivity was initially used to understand the coherence in the spiking activity of neurons and was defined as the minimum neuronal circuit model that can reproduce the observed signal coherence. Effective connectivity checks if there is a flow between different components. Unlike functional connections, the direction of these connections is also specified and is consistently directed. This connectivity is mainly examined to find causal connections between different brain areas and can help researchers in finding the basic mechanisms of neuronal dynamics (Bullmore & Sporns, 2009).

5.8.6 *Definition of a Brain Graph*

Network science examines the macroscopic behavior complexity of a system of interconnected elements. Given the availability of high-quality datasets, they can be considered with respect to graph theory quantities. In other words, complex systems show common macroscopic features despite having completely different microscopic features. In many studies, one of these important features, small-world, has been seen in the functional and structural networks of the brain of humans and other organisms in different dimensions.

In network science, other topological features of complex systems such as modularity, hierarchy, centrality, and the distribution of network hubs can be quantified.

The structural and functional networks of the brain can be investigated using graph theory through the following four steps:

I. Defining nodes using imaging techniques.
II. Estimating the continuous size of the inter-node connection through spectral coherence or Granger causality measures between two magnetoencephalography sensors, the probability of connection between two regions of an individual diffusion tensor imaging data set, or inter-regional correlations in MRI measurements of the thickness or volume of the cerebral cortex.
III. Linking all connections between nodes and creating a connectivity matrix where a threshold is (usually) applied to each element of this matrix to create a binary adjacency matrix or undirected graph.
IV. Analysis and calculation of network parameters in the brain network and comparison with the equivalent parameters of a population of random networks (Bullmore & Sporns, 2009).

Graph Analysis Metrics in the Brain

1. Node Degree, Distribution Degree, and Assortativity
The degree of a node is its number of connections with the rest of the network. This is one of the most fundamental metrics of the network that other network metrics are also related to. In random networks, all connections are equally likely, resulting in a Gaussian and symmetric degree distribution. Networks of complex systems usually have non-Gaussian degree distributions, which often have long tails toward nodes with high degrees. The degree distribution of scale-free networks follows the power law. Assortativity is the correlation between the degrees of connected nodes. Positive clustering indicates that high-degree nodes tend to connect. The following will explain assortativity and its application in more detail.

2. Clustering Coefficient and Motifs
If a node's nearest neighbors are also interconnected, a cluster is formed. The clustering coefficient is the number of connections between a node's nearest

neighbors relative to the maximum possible number of connections. Random networks have low clustering averages, while complex networks have high clustering (related to high local efficiency of information transfer and robustness).

The topological motifs of a network are basic building blocks where connections between small sets of three or four nodes are repeated in the network with a frequency greater than chance.

3. Path Length and Efficiency

The path length is the minimum number of edges to be skipped when going from one node to another. Random and complex networks have a short average path length (high global efficiency of parallel information transfer), while normal networks have a long average path length. Efficiency is inversely correlated with path length.

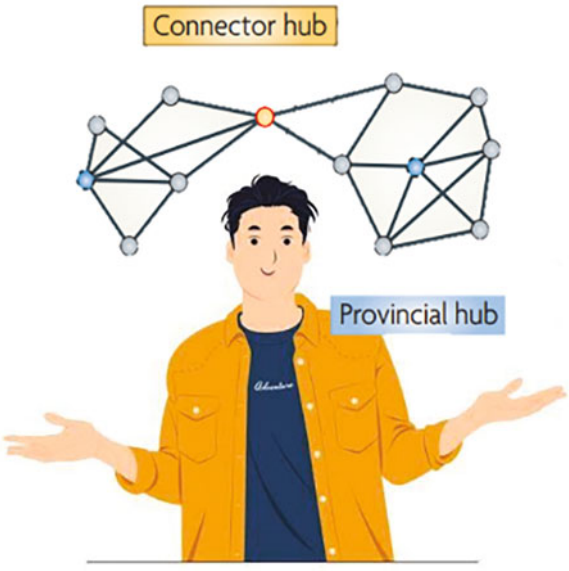
4. Connection Density or Cost

Connection density is the actual number of edges in a graph as a ratio of the total number of possible edges. It is the simplest estimator of a network's physical cost. For example, in fabricating a computer chip, if each logic gate is assumed to be a node and wires to connect them, the amount of wiring between the nodes should be minimized to reduce thermal noise and additional costs (Fornito et al., 2016).

5. Hubs, Centrality, and Robustness

Hubs are nodes with a high degree or high centrality. A node's centrality measures the number of shortest paths between all pairs of nodes in the network that pass through it. Thus, a node with high centrality is essential for efficient communication. An individual node's importance for network efficiency can be evaluated by removing it and estimating the resulting network's efficiency.

Robustness either refers to the network's structural integrity after node or edge removal or the effects of disruption on local or global network states.



6. Modularity

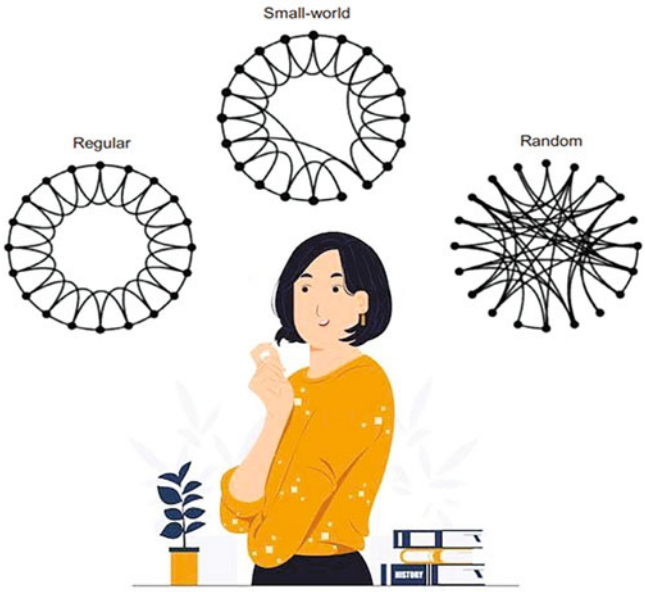
Many complex networks are made up of a number of modules. There are various algorithms for network modularity estimation, many of which are based on hierarchical clustering. Each module includes several densely interconnected nodes, and there are relatively few connections between nodes of different modules. Provincial hubs typically connect to nodes in their own modules, whereas connector hubs connect to nodes in other modules.

In most cases, correlations, coherence, and mutual information are used to create an undirected graph, and correlations, coherence, and mutual information methods are used to create directed graphs.

MRI measurements suggest that centrality and modular organization can be significant biomarkers for early diagnosis and prediction of clinical outcomes in neurology and psychiatry (Fornito et al., 2016).

7. Random, Scale-Free, and Small-World Networks

In random graphs, each pair of nodes has an equal probability of p to be connected and a Gaussian distribution, but most of the descriptive graphs in the real world deviate from the random graph model.



The "small-world" feature has high levels of local clustering among the nodes of a network and short paths that globally connect all network nodes. In fact, all nodes in a large system are connected through relatively few steps. Small-world features are between random networks and regular networks. Research on genetics, signaling, communication, computing, and the neural network has provided evidence for this

feature. These studies reveal that almost all networks in natural and technological systems have non-random/irregular or small-world architectures, and this feature reflects their specific performance (Bullmore & Sporns, 2009).

5.8.7 Brain Network Identification and Analysis in Graph

The use of graph theory in brain network analysis dates back to the introduction of the "human connectome" (Sporns et al., 2005). Here, there is an N*N matrix (also called connectivity matrix or connection matrix) with 0 or non-zero values denoting whether or not there is a connection between the two regions. A brain network analysis of humans or other animals becomes possible upon obtaining the mentioned metrics in this network.

Generally, there are two calculation methods used in connectivity identification: functional connectivity and effective connectivity. As mentioned earlier, functional connectivity provides information about temporal coherence between distant regions, and effective connectivity is the direct effect of regions on each other.

Research on functional connectivity studies can be divided into two categories: model-based and model-free. The model-based mode employs methods such as cross-correlation, coherence analysis, and statistical parametric mapping, whereas the model-free mode uses decomposition-based analysis, clustering, and mutual information methods.

The model-based mode considers as seed a region of the brain to check whether or not that region is related to other regions. This method needs background knowledge for the correct selection of seeds and related areas. In other words, there is a need for a hypothesis, which is verified based on experiments. At the same time, this method may destroy and fail to examine the useful information in the communication neglected by the researcher.

Meanwhile, the model-free mode does not select any specific region or seed and considers the whole brain and its connectivity without a hypothesis.

As mentioned, there can be two types of criteria for the examination of the brain graph: global and local criteria (Fornito et al., 2016).

Global Criteria

The global criteria are aimed at finding functional segregation, information flow in the brain, functional integration, finding the small-world feature, and checking the network's resilience to failure (Sporns, 2014).

In fact, segregation means the degree to each element in the brain that is assigned to populations. In this case, the clustering coefficient and modularity criteria provide good information about segregation. In brain networking, anatomically adjacent regions are called modules, and analysis with this mode has shown the small-world feature in the brain network. Moreover, integration provides insight into the efficiency of information connectivity throughout the brain, and the information can be measured as the length of the path between regions.

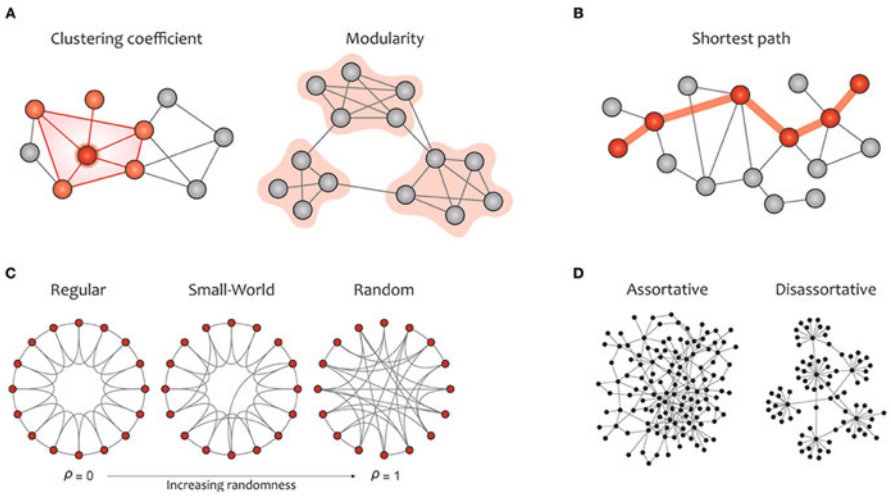


Fig. 5.6 Local criteria and global criteria. (Hu & Zeng, 2019)

In general, the small-world feature represents a balance between segregation and integration between networks.

Regarding the resilience feature, the assortative criterion can be put against failure. In fact, inter-hub connectivity in a network leads to the coverage of a particular hub's failures, and this criterion is also essential in examining the brain network (Fig. 5.6).

Local Criteria

In network science, a node with high centrality that greatly affects the network is called a hub. There are two types of hubs: namely connector hubs or provincial hubs, named based on their participation rate. Connector hubs have a high participation rate, and the opposite is true for the provincial hubs. In other words, connector hubs are responsible for connecting different modules of the brain while provincial hubs are responsible for connectivity between modules. The network's node degree criterion is one of the easiest criteria for finding hubs. Other criteria such as centrality, betweenness, and closeness also belong to this category.

5.8.8 The Brain's Important Networks

The first known brain network was the linguistic network discovered by Broca[1] and Wernicke[2] in the nineteenth century. The two separate parts in an interconnected

[1] Paul Broca

[2] Carl Wernicke

network are responsible for the single activity of language. In general, many networks of different scales have been identified in the brain. There are seven basic brain networks on a large scale:

The Sensorimotor Network

This network receives sensory information from inputs throughout the body and converts them into electrical signals in the brain. This network's responsibilities are divided into several categories: processing the brain's external physical signals, internal signals, examining sensations, and producing motor responses. This network is closely related to other networks such as the auditory network, visual network, limbic network (for a sense of taste and smell), salience network, and the default mode network (DMN). It is considered a transducer in the brain (ten Donkelaar, et al., 2020).

The Visual System

This network can be considered a spectator and is responsible for visual and vision-related processing. This system is very complex since it converts light into something recognizable. Despite the initial assumption that this task is accomplished in one area, it was later discovered that there is a system responsible for these processes. These processes include visual image enhancement and processing and detecting motion, patterns, faces, places, and more (Poggio, et al., 1988).

Limbic System Network

As one of the oldest networks in the brain, this system regulates many brain functions, such as memory, emotions, learning, and behavior. This network responds to stimuli such as smell, sound, and light. It is also responsible for behavior, reactions, and associated feelings, as well as memories of experiences or any type of learning (Sullivan, 2022).

The Central Executive Network (CEN)

Responsible for tasks and decision-making, this network operates at a high cognitive level and is considered one of the most important brain networks. Its responsibilities cover memory, processing, controlling, and combining information from other areas and networks, organizing behavior based on internal motivations, mental preferences, and choices. This high-level network receives and integrates input from other networks to process a variety of information that includes flexibility, working

memory, initiation, and inhibition, which were previously thought to have different networks (Dosenbach, et al., 2007).

Default Mode Network

This is one of the most well-known brain networks that is activated when resting, daydreaming, or contemplating a new idea. This network is active during sleep and rest and becomes more active during inner thinking. These internal thoughts can be about reminiscing about childhood memories, planning for the next vacation, or hunger and bowel movements, which indicate an active DMN. Given the unconscious analysis and contemplation about oneself and the world, it is considered one of the most active and stable brain networks. Many studies on biomarkers use this network (Sambataro, et al., 2010).

Salience Network

This network accurately considers the outside world and determines the brain's response to stimuli. This network acts as a gear for switching from the executive network to the DMN network and back. In other words, it switches between internal and external processes. The reason for the existence of this network is that in healthy brains, the DMN and CEN networks do not activate simultaneously; the regulation is handled by SN (Elton & Gao, 2014).

The Dorsal Attention Network

This network supervises human attention and is bidirectional and coherent for remaining attentive. DAN is often activated in conjunction with other active networks in the brain. DAN directs attention to whichever network (or networks) is the most salient and active. The human brain continuously receives sensory input and cannot consistently pay equal attention to all sensory signals. Instead, DAN focuses the brain's attention on the most important sensory input at a given moment (Corbetta, et al., 1995).

5.8.9 Applications of Graph in Cognitive and Behavioral Science

Cognition covers neural actions that lead to thinking, feeling, or experiencing and includes problem-solving, attention, memory, executive functions, and reasoning (Farahani, et al., 2019).

Intelligence is one area of interest in the connection between the brain network and cognition. Human intelligence refers to various functions, including logical reasoning, rapid learning, and thinking, and studies examining the structural and functional network of the brain have indicated the connection between these networks and intelligence. For instance, van den Heuvel et al. (2009), Langer et al. (2012), and Hilger et al. (2017) mentioned the relationship between intellect and the small-world criterion in the brain's intrinsic networks. These findings reveal the correlation of intellectual performance with the shortness of the feature path and the hub centrality value in the salience and integration network between the frontal and parietal regions. Also, Wu et al. (2013) showed that IQ is positively correlated with nodal features in the attention-related network and negatively correlated with nodal features in the default mode, emotions, and language systems. Although these findings show that general intelligence is deeply influenced by the functional integration of spatially distributed regions, they do not provide sufficient information about whether and how human IQ is related to the brain's modular architecture.

Another field of graph study concerns brain changes during life. The human brain undergoes many functional changes from birth to adulthood. Numerous studies on this subject have shown that local efficiency and the rich club coefficient increase until adulthood in healthy people and then decrease with age. At the same time, regardless of the initial post-birth years, global efficiency remains unchanged (Gao et al., 2011). Moreover, the reverse paths between short and long connections indicate the gradual change in the brain's functional network, which likewise causes behavioral and cognitive changes in the person.

There have also been studies on the relationship between working memory and brain network. Stanley et al. (2015) showed that local efficiency is less associated with better working memory and greater global efficiency is associated with improved performance in young people and deficiency in the elderly.

Another interesting subject is the examination of changes in the person's daily functions in natural environments, on which basis brain function is measured in routine environments. For example, Petruo et al. (2018) showed that mental fatigue is associated with topological changes in the brain such as a decrease in the small-world feature and global efficiency and functional changes in the frontoparietal network and connected areas in the thalamus and the striatum.

One of the most significant applications of this science is in the diagnosis of biomarkers and mental disorders. These criteria and examination are applied to many diseases, including epilepsy, Alzheimer's disease (AD), multiple sclerosis (MS), autism spectrum disorder (ASD), attention-deficit/hyperactivity disorder (ADHD), schizophrenia, Parkinson's disease, insomnia, major depression, obsessive-compulsive disorder (OCD), borderline personality disorder (BPD), and bipolar disorder, and in many, the biomarkers of connections in brain function can be easily identified.

5.8.10 *Machine Learning in Analysis of Resting-State fMRI (Rs-fMRI) Data*

Most applications of machine learning in the analysis of rs-fMRI data are related to unsupervised learning approaches. Contrary to task-driven studies, modeling resting-state activity is not straightforward since there is no controlled stimulus to justify and drive these fluctuations. Therefore, the analytical methods used to describe the spatiotemporal patterns observed in task-based fMRI are typically ill-suited for rs-fMRI (Khosla, et al., 2019).

Unsupervised Learning

Given the numerous dimensions of fMRI data, it is no surprise that the primary analytical approaches are based on decomposition or clustering techniques to provide better descriptions of spatial and temporal data. Unsupervised learning approaches such as ICA accelerated the discovery of so-called resting-state networks or RSNs. It was also developed in resting-state brain mapping with the main goal of creating brain divisions, that is, the optimal grouping of voxels that functionally define coherent spatial sections in the brain. These maps help to better understand the brain's networks and connections. In addition, they serve as a feature reduction technique for statistical analysis or supervised machine learning. The literature has indicated that functional connectivity in the resting state undergoes significant changes during a typical rs-fMRI scan, which interestingly shows the dynamics of the brain network. Unsupervised learning techniques demonstrated that resting brain network patterns change between multiple states, indicating a diversity of mental processes (Khosla, et al., 2019).

The unsupervised learning approach employs k-means, Gaussian mixture models, hierarchical clustering, and graph-based clustering methods, as well as latent variable models, decomposition methods, independent component analysis (ICA), PCA, and hidden Markov models.

Applications of unsupervised algorithms in rs-fMRI data include the following:

Most unsupervised learning approaches in rs-fMRI aim to segment the brain into discrete functional subunits akin to atlases. Unlike approaches that use atlases, these segmentations are driven by functional data.

The second set of these applications is an exploration of the dynamics of brain networks. Recently, unsupervised learning has been applied to the analysis of the dynamic functional connectome with promising results.

Meanwhile, as mentioned earlier, one application of machine learning in rs-fMRI is in the clinical field. These supervised machine-learning applications are used for personal-level predictions. As a sensitive biomarker in disorders, many researchers have developed an interest in "Connectome," and many studies have further suggested that like fingerprints, these connections and connectomes are unique to each individual. Given the importance of deep learning, several new neural network-based approaches for the analysis of rs-fMRI data have also been developed. Most of these approaches extract connectomic features for individual-level prediction. Deep learning algorithms (Tzourio-Mazoyer, et al., 2002) will be explored in more detail in the following chapter.

Supervised Learning

This approach to learning deals with problems that have features and predictions and labels with the aim of learning the mapping between the input and target. This allows the system to estimate heretofore unseen input data points. An example is the prediction of autism through rs-fMRI correlations. Since intrinsic FC reflects interactions between cognitively relevant functional networks, there is a hypothesis that systematic changes in resting state patterns could be associated with pathology or cognitive features. The promising diagnostic accuracy of supervised algorithms using rs-fMRI is strong evidence for this hypothesis. This approach also uses methods such as ridge regression, LASSO regression, elastic-net regression, logistic regression, SVM, random forest, and deep learning.

This approach is under intensive development and has been discussed in many studies. For instance, the study of brain development and aging is one application of using supervised machine learning on rs-fMRI data. In this context, Duesenbach et al. used RSFC to predict brain maturity based on the chronological age of adolescents. Hence, these activities can be introduced as a valuable tool for predicting healthy neurodevelopment. This method can also be used to identify the unusual neurodevelopmental changes associated with normal aging (Abraham, et al., 2017).

As mentioned, in this field, the machine learning method can also be used for identifying biomarkers in mental disorders. The biological basis of psychiatric disorders has been unclear, and their diagnosis is currently guided entirely by behavioral assessment. rs-fMRI has emerged as a powerful imaging method for biomarker extraction for the diagnosis of psychiatric disorders. Supervised learning algorithms using RSFC have exhibited promising results for classifying or predicting symptom severity in a variety of psychiatric disorders, including schizophrenia, depression, autism spectrum disorder, attention-deficit hyperactivity disorder, social anxiety disorder, post-traumatic stress disorder, and obsessive-compulsive disorder.

It has also received a tremendous amount of attention in the field of cognitive abilities and personality traits. Analysis of functional connectivity can predict individual differences in cognition and behavior. Due to its uncontrolled nature, resting-state imaging covers a wide range of inherent cognitive states. Currently, machine learning models have been used for predicting some individual characteristics such as fluid intelligence, sustained attention, memory performance, and language scores from RSFC-based biomarkers. Also, some studies have used these models to identify personality traits such as neuroticism, extroversion, agreeableness, and openness.

Another important application concerns studies on sleep and fluctuating levels of consciousness. Nevertheless, few studies have used machine learning to predict levels of consciousness during rs-fMRI scans and have classified levels of consciousness during rs-fMRI.

Many studies have also implemented this for investigating genetics and inheritance. Understanding the influence of genetics on brain structure and function has been a long-standing goal in neuroscience. Research in genetic and environmental RSFC is also ongoing in the framework of machine learning. For instance, Miranda-Dominguez et al. applied an SVM classifier to twin FC with the remarkable result of the feasibility of successfully predicting family relationships from resting-state fMRI and forming aspects of functional connectivity by unique genetic or environmental factors (Khosla, et al., 2019).

5.9 Designing Conceptual Networks

An essential first step to use a network model, or any model, involves the choice of which variables should be included in the model. The variables are called nodes in network analysis. In other words, network analysis starts with network design and network design begins with defining and describing variables (nodes) in networks (Bringmann et al., 2022).

To define the variables as nodes of a network, Bringmann et al. (2022) recommend the process of node validity, which involves two steps. The first is node selection and the second is node assessment.

Nodes can be any psychological variables, but clearly the theoretical or clinical assumptions of researchers play a key role in selecting a variable as a node in a network. That is, the chosen node should be minimally complete. It means that all nodes necessary to model the intended phenomena should be contained in the network, while excluding superfluous nodes. This feature will differ across contexts. The researcher's insight is of utmost importance. Nodes should be sufficiently distinct as well, especially if the researcher's assumptions are casual. To be sufficiently distinct, nodes should be

(a) Separately identifiable (i.e., at least in theory, they can be assessed independently of one another).
(b) Independently manipulable (i.e., at least in theory, one should be able to intervene on a node without intervening on other nodes).

More importantly, the validity and reliability of selected nodes should be assessed.

It should be borne in mind that network analysis in this book is cross-sectional, and most of network analysis in psychology is of this type (Robinaugh et al., 2020).

The main aim of these networks is to study between-person differences, and sometimes with the aim of generating hypotheses on within-person dynamics. We know that cross-sectional analyses do not statistically separate between-person variability (for example, typically stable trait-like features) from within-person variability (for example, state-like features). Thus, analyses based on cross-sectional data may result in networks with edges that reflect a mix of between and within effects. Cross-sectional data can usually be directly used to identify within-person

dynamics if stationarity is yielded. This specific circumstance is called ergodicity (Molenaar, 2004). It requires that individuals are independent and that the same data-generating process applies to all individuals (homogeneity). Moreover, the statistical characteristics of the data, like means, variances, and auto- and cross-validation, do not change over time. In other words, no trends could be detected in the data.

We know that in many cases these assumptions are not met and therefore network analysis based on time-series data is carried out. This analysis is called temporal network analysis, which will not be discussed in this volume.

In sum, network science provides tools, by which the difference between people in the structure of different networks becomes evident. These findings lead to the formation of an assumption among artificial psychology researchers: network processes are dynamic by nature.

Network science is quite useful in cognitive psychology particularly and in psychological sciences generally in at least three aspects. Siew et al. (2019) summarize these three aspects:

1. Network science provides a quantitative approach to represent cognitive systems. Network science provides an approach to model several cognitive and psychological networks, such as language, semantic memory, traits, and linguistic environment. It enables micro-meso and macroscopic network analysis, leading to a new understanding of the structures of psychological networks.
2. Network science facilitates a deeper understanding of human cognition so that researchers consider how network structure and the processes operating on the network structure interact to produce behavioral phenomena.
3. Network science provides a framework to model structural changes in cognitive systems on multiple scales. Network science enables the studying of the development of cognitive and psychological systems. Thus, researchers might gain a deeper understanding of the early and late stages of human life and study the structural changes across time. Network science helps to quantify the structural and dynamic changes of cognitive systems.

5.10 Sample Size in Network Analysis

Advances in the accurate estimation of psychological network structures have been considered in psychological advances (William & Rast, 2020). One crucial factor to accurately estimate network parameters is the sample size (Epskamp et al., 2018). In psychological studies with a limited number of participants, and the convenience sampling method, the sample size is of great importance (Shen et al., 2011), as the number of parameters to be estimated would be increased if nodes are increased (Ryan et al., 2022).

Ryan et al. (2022) introduced a new method, based on Monte Carlo method, to determine the sample size in network analysis. They have used R software, similar to G*power software, for sample size recommendations. Examples are provided in this chapter. In this context, by utilizing necessary factors, consisting of the number of nodes, assumed density, statistic value (similar to statistic power), and measure value, one can estimate the sample size necessary for network analysis. The R package named powerly is used here.

5.11 Moderated Psychological Network Analysis

It should be remembered that an artificial psychologist is a psychologist who deepens their analyses with AI-based methods. In psychological studies, the comparison of groups and sub-groups is significant. In other words, moderating variables are important. Many researchers are interested in studying the structure of a network to see whether nodes and edges and centrality indices, depending on the levels of moderating variable, change or not. They might also ask if cognitive parameters differ in a psychological network for men or women. Answering these questions might be of interest to cognitive and evolutionary psychologists. After estimating networks, one can compare psychological networks. These comparisons open new doors to artificial psychology. Such analysis for cross-sectional data is called moderated psychological network analysis.

In R software, there are libraries named networktoolbox and bootnet, which enable this type of analysis. An example is provided below.

5.11.1 Practical Example Using R

Example 5.1
An artificial psychologist first tries to define three networks. This method can be considered a confirmatory mode of network analysis. We call it an ad hoc network analysis. This definition is based on theoretical foundations. He /she defines three networks:

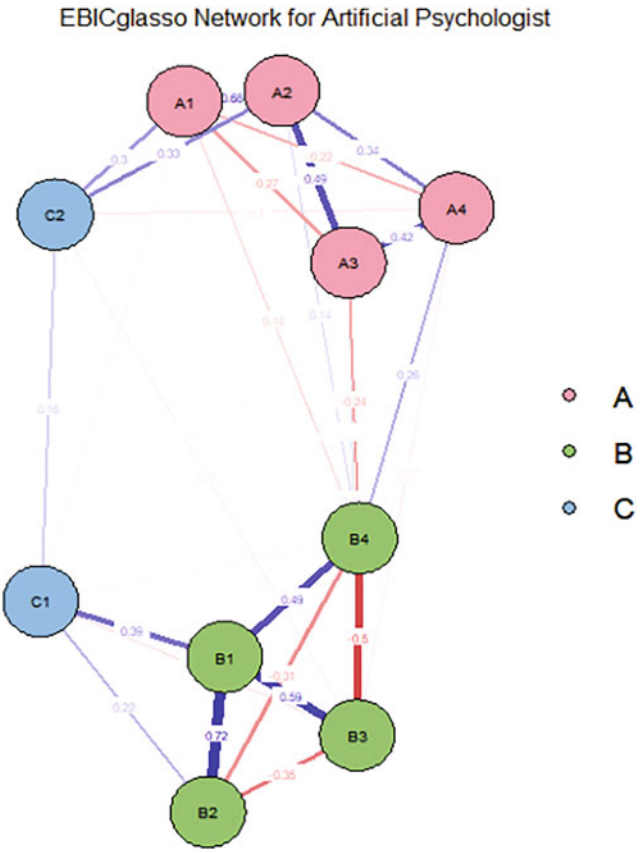
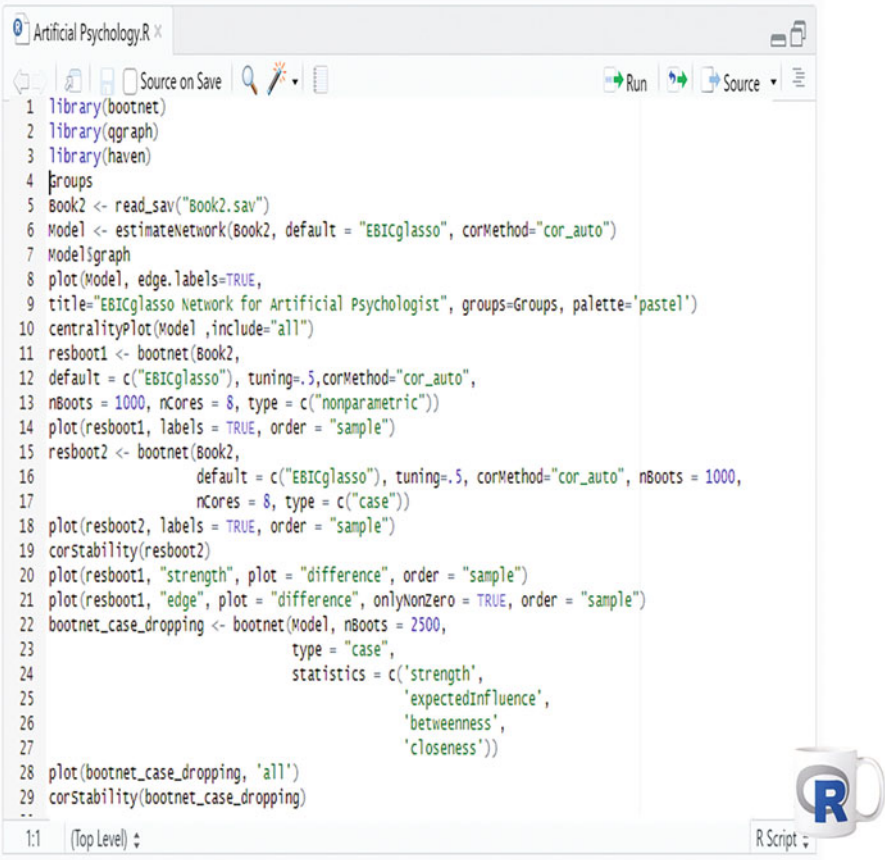


Fig. 5.7 The network plot of three networks (A, B, and C)

Network A includes Nodes A1, A2, A3, and A4, which are shown in pink in Fig. 5.7 Network B includes Nodes B1, B2, B3, and B4, and Network C includes Nodes C1 and C2. Network B in the output plot from the R software is marked in green color and network C is marked in blue. R Codes for the ad hoc network analysis are in Listing 5.1.

Listing 5.1 R codes for network analysis (ad hoc network analysis)



In network analysis diagrams, blue edges show a positive relationship and red edges show a negative relationship, and the thickness of each edge shows its strength.

The result of the analysis of this network in Fig. 5.8 showed that the highest strength is related to nodes B1 and B2, which have the most effect in the network and show that B4 is the bridge node between network B and network A, and also C1 is the node that links network B and C, and C2 is the node that connects network C and A.

In Fig. 5.9, the strongest positive connection in network A is between A1 and A2 and the strongest negative connection in network B is between B3 and B4 and the strongest positive relationship can be seen in that network between B1 and B2.

The plot in Fig. 5.9 also checks network stability. This graph shows the results of the samples obtained by the bootstrapping method from 1000 samples with the results of the main sample used in the network analysis and shows that there is stability of the edges. The path of red dots (resulting from the original sample) and

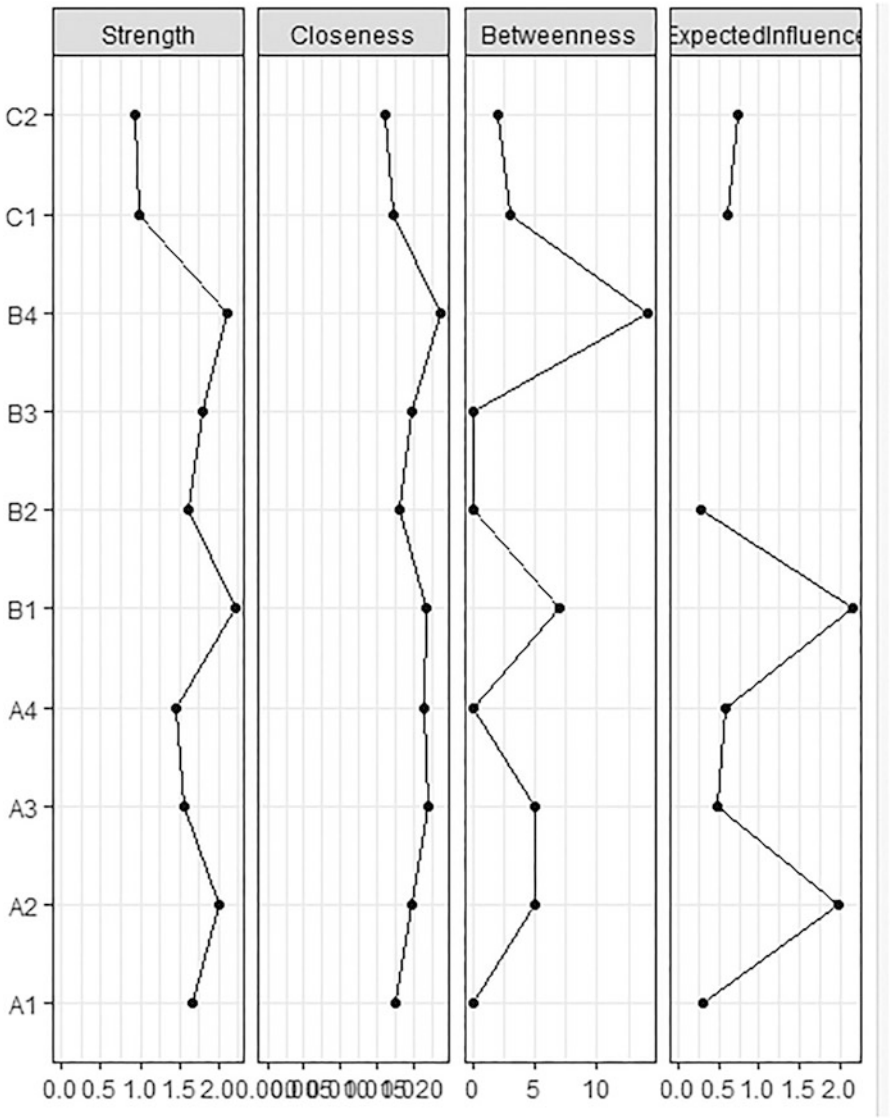


Fig. 5.8 The network plot for centrality indices

black dots, which are the average edges between nodes in the bootstrapping method, corresponds to a great extent, and the gray area around the two lines of the graph indicates the 95% confidence interval resulting from bootstrapping. As you can see, some edge weights, particularly those in the middle, which are smaller in absolute value, are more accurate than other edges. Also, this diagram shows how most of the edges are close to zero; that is, most of them intersect the 0 (zero) point in the bootstrapping samples.

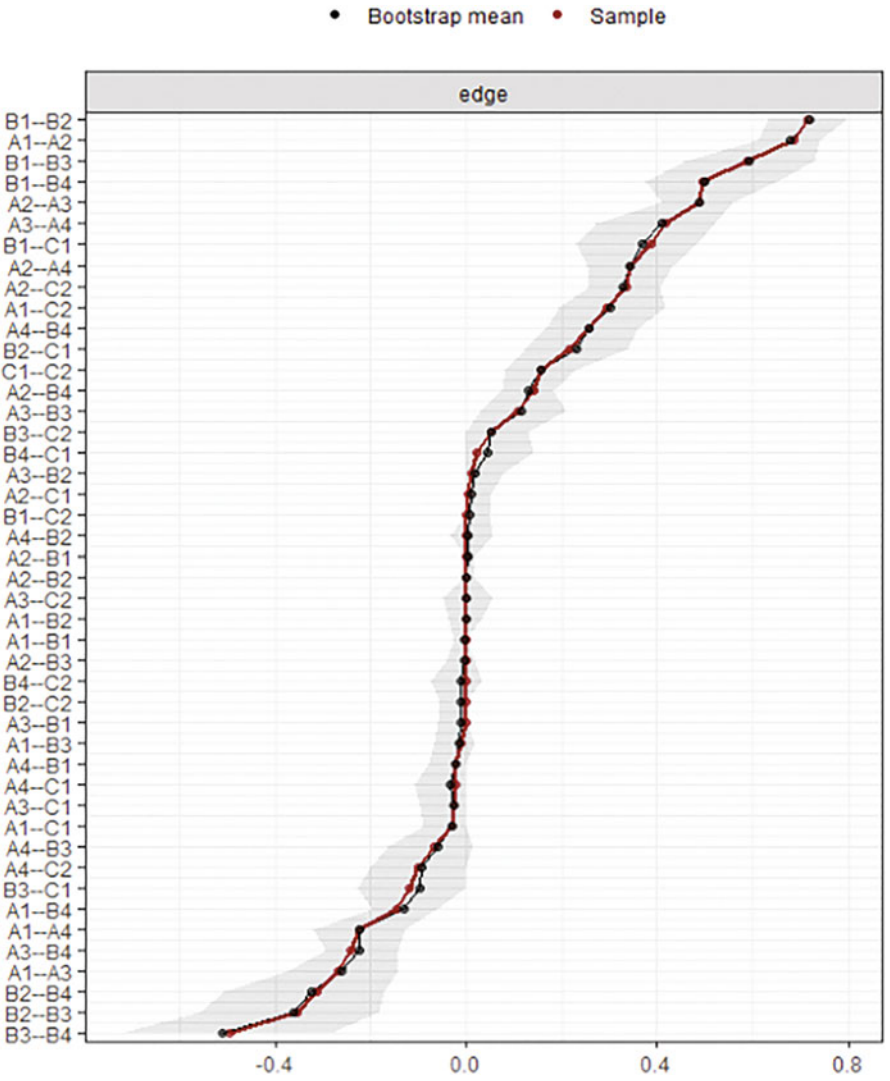


Fig. 5.9 The network stability

Figure 5.10 shows the average correlation of the original sample with 1000 bootstrap samples between each of the four centrality indices. This graph shows the robustness of these correlations to reduced sample size by looking to see whether the average correlation between the four centrality indices will be the same and will remain constant when dropping cases. This plot suggests the average correlation is robust to the sample size.

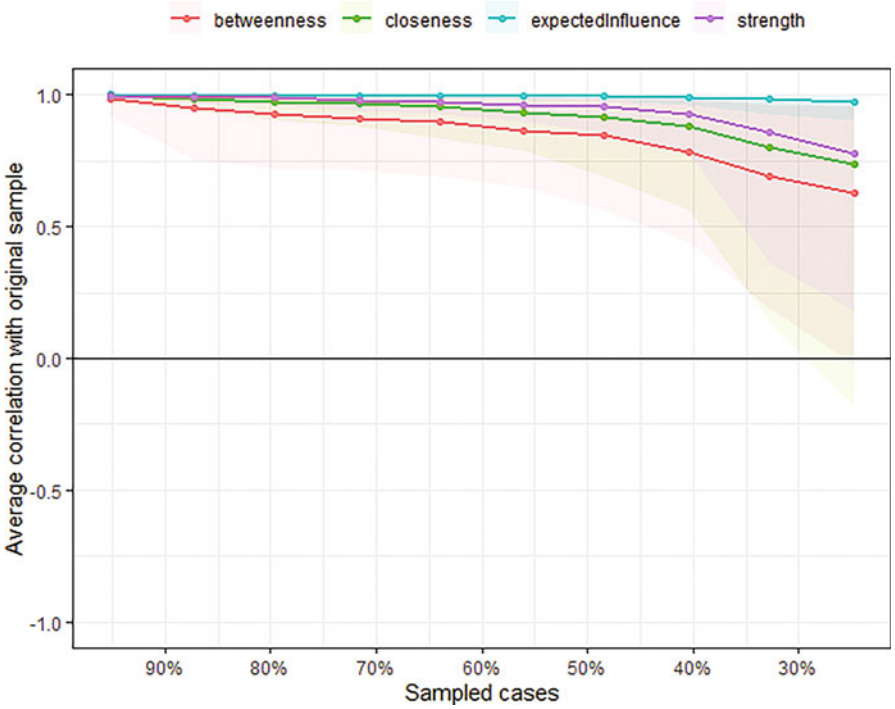


Fig. 5.10 R results for the average correlation between each of the four centrality indices and the correlation among nodes (a red square signifies a negative node and a blue one shows positive edge)

Figure 5.11 shows the negative or positive connection of edges between the nodes.

To avoid subjective error in viewing the graph by eye, Epskanp et al. (2018) invented a statistical index called a correlation-stability coefficient (c-s), which was explained in the previous sections of this chapter.

This index shows the average percentage of the sample that can be dropped to maintain a specified correlation of, for example, $r = 0.7$ between the central indices of the sample and the central indices obtained from Case-dropped bootstraps. $r = 0.7$ is not a fixed number and can be changed. If by dropping a large number of cases from the main sample, the correlation between the centrality indices is still high, it can be said that the network nodes are not affected by the characteristics of the sample and it is believed that they will show stability in the population.

This coefficient should not be below 0.25 and preferably above 0.5 (Epskanp et al., 2018). It can be said that the result of this analysis indicates a favorable interpretation of centrality indicators, especially for expected influence and strength.

Fig. 5.12 shows the adjacency matrix for each node in the output of R software. This figure represents the covariance of the node, which is considered to be the same as the adjacency matrix (Fig. 5.13).

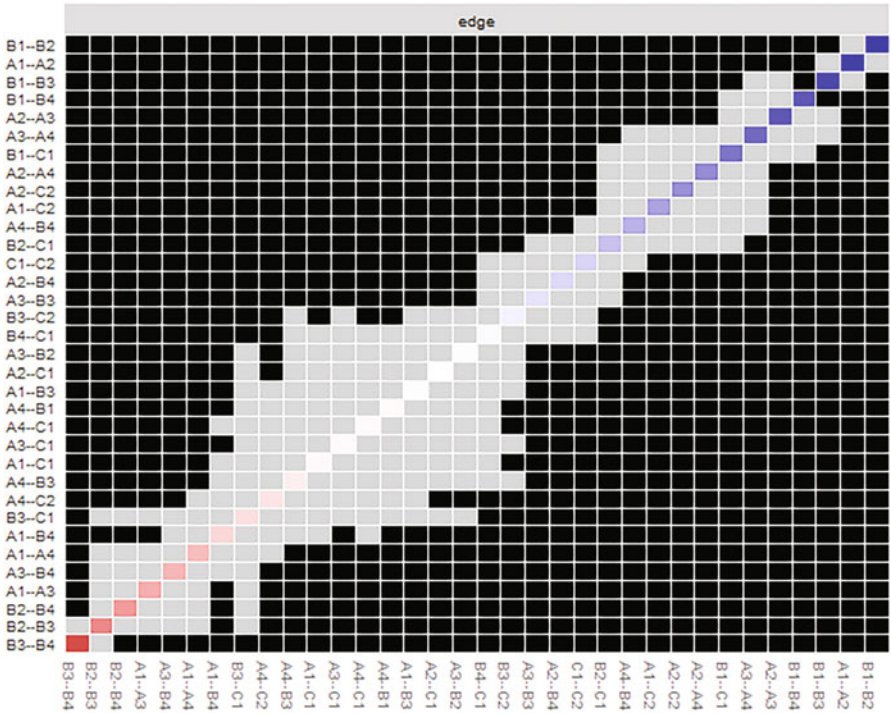


Fig. 5.11 The correlation among nodes (a red square signifies a negative node and a blue one shows a positive edge)

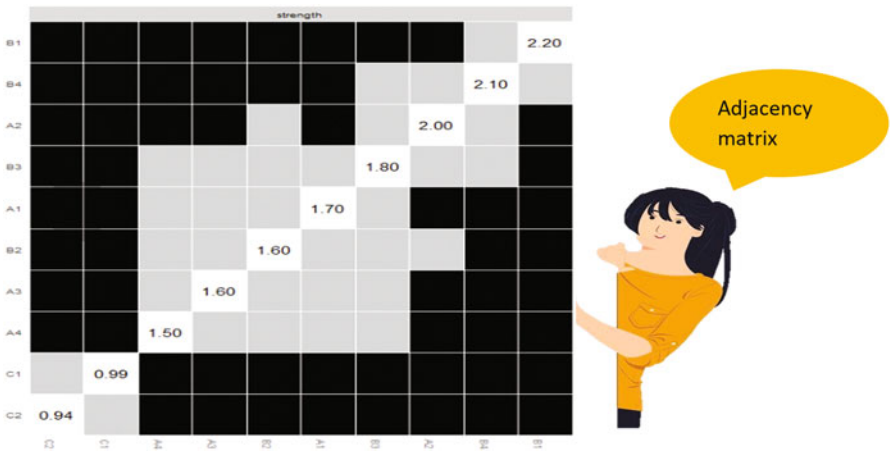


Fig. 5.12 R results for adjacency matrix

```
=== Correlation Stability Analysis ===

Sampling levels tested:
   nPerson Drop%   n
1       62  75.2  98
2       82  67.2 110
3      101  59.6 116
4      121  51.6  90
5      140  44.0 101
6      160  36.0  91
7      179  28.4  95
8      199  20.4  97
9      218  12.8 103
10     238   4.8  99

Maximum drop proportions to retain correlation of 0.7 in at least 95% of the samples:

edge: 0.752 (CS-coefficient is highest level tested)
  - For more accuracy, run bootnet(..., caseMin = 0.672, caseMax = 1)

strength: 0.596
  - For more accuracy, run bootnet(..., caseMin = 0.516, caseMax = 0.672)
```

Fig. 5.13 R results for the correlation stability

```
> perm.1$result
     part1 part2 p-value
c1    0.85  0.78   0.718
c2    0.67  0.85   0.413
c3    0.72  0.83   0.532
c4    1.04  1.07   0.914
c5    1.13  1.23   0.566
c6    0.63  0.55   0.675
c7    1.07  0.90   0.363
c8    1.47  1.22   0.263
c9    0.81  1.01   0.357
c10   0.62  0.91   0.273
> perm.hyb$result
     part1 part2 p-value
c1   0.667 0.375   0.328
c2   0.306 0.500   0.510
c3   0.708 0.389   0.253
c4   0.625 0.625   1.000
c5   0.764 0.944   0.349
c6   0.375 0.444   0.769
c7   0.792 0.667   0.709
c8   0.556 0.389   0.513
c9   0.278 0.597   0.258
c10  0.333 0.708   0.173
```

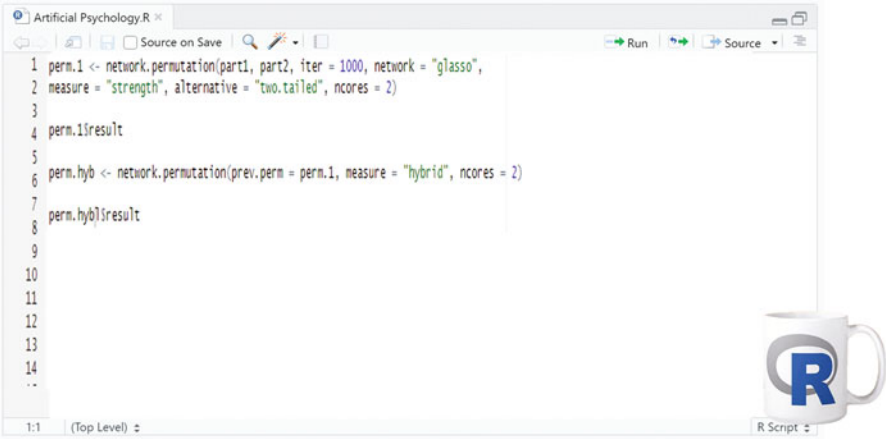
Fig. 5.14 R results for the permutation test

The correlation shown by the correlation-stability coefficient (c-s) is higher than 0.5 in the case of the two indicators of closeness and strength, and only the case of betweenness is below 0.5. Among these indices, the most important correlation-stability coefficient (c-s) is the one for strength (Fig. 5.14).

A permutation test is used to find out if there is a difference between measures obtained from a global network and one using centrality indices. For this purpose, two samples are needed. In this example, the investigated sample ($n = 250$) is randomly divided into two parts. It is better to choose two different samples from the

same community. The artificial psychologist then decides to compare the most important index under investigation, which is strength, in two groups using 1000 bootstrap iterations. The p-value indicates the difference in this index between the two parts of the sample. The results show that there is no significant difference between the measures of strength in the two parts of the sample. In this example, the investigated network has 10 nodes named from C1 to C10 (Listing 5.2).

Listing 5.2 R code for permutation test



Example 5.2
An artificial psychologist decides to use network analysis to examine the between different nodes. She examines many cognitive variables. For this purpose, she measures the components of attention, cognitive control, and cognitive regulation of emotion in adolescents. She measures cognitive variables including attention and response speed. Table 5.1 shows their abbreviations and general descriptions.

Table 5.1 Abbreviation and general description of the variables

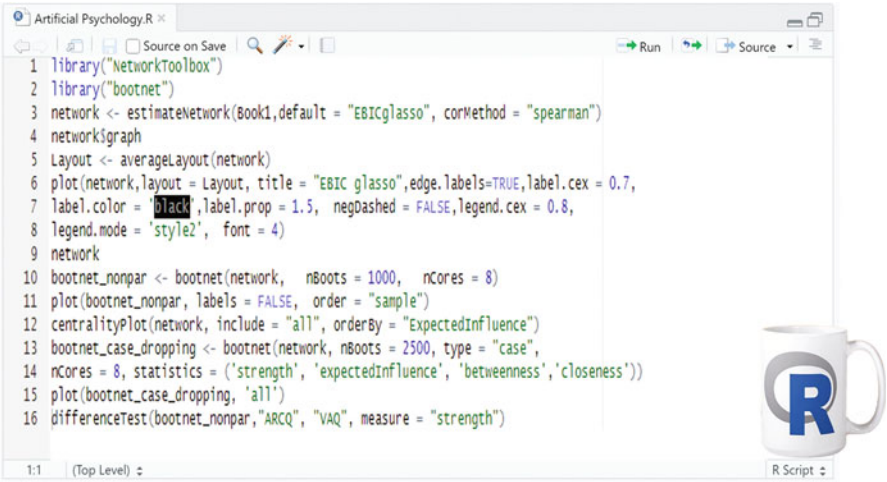
Abbreviation	Description
AAQ	Combination of vigilance, focus, and speed (auditory)
VAQ	Combination of vigilance, focus, and speed (visual)
QVIA	Vigilance Measure of inattention as evidenced by two different types of errors of omission (auditory)
QVIV	Vigilance Measure of inattention as evidenced by two different types of errors of omission (visual)
QFOCA	Total variability of mental processing speed for all correct responses during the test (auditory)

(continued)

Table 5.1 (continued)

Abbreviation	Description
QFOCV	Total variability of mental processing speed for all correct responses during the test (visual)
QMNA	Average reaction time for all correct responses (auditory)
QMNV	Average reaction time for all correct responses (visual)
SAAQ	Global measure of ability to respond to stimuli under low demand conditions accurately, quickly, and reliably (auditory)
SVAQ	Global measure of ability to respond to stimuli under low demand conditions accurately, quickly, and reliably (visual)
ARCQ	Combination of prudence, consistency, and stamina (auditory)
VRCQ	Combination of prudence, consistency, and stamina (visual)
QPRA	Measure of impulsivity (auditory)
QPRV	Measure of impulsivity (visual)
QCONA	Measure of ability to stay on task and sustain a reliable effort (auditory)
QCONV	Measure of ability to stay on task and sustain a reliable effort (visual)

Listing 5.3 R code for post hoc network analysis



In R software, there are libraries named networktoolbox and bootnet, which enable this type of analysis (Listing 5.3). We call this exploratory network analysis as post hoc network analysis.

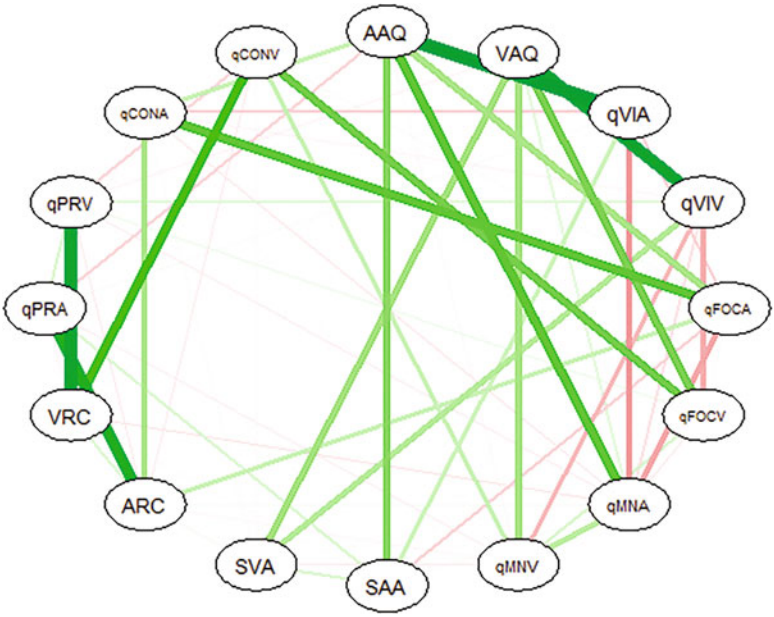


Fig. 5.15 The relationships of the nodes

She then estimates the network using the EBicglasso algorithm, which uses the LASSO method to produce a parsimonious network (Fig. 5.15). Figure 5.16 shows the relationship of each node with other nodes. The centrality indices are shown in Fig. 5.17. The three columns show the Closeness, Betweenness and Expected influence of each nodes in the network. As seen in this figure, VAQ and AAQ have the highest strength in the network. VAQ and AAQ also have relatively large numbers of thicker edges linking them to other nodes. The lowest strength is related to qPRA.

This researcher who is interested in artificial psychology decides to compare two networks based on the sex of teenagers. It is, therefore, necessary to design a network for male teenagers and a separate network for female teenagers. Then, centrality indices of the boys' and girls' networks are able to be compared using the compare centrality command in the R code in Listings 5.4 and 5.5. As can be seen, there is a difference between the girls (dashed line graph) and the boys (solid line graph) (Fig. 5.18).

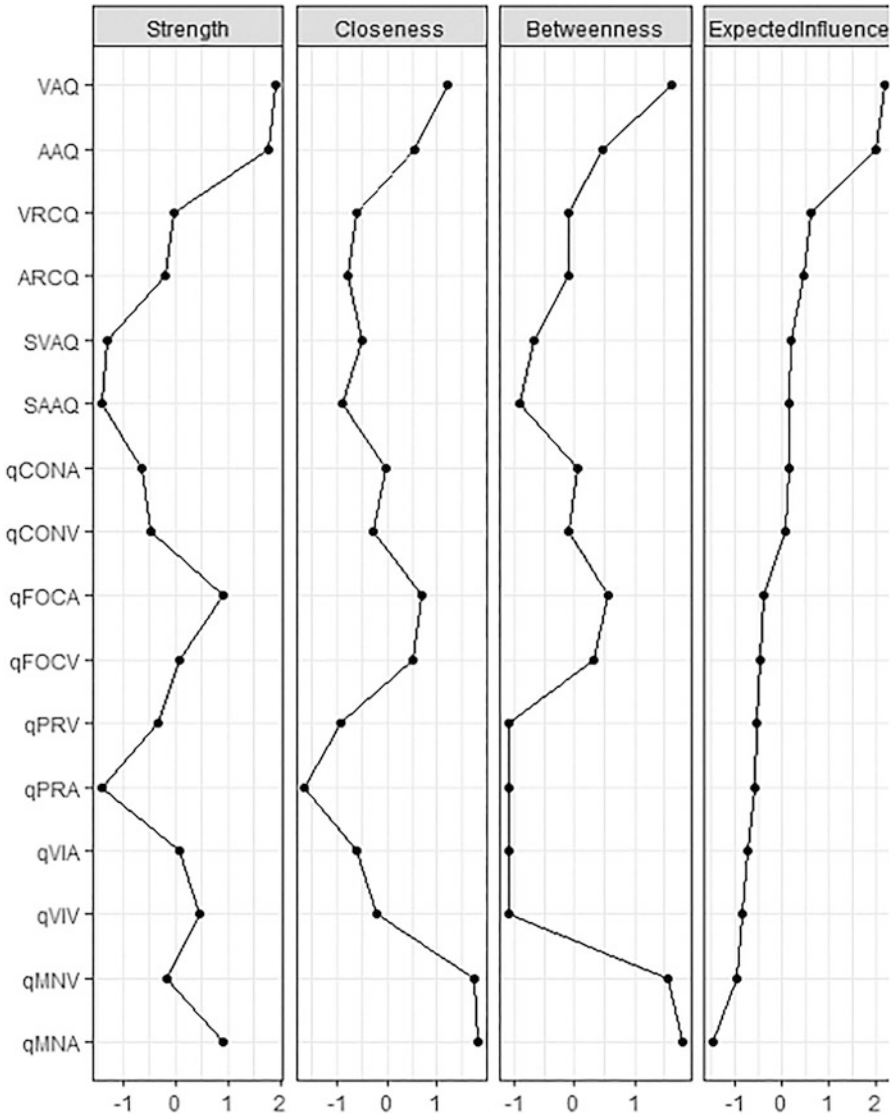


Fig. 5.16 The centrality indices for each node

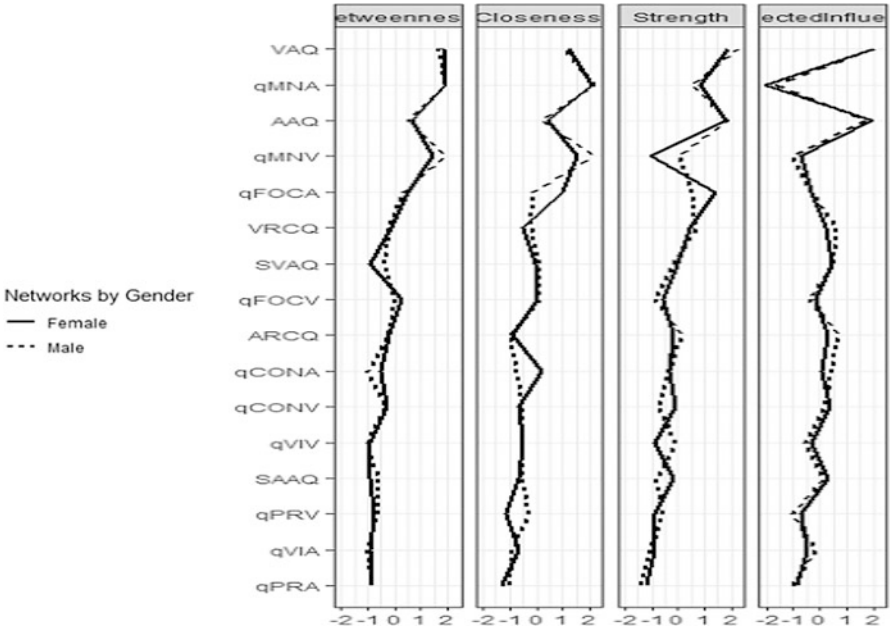


Fig. 5.17 R results for networks categorized by gender

Listing 5.4 R code for comparing two network analysis

Artificial Psychology.R

Source on Save | Run | Source

```
1  book$gender %<>% factor(levels = 1:2, labels = c("Male", "Female"))
2  network_male <- estimateNetwork(Y %>%
3  filter(gender == "Male") %>% select(-gender),
4   default = "EBICglasso",corMethod = "spearman")
5  network_female <- estimateNetwork(Y %>%
6   filter(gender == "Female") %>%
7   select(-gender), default = "EBICglasso",
8    corMethod = "spearman")
9  compareCentrality <- function(net1, net2, include = c("Strength",
10  "Closeness", "Betweenness",
11  "ExpectedInfluence", "all", "All"), orderBy = c("Strength",
12  "Closeness", "Betweenness","ExpectedInfluence"),
13  decreasing = T,legendName = '',net1Name = 'Network 1',
14  net2Name = 'Network 2'){
15   library(ggplot2)
16   library(forcats)
17
18   if(include == "All" | include == "all"){include = c("Strength", "Closeness",
19  "Betweenness", "ExpectedInfluence")}
20   df <- centralityTable(net1, net2) %>% filter(measure %in% include)
21     df %>%
22  mutate(graph = case_when(graph == 'graph 1' ~ net1Name,
23  graph == 'graph 2' ~ net2Name),
24  graph = as.factor(graph),
25  node = as.factor(node)) %>%
26
27  mutate(node = fct_reorder(node, value)) %>%
28
29   ggplot(aes(x = node, y = value, group = graph)) +
30
31  geom_line(aes(linetype = graph), size = 1) +
```

1:1 (Top Level) R Script

Listing 5.5 (Continued) R codes for comparing two network analysis

```
#continued codes
geom_line(aes(linetype = graph), size = 1) +

 labs(x = '', y = '') +

scale_linetype_discrete(name = legendName) +

 coord_flip() +

facet_grid(~measure) +

 theme_bw()}
compareCentrality(network_male, network_female,
                  include = "all",
                  legendName = "Networks by Gender",
                  net1Name = "Male",
                  net2Name = "Female")
```

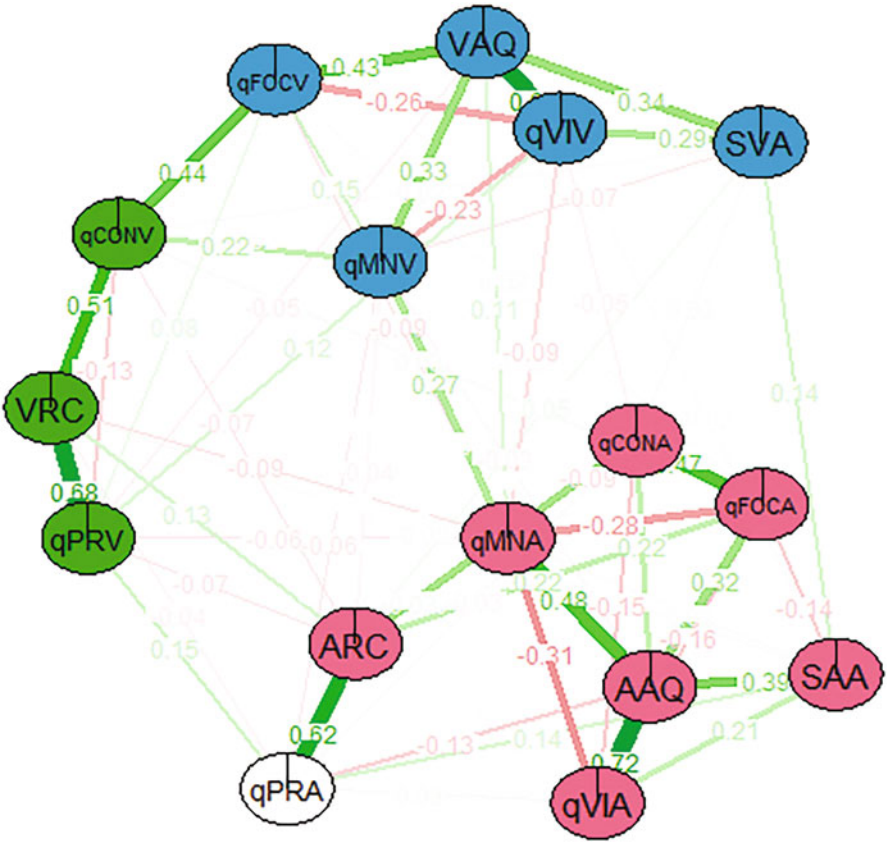
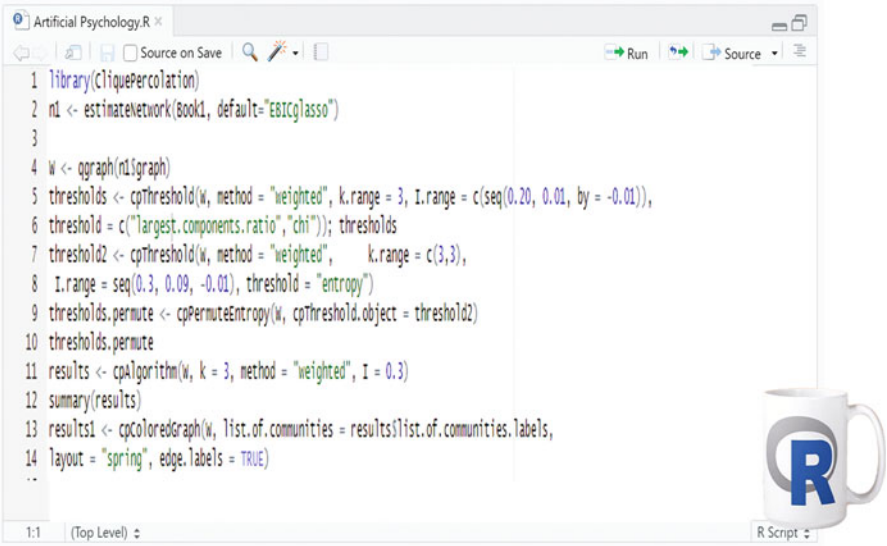


Fig. 5.18 R results for networks analysis for discovering the communities

One of the interesting features of network analysis is finding communities based on the relationship between nodes. This is an exploratory method. As the output of the software shows, three communities have been extracted in which there are no shared nodes, which means that a number of communities are created that are independent of each other (Listing 5.6).

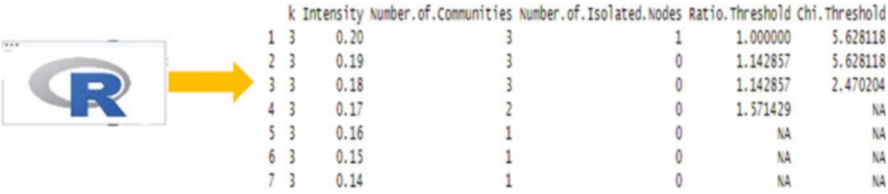
Listing 5.6 R codes for discovering communities using network analysis



Community1: AAQ, qVSI, qFOCA,qMNA, SAA, ARC, qCONA.
Community2: VCR, qPRV, qCONV.
Community3: VAQ, qVIV,qFOCA, qMNV, SVA.

In this exploratory analysis, which extracts communities based on the connection of nodes, there is only one separate node from the rest, qPRA (Fig. 5.19).

In the network diagram, community 1 is marked in blue, community 2 is marked in green, and community 3 is marked in pink, with a single isolated node, which is shown in white in the diagram (Fig. 5.20).



	k	Intensity	Number.of.Communities	Number.of.Isolated.Nodes	Ratio.Threshold	Chi.Threshold
1	3	0.20	3	1	1.000000	5.628118
2	3	0.19	3	0	1.142857	5.628118
3	3	0.18	3	0	1.142857	2.470204
4	3	0.17	2	0	1.571429	NA
5	3	0.16	1	0	NA	NA
6	3	0.15	1	0	NA	NA
7	3	0.14	1	0	NA	NA

Fig. 5.19 The results of discovering the communities using the network analysis

```
--------------------
Communities (labels as identifiers of nodes)
--------------------

Community 1 : AAQ qVIA qFOCA qMNA SAA ARC qCONA
Community 2 : VRC qPRV qCONV
Community 3 : VAQ qVIV qFOCV qMNV SVA


--------------------
Shared nodes (labels as identifiers of nodes)
--------------------

NULL

--------------------
Isolated nodes (labels as identifiers of nodes)
--------------------

qPRA
```

Fig. 5.20 The results of discovering the communities ($K = 3$, $I = 0.2$)

To run the clique percolation algorithm for weighted networks, we initially need to optimize k and I. In order to do this, the cp threshold function can be used. By default, this function provides the communities, shared nodes, and isolated nodes with labels as identifiers of the nodes. As we can see, the results indicate that three communities have been discovered with qPRA as an isolated node. We run this analysis using k equal to 3 and I taking values of 0.20 to 0.01 in steps of 0.005. There is no shared node. It is also possible to use the numbers as identifiers of the nodes or to restrict the output. The range of I values was chosen based on the mean edge weight of the network, and it was set to 0.3 when generating the network. Thus, $I = 0.40$ should allow the artificial network to find a broad range of community sizes. However, Farkas et al. (2007) recommended starting the analysis by setting the highest tested value of I to the maximum edge weight in the network.

Example 5.3

An artificial psychologist is going to determine the sample size for implementing network analysis. Based on the initial network, He knows the network has 10 nodes and he supposes the density is 0.4. The network density implies the proportion of present edges in the network that impact the required sample size (Constantin et al., 2021). He includes the measure value equal to 0.6. This measure is similar to effect size. We use a statistic value of 0.8. This measure is similar to the statistical power. Listing 5.7 indicates the R code for determining the appropriate sample size to analyze a GGM network with 10 nodes and a density of 0.4.

Listing 5.7 R codes for determining the appropriate sample size to analyze a GGM network

```
library(powerly)
results <- powerly(
  range_lower = 150,
  range_upper = 800,
  samples = 30,
  replications = 50,
  measure = "sen",
  measure_value = 0.6,
  statistic = "power",
  statistic_value = 0.8,
  model = "ggm",
  nodes = 10,
  density = 0.4)
summary(results)
```

A higher density value implies that more of these pairwise connections will be present in the network. Figure 5.21 shows the results.

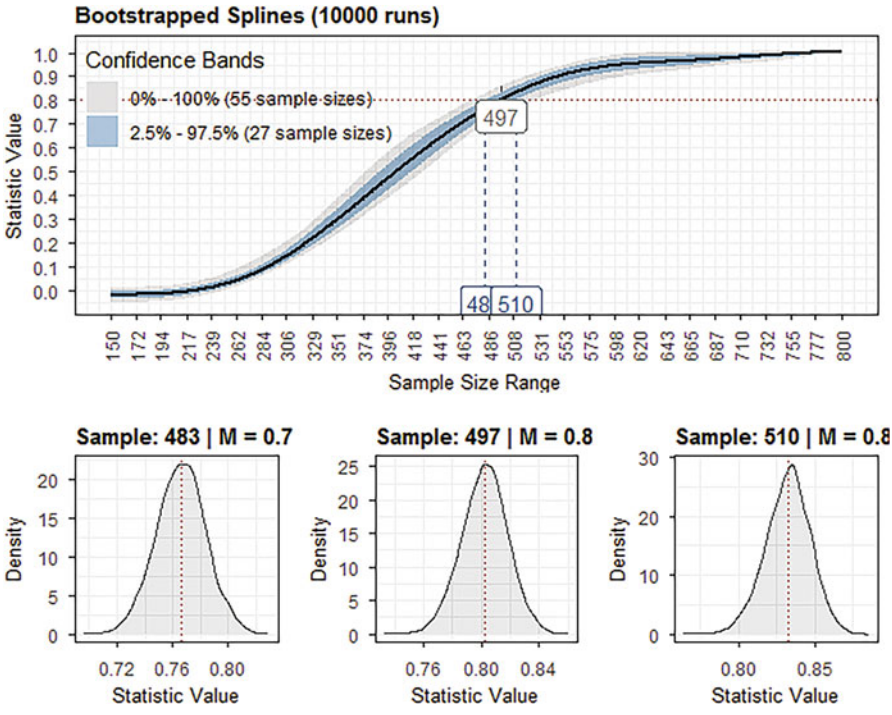


Fig. 5.21 The results of powerly package for determining the appropriate sample size

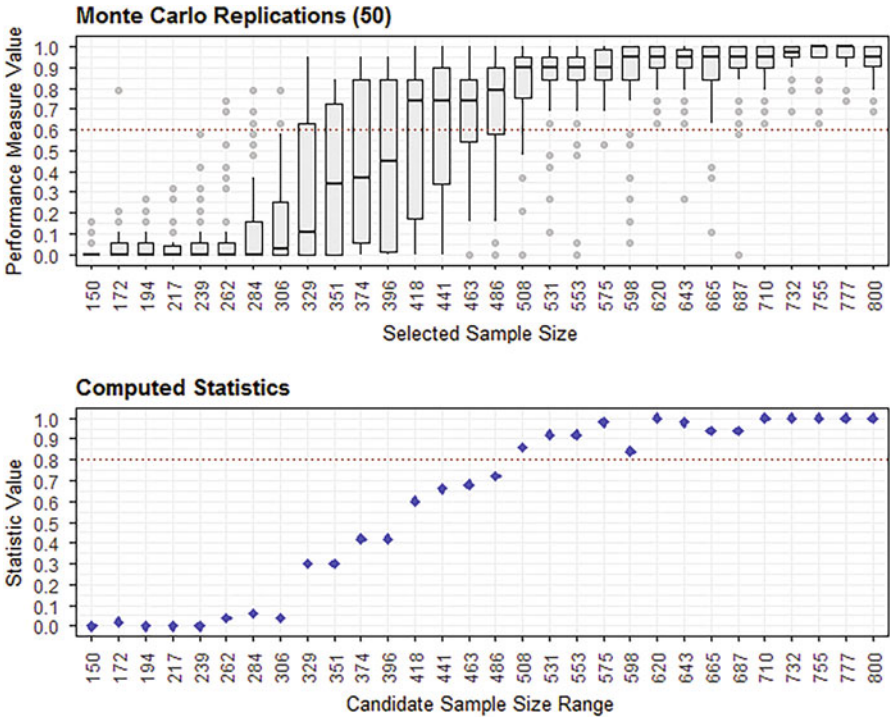


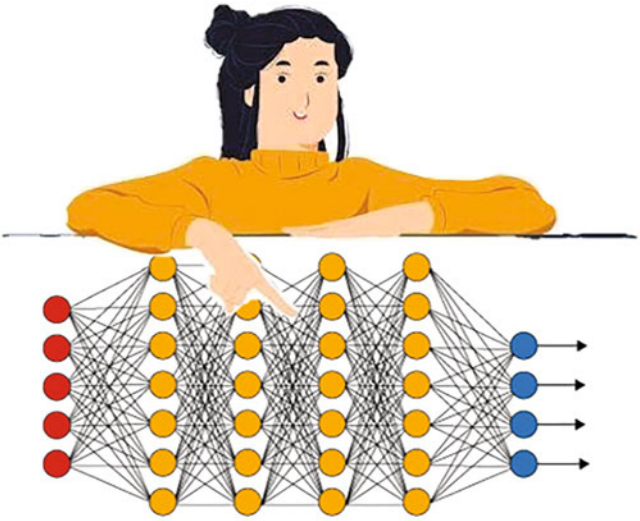
Fig. 5.22 The results of Box plot and Statistic value

The horizontal red dotted line illustrates the desired target for the statistic (0.8), and the vertical red dotted line shows the mean (0.8) of the bootstrapped statistics for the median sample size of 497. The horizontal blue dashed lines indicate the lower and upper bound sample sizes based on the 95% CI. The CI was constructed using the percentile method (Diciccio & Romano, 1988) based on 10,000 bootstraps.

Box plots provide a representation of how much the performance measure varies for each sample size based on the number of replications performed.

Fig. 5.22 shows the performance measure values obtained from the Monte Carlo replications (Box plot) and the curve line shows the values for the statistic computed on the performance measure values and the monotone spline used to interpolate the statistic across the entire candidate range of sample sizes. A sample size of about 500 is suggested based on both of the above plots.

Chapter 6
Deep Neural Network



6.1 Deep Neural Network (DNN)

In 1981, Teach and Shortliffe emphasized the importance of explaining decisions in decision-making systems. In recent years, XAE has grown in popularity. Explanation and interpretation aim to reduce the opacity of black box models and to improve the perception of reasoning in predictive and classification models. This reasoning precedes predictions. This makes it possible to promote the explanation and interpretation of transparency in AI/ML-based models; therefore, people will gain more confidence in such systems. First, artificial neural networks (ANNs) are analyzed in this chapter.

H. Farahani et al., *An Introduction to Artificial Psychology*,
https://doi.org/10.1007/978-3-031-31172-7_6

6.2 Neural Network

Neural networks or artificial neural networks) are a data processing system somewhat inspired by the human nervous system. They can also be considered an electrical model of the brain's nervous system. Like the learning mechanism, these models are also experiential. However, these models and problems employ different methods. This system uses smaller interconnected processors called neurons to create a network that is trained using training algorithms (Lawrence, 1993).

This technique uses the principles of the human brain and its structure to develop data processing strategies. The essence of the approach is reflected in the parallel processing of data. The procedure by which training is performed is the training algorithm. Through this process, the weights of the synapses are systematically changed in order to achieve the desired performance of the network.

Properties of neural network (Milan Milosavljević, 2019):

- Nonlinearity.
- Input–output mapping.
- Adaptability – the ability to change the strength of synaptic connections.
- Evidentiary response.
- Contextual information.
- Failure resistance.
- Possibility of realization in VLSI (very large scale integration) technology.
- Uniformity of analysis and synthesis.
- Neurobiological analogies.

6.3 Neuron

As mentioned earlier, the neural network in living organisms and the artificial neural network are both based on neurons. In their natural state, neurons have a biological structure comprised of dendrites, axons, and somaxfigure (Fig. 6.1).

Dendrites are the communication appendages of cells responsible for the communication and reception of signals from their surroundings, which are ultimately processed in the soma in the center of cells. The axon then sends the output signal to other neurons. Despite the greater complexity of the natural neuron model compared to this simple computational model, this model is the main basis of artificial neural networks (McCulloch & Pitts, 1943).

In the mathematical model of the artificial neural cell, the processing unit or neuron can be shown as follows (Fig. 6.2).

An artificial neuron is the basic element of a neural network. These neurons contain the following (Pamučar, 2010):

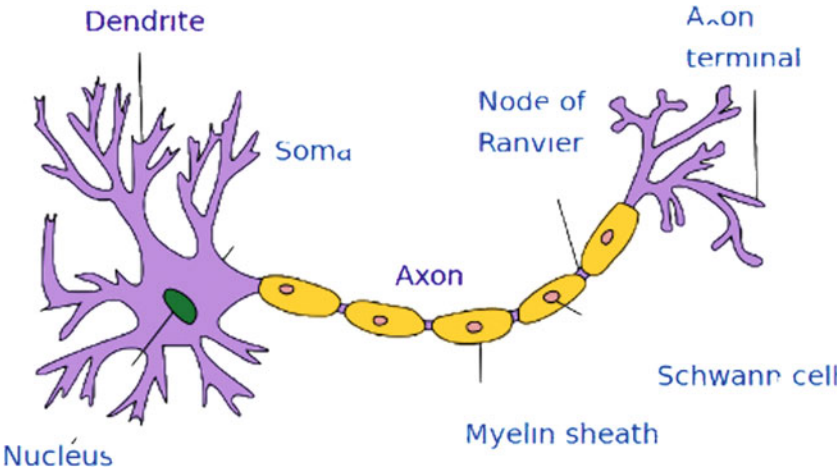


Fig. 6.1 Structure of neuron

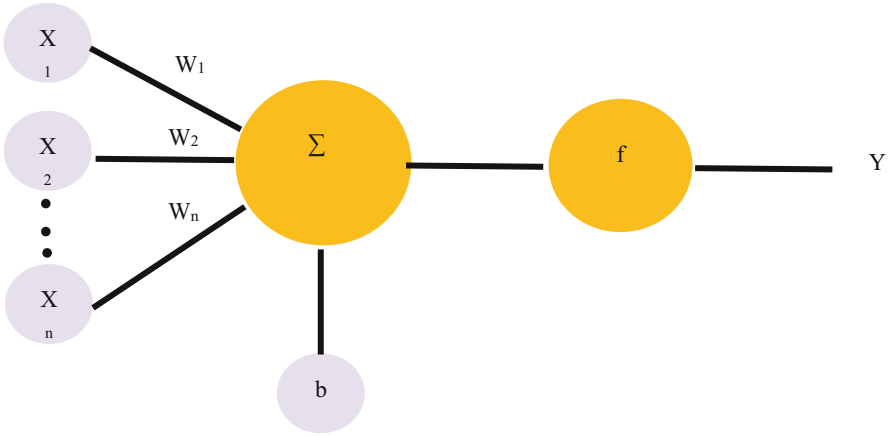


Fig. 6.2 An artificial neuron structure

- Inputs – x_i.
- Synapses (input weighting factor) – w_i.
- State of activation – z.
- Output function – f.
- One output – o.
- Threshold – T.

Here, the inputs acting as dendrites are considered the features of a problem. In these models, input signals are multiplied by weight values (each input has its own weight). Thus, all inputs can be considered as input vectors, and the corresponding weights are the weight vector. The weights are usually adjusted during the training phase. Another value that may also be added to this set, the bias parameter, is also

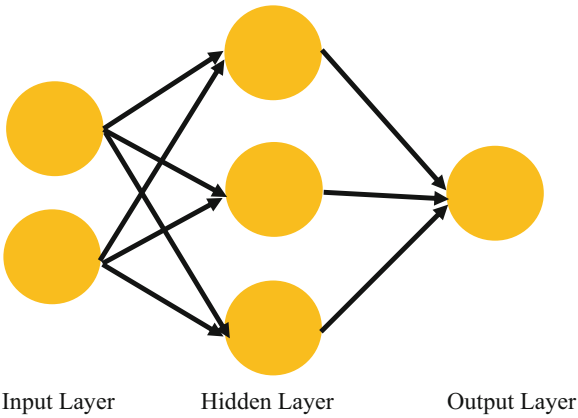


Fig. 6.3 A network of neurons

adjusted during training. Finally, the actual output is applied to the input based on the Φ activation function.

The simplest activation function is a binary function. If the total value of inputs exceeds the threshold value, the neuron spikes and the output function becomes 1, which is otherwise 0.

$$\phi(x) = \begin{cases} 1 & wx + b > \text{threshold} \\ 0 & O.W \end{cases} \tag{6.1}$$

A Network of Neurons: The neural network is considered to have a network of neurons. A row of neurons is considered a layer, of which there are three in the neural network: an input layer, an output layer, and a hidden layer. The network can have different structures, and layers and neurons might vary in numbers (Fig. 6.3).

The neural network consists of the following (Pamučar, 2010):

- The architecture (topology) of the network, i.e., the neuron connection scheme.
- Transmission functions of neurons.
- Laws of learning.

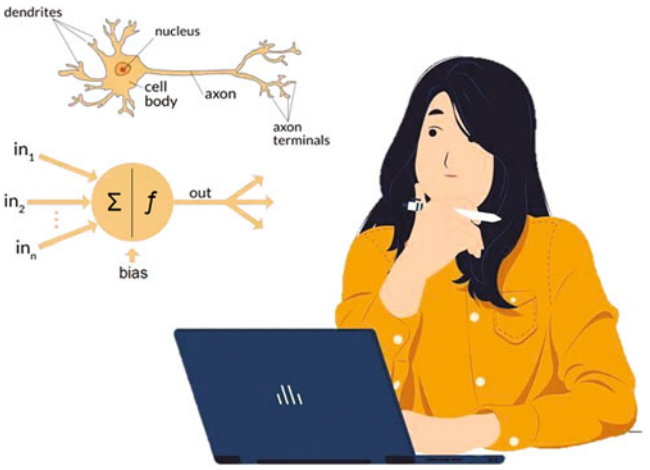
The architecture of an artificial neural network is represented by the specific arrangement and connection of neurons in the form of a network. By architecture, neural networks differ according to the number of neural layers. Usually, each layer receives inputs from the previous layer and sends its outputs to the next layer. The first layer is called input, the last is output, and the other layers are called hidden layers. One of the most common neural network architectures is a three-layer network. The first layer (input) is the only layer that receives signals from the environment. The first layer transmits signals to the next layer (hidden layer), which processes this data and extracts features and patterns from the received signals. Data that is considered important is sent to the output layer, the last layer of the network. The final processing results are obtained at the outputs of the neurons

of the third layer. More complex neural networks can have multiple hidden layers, feedback loops, and time delay elements.

Learning neural networks boils down to learning from examples, which should be as many as possible so that the network can behave more precisely.

6.4 Artificial Neural Network (ANN)

Some human inventions have been inspired by nature. In fact, nature is full of inspiration, and those who are more prepared than others will benefit from intuition.



A neural network is a conceptual model based on the human brain, which consists of nearly 10 billion neurons and 60 trillion synapses (Shepherd & Koch, 1990).

The human brain is considered to be a very complicated nonlinear parallel information processing system.

Therefore, scientists have tried to mimic the behavioral pattern of a neuron and design an artificial neuron called a perceptron. This chapter does not intend to review the history of the perceptron but aims to address its application to explainable artificial psychology presented in this book. The idea proposed by Pitts and McCulloch in 1943 was used as the basis for artificial neural networks (ANNs).

An ANN contains a large number of very simple interconnected processors called neurons. These artificial neurons serve as biological neurons. They receive inputs and provide outputs.

In ANNs, neurons are connected by links, each of which is given a numerical weight. These weights serve as long-term memory in ANNs. They represent the strength or importance of each input neuron. With ANNs, learning means nothing more than adjusting these weights in an iterative process.

Table 6.1 draws a brief comparison between a biological neural network and an artificial neural network.

Table 6.1 A brief comparison between a biological neural network and an artificial neural network

Biological neural network	ANNs
Soma	Neuron
Dendrite	Input
Axon	Output
Synapse	Weight

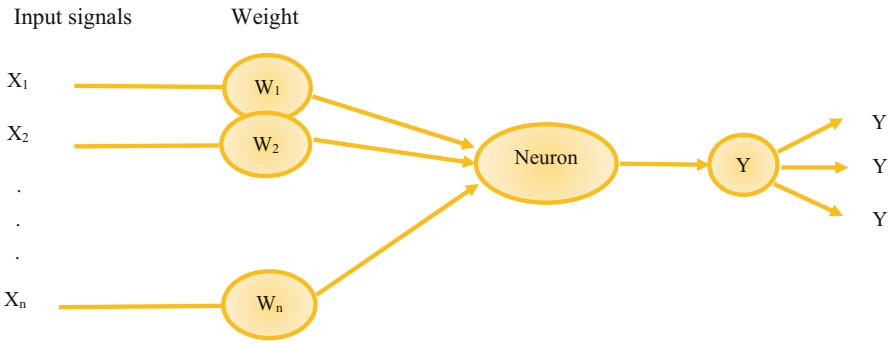


Fig. 6.4 An artificial neuron's elements

Figure 6.4 demonstrates an artificial neuron as a computing element of ANNs.

As already mentioned, the main idea of ANNs in 1943 was based on the seed presented by Pitts and McCulloch. They briefly pointed out that the weighted sum of the inputs would be calculated by a mathematical formula called neuron and its output would be compared to a threshold (θ). If the final value is less than the threshold, the neuron output is -1; however, it is $+1$ if the final value is greater than or equal to the threshold. There is an activation function in these sets, which is computed as follows:

$$Y = \text{Sign}\left[\sum_{i-1}^{n} \alpha_i w_i - \theta\right] \tag{6.2}$$

$$y = \begin{cases} +1 & \text{if} \& x \geq \theta \\ -1 & x < \theta \end{cases} \tag{6.3}$$

Nevertheless, the sign function is not the only activation or transfer function. There are several active functions, a limited number of which are practically applicable.

$$(n) = \begin{cases} 1, & n < 0 \\ 0, & n \geq 0 \end{cases} \tag{6.4}$$

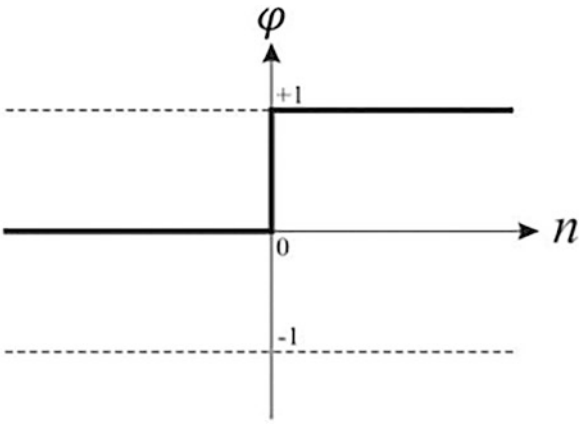


Fig. 6.5 Threshold function

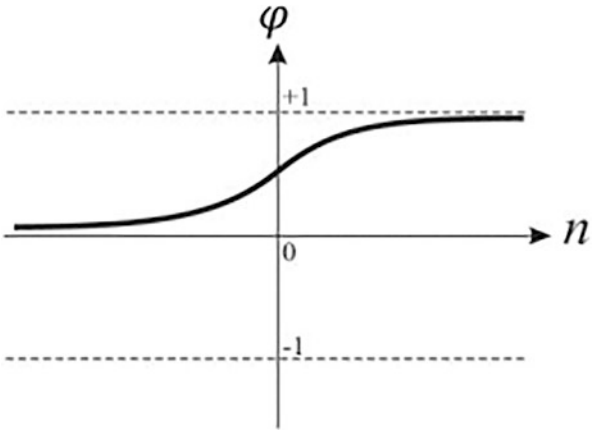


Fig. 6.6 Sigmoid function −1

This is a widely used function in an artificial neural network. The following are two examples of these functions, which are widely used since they are derivable, ascending, and continuous (Fig. 6.5).

$$\phi(n) = \text{logsig}(n) = \frac{1}{1 + e^{-n}} \tag{6.5}$$

$$\phi(n) = \text{tansig}(n) = \frac{2}{1 + e^{-2n}} - 1 \tag{6.6}$$

This function uses the simple $y = x$ equation for linear approximation. Its diagram is shown in Figs. 6.6, 6.7 and 6.8.

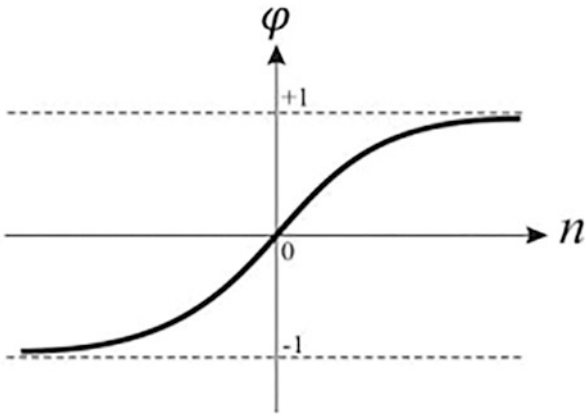


Fig. 6.7 Sigmoid function-2

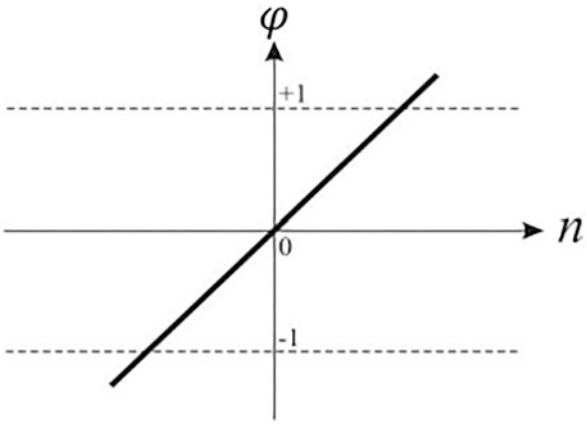


Fig. 6.8 Linear activity function

$$\phi(n) = \text{tribas}(n) = \begin{cases} 1 - |n|, & -1 \le n \le 1 \\ 0, & O.W \end{cases} \tag{6.7}$$

The level and sign functions are also known as hard limit functions, which are commonly used for pattern recognition and classification. In addition, the sigmoid function can take any input value from $-\infty$ to $+\infty$ and convert it into a value between 0 and 1. This activation function is typically used in backpropagation networks (Fig. 6.9).

In 1958, Frank Rosenblatt presented a training algorithm that can be considered the first training method for a simple ANN.

He named this simple single-neuron network the perceptron, which consisted of a single neuron with tunable synaptic weights and a hard limiter. This model was based on the idea of Pitts and McCulloch. Over time, this single neuron ANN evolved and multilayer neural networks (MLNNs) emerged in scientific fields.

Classification is very common in psychological studies, and this purpose fully justifies and proves the application of explainable AI through the use of MLNNs in the psychological sciences.

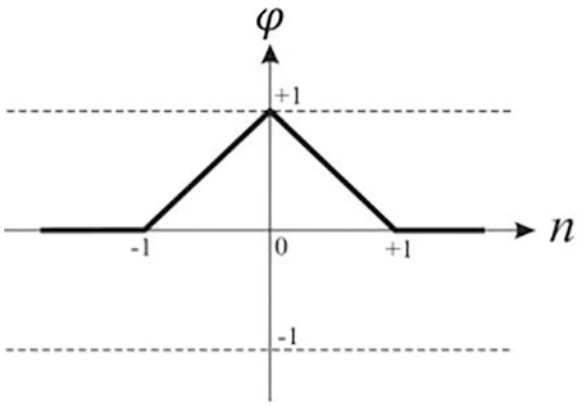


Fig. 6.9 Triangular activity function

The multilayer perceptron neural network (MLPNN) is an MLNN in which input signals (i.e., input data) propagate in the forward direction layer by layer. This model consists of an input layer, a middle or hidden layer, and an output layer.

In MLPNNs, each layer has its own specific function. The input signals from the outside world are fed into the input layer and then redistributed to all the neurons of the hidden layer. The input layer either lacks computational neurons or rarely has any; therefore, no processing takes place in this layer. The output layer receives the output signals from the hidden layer and creates the output pattern of the entire network.

The neurons of the hidden (middle) layer recognize some features and the weights of the neurons are considered as features for the hidden layer, which is called hidden because it hides the output of interest. In other words, the hidden layer determines its desired output. It is possible to increase the number of hidden layers. Deep learning is achieved when the network has more than one hidden layer. The number of neurons in each hidden layer can range from 10 to 1000 layers. Although it is possible to increase the number of hidden layers arbitrarily, this increases the computational load.

Learning is very important in MLNNs and deep learning. There are more than a hundred different algorithms for network learning. The most common algorithm is the backpropagation method. Learning in a multilayer network resembles a perceptron. The network calculates the output pattern. If there is an error (i.e., a difference between arbitrary and real output patterns), the weights are adjusted to reduce this error.

The goal of this learning algorithm is to minimize the sum of squared errors. A network is said to have converged if the sum of the squared errors in the learning sets or in the epoch is sufficiently small.

Furthermore, learning in an MLNN does not exactly emulate the behavior of a biological neuron, for these neurons do not have backpropagation mechanisms.

6.5 Types of Training

There are three types of training (Anđelković, 2022):

- *Supervised training* – Input data and expected output data are presented to the network.
- *Training by evaluation* – The network is not presented with the expected output data but, after some time, is presented with the evaluation of the previous work.
- *Self-organization* – Only input is presented to the network.

Initially, neural networks were used to model the nervous systems of living organisms. Today, neural networks are applied for the following:

- Shape recognition.
- Handwriting recognition.
- Speech recognition.
- Financial and economic models.
- Predicting price movements on the market.
- Systems management.
- Management of production processes.
- Analysis of electric circuits.
- Psychiatric assessments.
- Data compression.
- Oil research.
- Criminological research.
- Analysis of medical tests.
- Examination of EEG and ECG signals.
- Finding the optimal solution.
- Managing robots.
- Analyzing data during pyrolysis and spectroscopy.
- In biocomputer systems.
- Weather forecast in other areas as well.

Neural networks have been successfully applied in supervised and unsupervised learning. Neural network techniques belong to the so-called nonlinear techniques that can model complex functions. Generally, they are applied in the fields of prediction, classification, or control in a number of fields.

A key feature of neural networks is related to learning the relationship between input and output parameters through the network training process.

Network training is used to learn behavioral patterns and the main goal of training is to find a set of weights between neurons that determine the global minimum of the "error function." This procedure includes decisions in terms of the number of training iterations, that is, the point at which training stops (Anđelković, 2022).

The most famous training algorithm is known as backpropagation. During the training phase, the training data is in the input layer. In this algorithm, each node in the hidden layer receives input from each node in the input layer, which is multiplied

by the appropriate weights and then summed. The output from the hidden node is a nonlinear transformation of the resulting sum. Similarly, each node in the output layer receives input from all nodes from the hidden layer, which are multiplied by the appropriate weights and then summed.

Then, the resulting output values are compared with the target output values. The target output values are those values that the neural network is trying to teach.

Then, the error between the obtained output values and the target values is calculated and fed back through the hidden layer. This procedure is called the backward procedure through the hidden layer. The error is used to correct the strength of the connection between the nodes and the weights between the input and hidden layers, and in this way, the hidden and output layers are updated.

6.6 Usage of Neural Network

Neural network models can be used for the following:

- *Classifications* – If the object should be associated with one of the existing, predefined groups or classes.
- *Predictions* – If an object that is not in the set of existing classes is predicted.

Before using software that includes neural network techniques, it is necessary to determine the reliability of the results obtained. Reliability is determined through procedures:

- *Validation* (This procedure provides an answer to the question: "Is the appropriate product made?")
- *Verifications* (This procedure answers the question: "Is the product made in the right way?")

Validation is a process in which the degree of accuracy of the model representation is determined in relation to the purpose of using the model (AIAA G-077-1998 2002).

Verification provides an evaluation of the accuracy of the model. Also, the verification process analyzes the test results and provides results that measure the reliability of the system. Verification is the process of determining whether the model implementation accurately represents the conceptual description of the model as its solution.

Verification and validation processes include model evaluation using real data. In order for the neural network to be successfully applied in real situations, it is necessary to evaluate the test results.

Verification ensures the complete specification and ensures that there are no errors in the implementation of the model. However, verification does not ensure that the model solves the problem, and that it accurately represents the real processes that occur.

There are several approaches for obtaining information about the quality of information obtained through data analysis models (model validation):

- Use of statistical indicators.
- Division of the data set into two parts: a part for training and testing.
- Consulting with experts and reviewing the results obtained to determine whether the discovered behavior patterns make sense in the specific use case.

This procedure is carried out in order to determine the fulfillment of the following characteristics of the created model:

- Accuracy.
- Reliability.
- Usefulness.

6.7 The Artificial Neural Network Structure

The arrangement of neurons and layers in the network represents the neural network's structure. An artificial neural network includes layers of one or several neurons. Since it is not involved in computations, the first layer is not considered a neuronal layer. The parameters of artificial neural networks and neuron behavior will be discussed in the following. In terms of the number of layers, there are two types of neural networks (Minsky & Papert, 1969).

1. Single-Layer Networks.
   This network includes a layer of neurons with weighted connections where n denotes the number of inputs and m is the number of outputs.
2. Multilayer Networks.
   These artificial neural networks include one or several hidden layers between input and output layers. Compared to single-layer networks, multilayer networks can solve more complex problems but are much more difficult to train (Camuñas-Mesa et al., 2019) (Figs. 6.10 and 6.11).

Therefore, the number of layers is an important efficiency parameter of neural networks. Selecting too few layers may render the network unable to provide a good solution, whereas selecting too many layers may significantly increase network complexity and training time. The number of neurons in each layer is another important parameter requiring careful selection since selecting the wrong number of neurons has the same consequences as the number of layers mentioned earlier (Ma & Ji, 1999; Camuñas-Mesa et al., 2019).

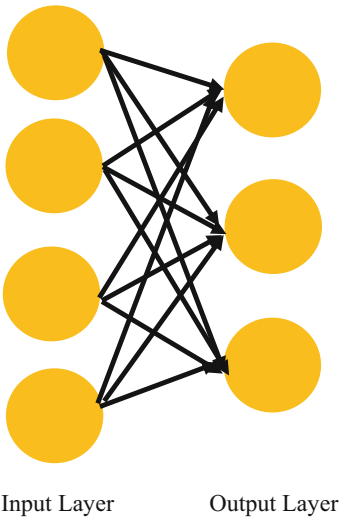


Fig. 6.10 Single-layer neural network

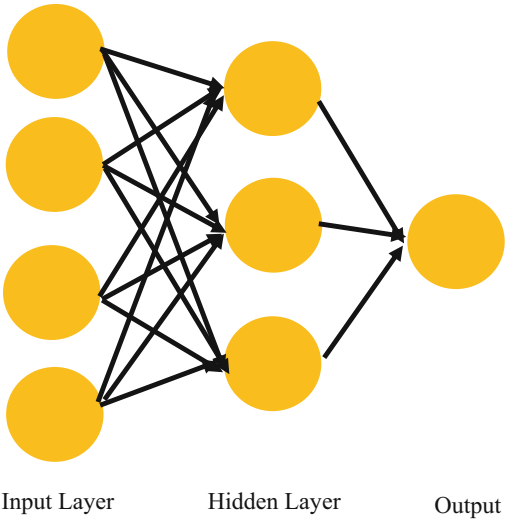


Fig. 6.11 Multilayer neural network

6.8 Modeling an Artificial Neural Network

As mentioned earlier, the elements of an artificial neural network include activation functions, weights, and biases. Generally, since the network's activation function is assumed to be unchangeable and constant, the variables are weights and biases. Therefore, the neural network can be generally considered a function with w and b parameters as follows (Fig. 6.12):

Fig. 6.12 Modeling an ANN

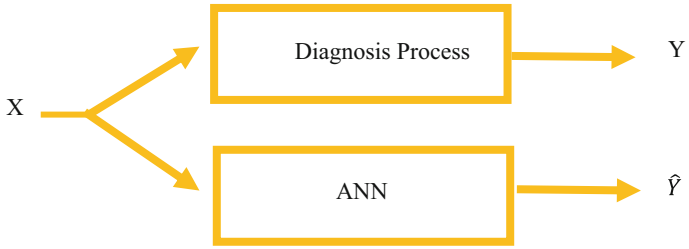


Fig. 6.13 Diagnosis process and ANN

$$y = f(x|W, B) \tag{6.8}$$

Now, using a neural network to model a process (any process, for example, the diagnosis of a mental disorder by a psychologist, can be considered a process) is shown in Fig. 6.13.

Here, Y is the output and $\widehat{Y}$ is the output the network aims to model in the shared input. Ideally, $Y = \widehat{Y}$ and modeling should be accurate. Nevertheless, natural and real problems are known to have many parameters that cannot be modeled completely. For example, is the parameter of the moon's size and its angle to the earth relevant to determining a person's mood or activity? Despite its apparent irrelevance to the problem, just as the moon's gravity affects the tides of oceans and seas, it also influences the amount of water in the human body and can affect the human body and mood. Nonetheless, these parameters are ignored without knowledge of their existence or application and measurement in the model. Thus, $Y = \widehat{Y}$ never occurs in normal problems. In this case, the goal is to minimize the error between the two outputs (Fig. 6.14):

Therefore, this is an optimization problem. To find the best model, the bias and weight values should be determined such that the value of e or its equivalent, e^2 or *MSE*, is minimized. Hence, the problem is as follows:

$$\min e^2 = \left(Y - \widehat{Y}\right)^2 \tag{6.9}$$

Since $\widehat{Y} = f(x|W, B)$, we have

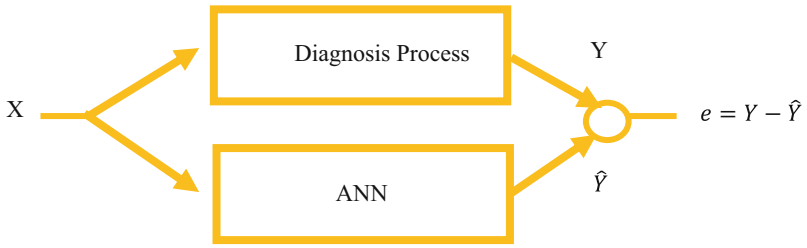


Fig. 6.14 Minimizing errors in ANN

$$\min e^2 = (Y - f(x|W, B))^2 \tag{6.10}$$

Here, we are after the correct W and B values, respectively, weights and biases.

Now, if the input is more than one number and the data covers several patients, the sum of errors, which creates the *MSE* problem, should be considered:

$$\min \frac{1}{N} \sum_{i=1}^{N} e_i^2 \tag{6.11}$$

Note that here, all inputs have the same value; if they do not, it is possible to apply weights to them.

Therefore, this is an optimization problem that requires two general optimization methods (Agatonovic-Kustrin & Beresford, 2000):

6.8.1 Classical Optimization Methods

Although this classification includes numerous methods, one of the best-known is arguably the gradient descent method. It is among the most widely used regression and classification methods and algorithms. The backpropagation method also uses this method and depends on the first-order derivative of the optimization function to determine how to change weights to minimize the function. Its advantages include ease of calculation, explanation, and implementation, but it is very memory intensive and could fall into a local minimum.

There are several measures used in the validation process of neural networks. The most commonly used measures are the following:

- Mean absolute error (MAE).
- Root mean square error (RMSE).
- Relative error (MARE).
- Correlation coefficient (R2).

Root Mean Square Error (RMSE):

- RMSE (Harper, 2022)

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=0}^{n-1}(t_i - o_i)^2} \tag{6.12}$$

where t_i is the computed output provided by the network, o_i is the desired (actual) output for case (input vector) i, and n is the number of cases in the sample. The error is averaged according to the number of output variables and in relation to the number of cases in the sample on which it is calculated.

In addition to the mentioned measure, the classification rate is also used in the classification problem as a measure of network validation. The classification rate shows the percentage or proportion of correctly classified cases.

6.8.2 Intelligent Optimization Methods

These methods use intelligent algorithms such as a genetic algorithm, PSO, ISA, and DE. Meta-heuristic algorithms can be used to determine the weights of edges and biases in neural networks.

6.9 Types of Data in Machine Learning Algorithms

Data is very important in machine learning algorithms, which actually extract relationships from existing data. In machine learning algorithms, data can be divided into three categories.

1. Training Data
Used for building the model, these data have known outputs to the model and are called learning or training data.

2. Test Data
These data are used for model evaluation and their class status is unknown to the model. After applying the data, the class status is compared to the predicted class status to evaluate the model's efficiency.

3. Validation Data
These data are used for testing various trained models in supervised learning. They are taken from training data and used for model evaluation before being returned to the model. Testing data are used with the assumption that the algorithm has used a series of questions for training, including questions and correct answers, and to test whether training was completed. In this case, validation tests are the midterms that show progress in education. Sometimes, however, these data are not considered and are only divided into test and training groups (Bonaccorso, 2017) (Fig. 6.15).

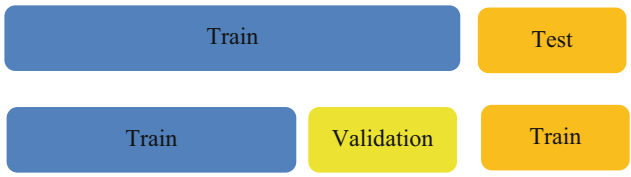


Fig. 6.15 Types of data in machine-learning algorithms

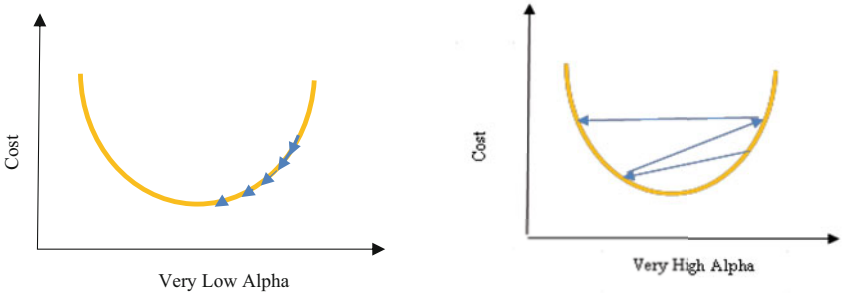


Fig. 6.16 The impact of alpha in cost function

6.10 Basic Concepts

Learning Rate

As one of the most important hyperparameters in neural networks, learning rate helps with the proper adjustment of the neural network. It indicates the magnitude of each step in each iteration for network training until the cost function is minimized. This concept is usually denoted by alpha, which varies between 0 and 1. Finding the optimal operational learning rate is difficult. If too high, that is, the steps are large, the algorithm operates faster with the risk of losing the minimum value for the cost function since the weights are updated quickly. If low, the calculation will be slower and produce results later, and if too low, there is a possibility of getting stuck in the local minimum. Thus, its size, which should be between the two states, depends on the data and network structure. This concept is shown in the following figure (Bengio, 2012) (Fig. 6.16).

Meanwhile, concepts such as epoch, batch size, or iteration in machine learning and deep learning are used when the dataset is very large and all the data cannot be entered into the model at once. Hence, the data must be divided (Wang, 2017).

Epoch: Epoch is when all the data have moved forward or backward once along the network. The epoch number is a hyperparameter related to forward and backward movement in the network with higher numbers meaning greater accuracy but also longer training time and possible network overfitting. A period is an opportunity for each sample to change the internal parameters. The number of epochs is

usually considered high so that the model continues until the error is minimized. (The number can be 0 to infinity and even greater than 100 or 1000.) If the epoch is too big to be given to the model at once, it will be divided into small batches. (Wang, 2017).

Batch: An epoch consists of one or several batches (batch size is also a hyperparameter that determines the number of samples). With this method, the model receives the data in batches instead of all at once.

Iteration: Iteration is the number of batches needed to complete an epoch; that is, the number of batches is the number of iterations in each epoch (Wang, 2017).

Network Operation Modes

Overfitting: It is one of the modes where machine learning underperforms. In overfitting, the model uses too much detail and noise from the training data, which negatively affects the model and its future predictions. In this case, model error is good for training data but poor for the test and new data, causing a generalization problem. The overfitting model is typically a nonlinear and non-parametric one. Fig. 6.17 clearly illustrates the concept of overfitting.

Underfitting: Like overfitting, it provides an inappropriate model that should be avoided. This model is inappropriate for training data and test data. In this case, different learning models should be tested to find the best model for the problem, which is shown in Fig. 6.18 (Brownlee, 2016).

Good Fit: Statistically, fit means the quality with which the model can approximate and model the target function. Machine learning algorithms include methods for this purpose, such as finding residual errors in supervised algorithms. Ideally, the model should be somewhere between overfitting and underfitting, which is a challenging goal (Brownlee, 2016) (Fig. 6.19).

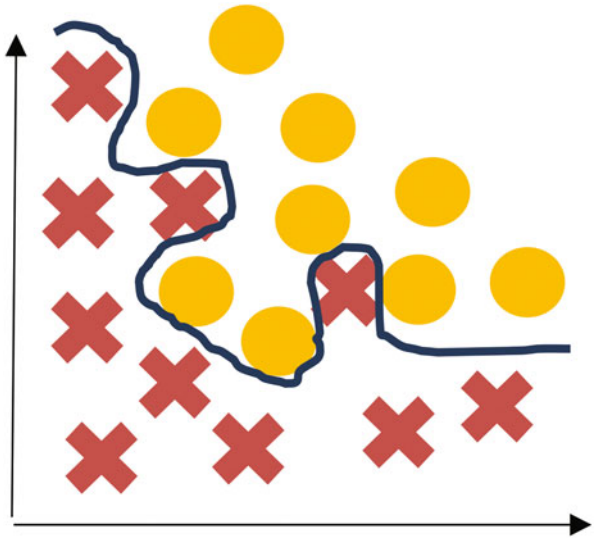


Fig. 6.17 Illustration of overfitting

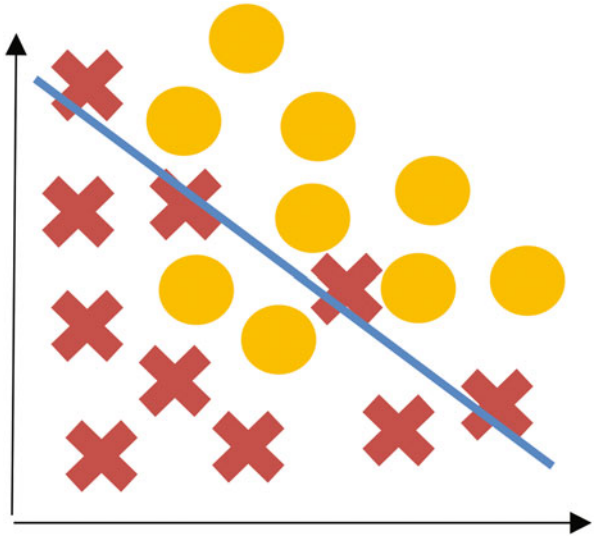


Fig. 6.18 Illustration of underfitting

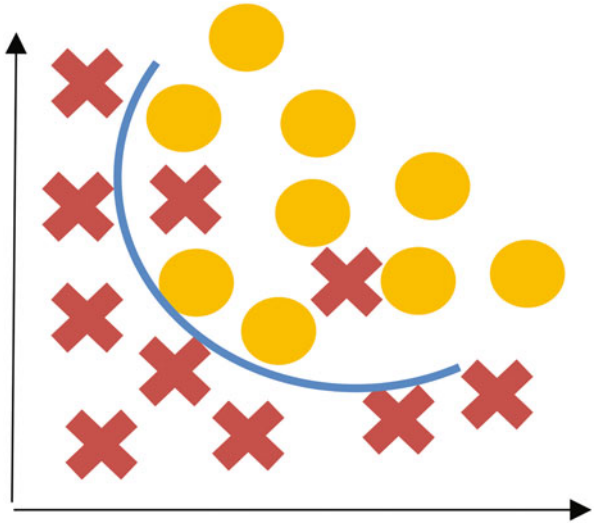


Fig. 6.19 Illustration of good fit

A model error can be measured for analysis. Ideally, training and testing data errors while training the network should be reduced together. Now, if training takes too long, the model pays too much attention to details and noise, which causes overfitting and increases test data error (Brownlee, 2016) (Fig. 6.20).

Cross-validation: As explained earlier, the model needs to estimate and check the parameters for performance evaluation (e.g., to avoid overfitting). In machine learning algorithms, the generalization feature, which is very important for inputs, is accomplished by cross-validation. Moreover, too many network parameters

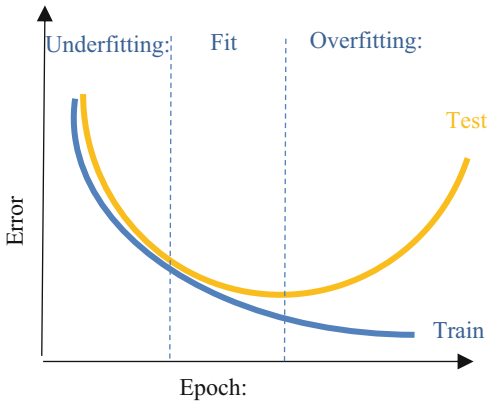


Fig. 6.20 Impact of overfitting and underfitting on error

introduce difficulty in the performance evaluation of the model and necessitate cross-validation. Cross-validation is easy to implement and understand, has less bias than other methods, and has different methods, such as the holdout method, leave-one-out method, leave-P-out method, and k-fold method (Russell, 2010).

6.11 Types of Artificial Neural Networks

There are various artificial neural networks for different applications, all inspired by the human nervous system. A close examination of learning in the human brain reveals a similar process via the weakening or strengthening of the connections between brain cells or neurons. In mathematics, these changes are known as weights. Some neural networks will be discussed in the following (Fig. 6.21).

Multilayer Perceptron (MLP) Neural Network
Considered one of the most basic artificial neural networks, it typically consists of an input layer, one or several hidden layers, and an output layer. This neural network considers the network behavior of humans and is also known as the feed-forward network. Here, neurons receive, process, and transfer signals between each other, which ultimately produces a result. This network uses the mean square error (MSE) as an efficiency index (Marques, et al., 2014).

Feed-Forward Neural Network
This is also one of the older neural networks with several main rules, such as the interconnectivity of all nodes, activation from the input layer to the output layer, and a hidden layer between input and output.

Deep feed-forward neural networks have more than one hidden layer, which is very time-consuming and impractical.

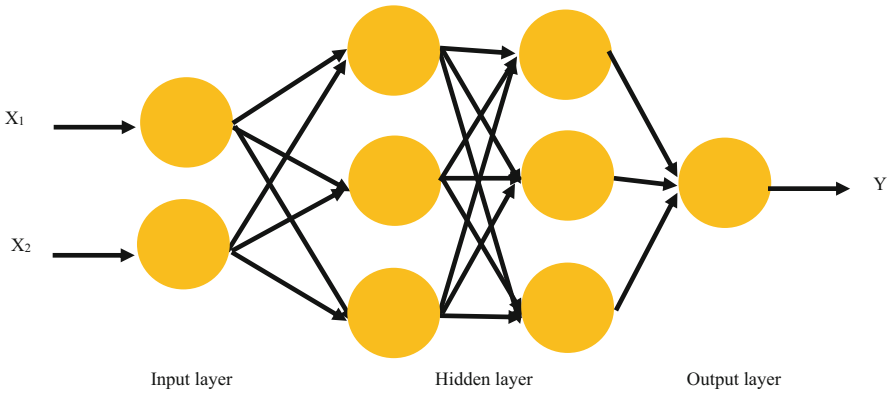


Fig. 6.21 Perceptron network

Radial Basis Function Neural Network
Despite its structural similarity to MLP, this network has different neuron processing. It employs the logistic function for the activation function. Due to its easier setup, this network provides faster training and is suitable for classification. However, it underperforms with regression.

Recurrent Neural Networks
In these networks, the neurons have a recurrent state. Basically, recurrent neural networks are used when the "context" is important and when decisions from previous iterations or samples can affect current examples.

Deep Learning
Generally, a higher number of layers and neurons in the artificial neural network make the network deeper and the model more complex. Having over three layers in the artificial neural network creates a deep network used for deep learning. In fact, deep learning is an artificial neural network with a large number of hidden layers. Nonetheless, the difference between deep learning and neural networks lies in the fact that deep learning is the learning mode, whereas neural networks are machine learning algorithms (Dangeti, 2017).

6.12 Comparing Multilayer Neural Network with Regression

If an artificial neural network has no hidden layers but has a sigmoid activation layer, the neural network is equal to a logistic regression. If it has a linear activation function, it is equal to a linear regression.

An artificial neural network designed for output layer classification has a class label.

In recent years, neural networks have been redesigned in deep learning format, often leading to better results (Goodfellow et al., 2016).

The aim of machine learning (ML) is to develop computing algorithms or statistical models that can automatically infer hidden patterns from data. In the latest technological wave of ML and AI, deep learning approaches aim to develop an end-to-end mechanism that can directly map input raw features to outputs and place a multilayer network structure between them to capture inputs from hidden patterns.

There are various deep learning (DL) strategies. This chapter addresses only the deep feed-forward neural network.

As previously mentioned, ANNs aim to mimic how the human brain works. From a mathematical point of view, an artificial neuron is viewed as a nonlinear transformation unit that captures the weighted set of all inputs and feeds the result to an activation function (e.g., sigmoid, rectifier (relu), and hyperbolic tangent).

The simplest ANN is a feed-forward neural network that stacks neurons layer by layer in a forward pattern. The first layer is the input layer. Each unit gets one dimension of the data vector. The last layer is an output layer that outputs the probability. It also assigns an individual to different classes. The middle layer is the hidden layer. There are usually several hidden layers, which is why a network is called a deep network.

The input features are considered as an input layer when designing deep artificial neural networks (DANNs). The hidden layer is mainly selected as dense or fully connected, and each neuron is assigned to another neuron in previous or next layers. The network is then trained to find the data pattern. This process results in a model that can be used to predict scores or classes (here classes) in the new untrained features. In every training process, there is an epoch. An epoch is the number of times a model has updated itself. Updating means calculating the error and backpropagating its gradient in previous layers. The class weight is a very important parameter in the network because it reduces the relative weights of false negatives in the total error.

It is important to manipulate this parameter because researchers often have to deal with false positives. Therefore, this ratio must be chosen carefully. The ratio is usually chosen as 1:4 because it is usually acceptable (Bailey et al., 2021).

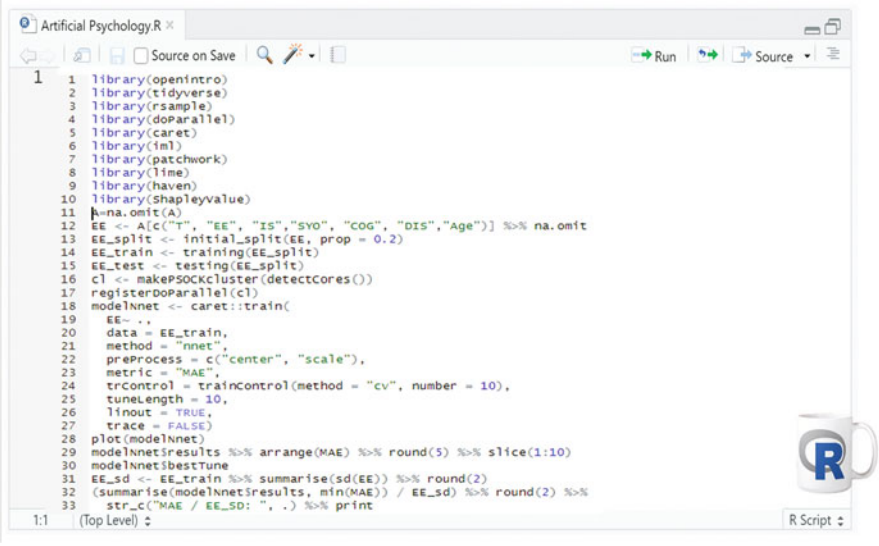
After the network has been trained, its appropriateness should be tested. For this purpose, it is imperative to use the data that has not been used to train the network. From a psychological point of view, testing all models would be the test of generalization to the untrained.

To test a model, the prototype is divided into two parts: 70% training and 30% testing.

6.12.1 Practical Example Using R

Example 6.1 In this research, an artificial psychologist using DNN artificial neural network tries to predict emotional competence (EE) based on two other variables. This model shows the application of DNN regression. The R codes for DNN are in Listings 6.1 and 6.2. In this example, in order to ensure the results, in addition to dividing the main sample into two training sample and the test sample, the researcher also performs cross-validation with $K = 10$. He also standardizes the predictor variables.

Listing 6.1 R codes for DNN implementation



Listing 6.2 R codes for DNN implementation (continued)

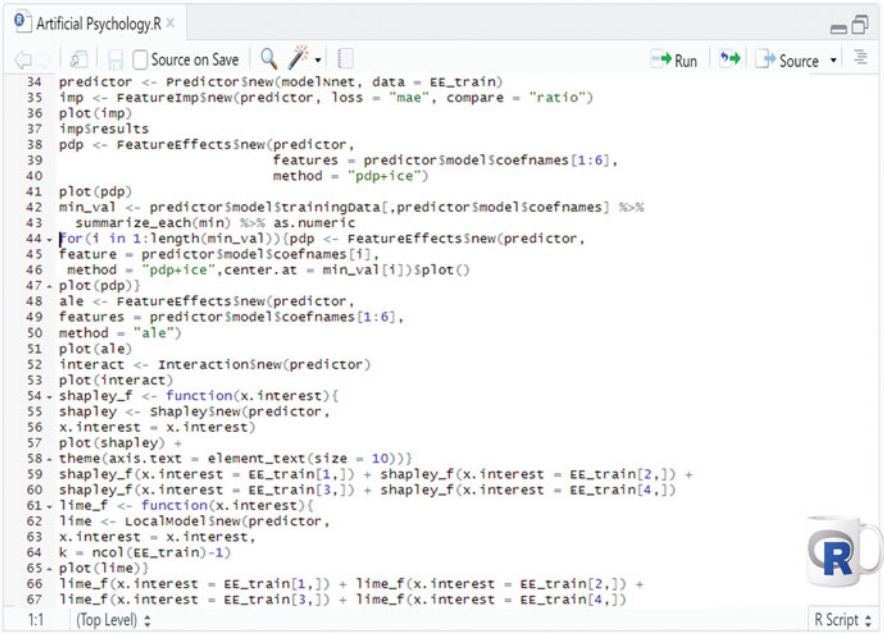


Table 6.2 shows the variable labels and their names. In this analysis, the cost function is the loss function and has been used by MAE. Figure 6.22 shows the median importance of the features.

This plot shows that the IS variable has the most relative importance in predicting EE and COG has the least importance.

The chart provides summary information about the behavior of the model. There are different models for interpreting features (variables). One of these methods proposed by Friedman in 2001 is the partial dependence plot (PDP), which shows the marginal effect of variables on the dependent variable(s) in the trained model (Listing 6.3). PD shows how much the predicted values change for a change in the value of a variable in the model.

Table 6.2 Abbreviated name and description of variables

Variable name	Variable label
Internalized shame	IS
Childhood trauma	T
Cognitive flexibility	COG
Distress tolerance	DIS
Alexithymia	SYO
Emotional competence	EE
Age	Age

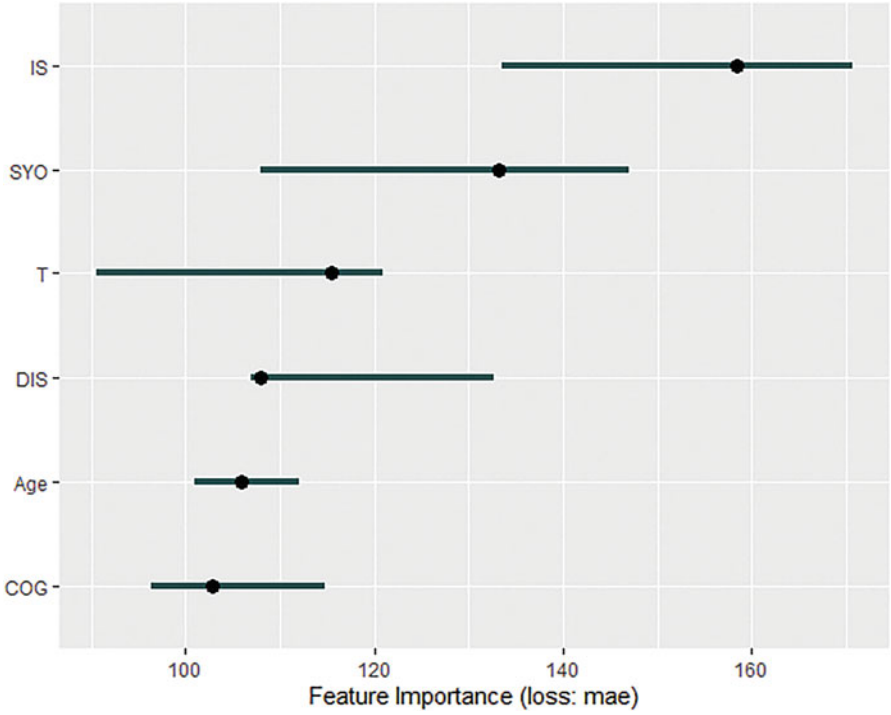


Fig. 6.22 The median importance

In 2015, Goldstein et al. introduced the individual conditional expectation (ICE) plot. This plot shows how the predictions of a fitted model change if only the variable in question changes compared to the rest of the variables. One condition is that the rest of the variables are fixed based on the value of a sample.

Listing 6.3 shows the R codes for drawing PDP and Rice. The results are shown in Fig. 6.23. In this PDP diagram, they are drawn as yellow curves and ICE curves are drawn in black, and the distributions of each variable are shown as short lines on the X axis. It can be seen that descending lines like IS and T show a negative relationship with EE and a variable like SYO shows a nonlinear relationship.

In 2016, Alpey introduced a new plot called accumulated local effect (ALE). The ALE plot calculates the average effect of predictor variables on model predictions. It is faster and more unbiased than PDP. This chart includes the interaction of variables even if they are correlated. The ALE plot related to the DNN regression model is shown in Fig. 6.24.

Figure 6.24 indicates that predictor variables have nonlinear relationships with EE, which was not easily observed in PDP.

Figure 6.25 shows feature interaction. This diagram shows a way to show the interaction strength between two features. This value was given by Friedman and

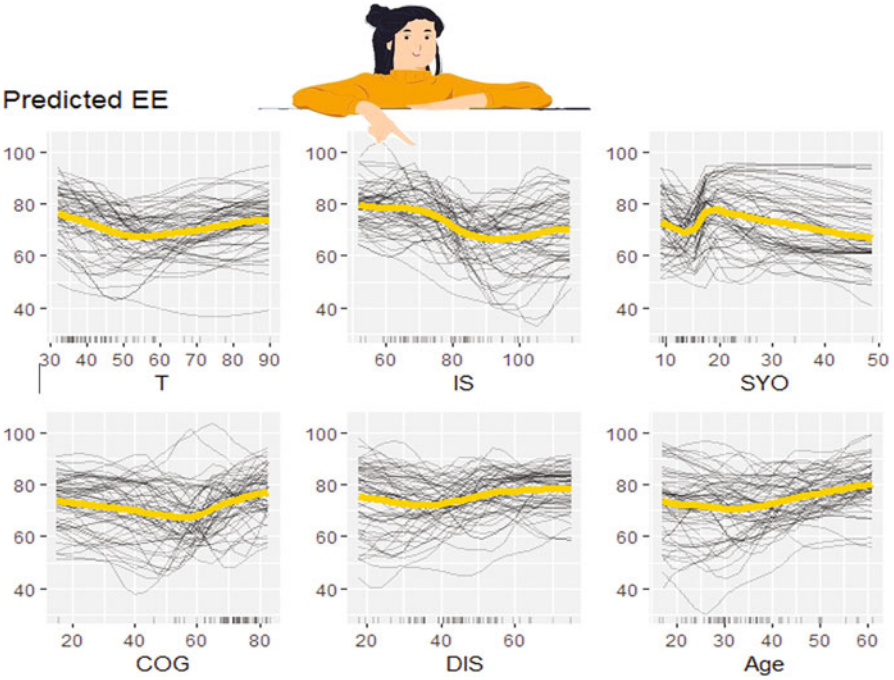


Fig. 6.23 PDP and ICE plot

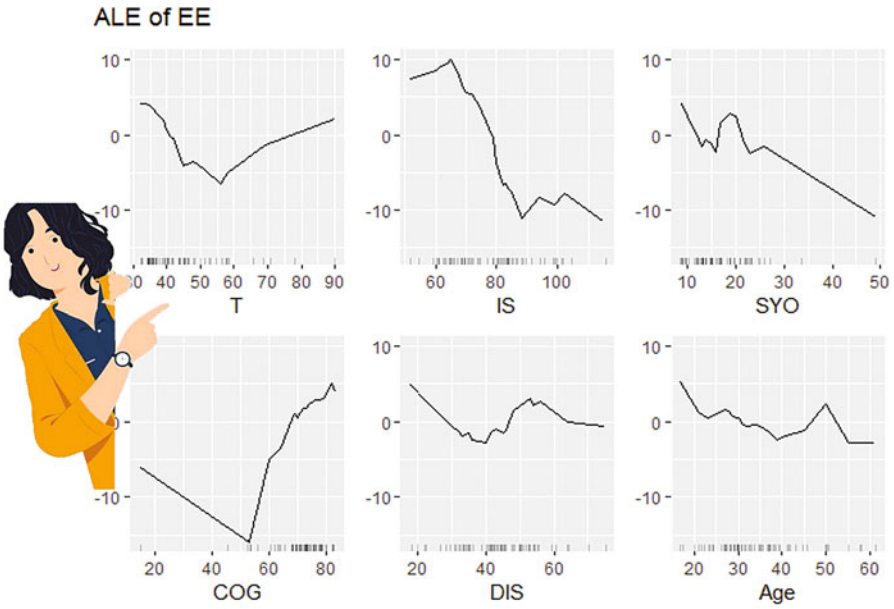


Fig. 6.24 ALE plot

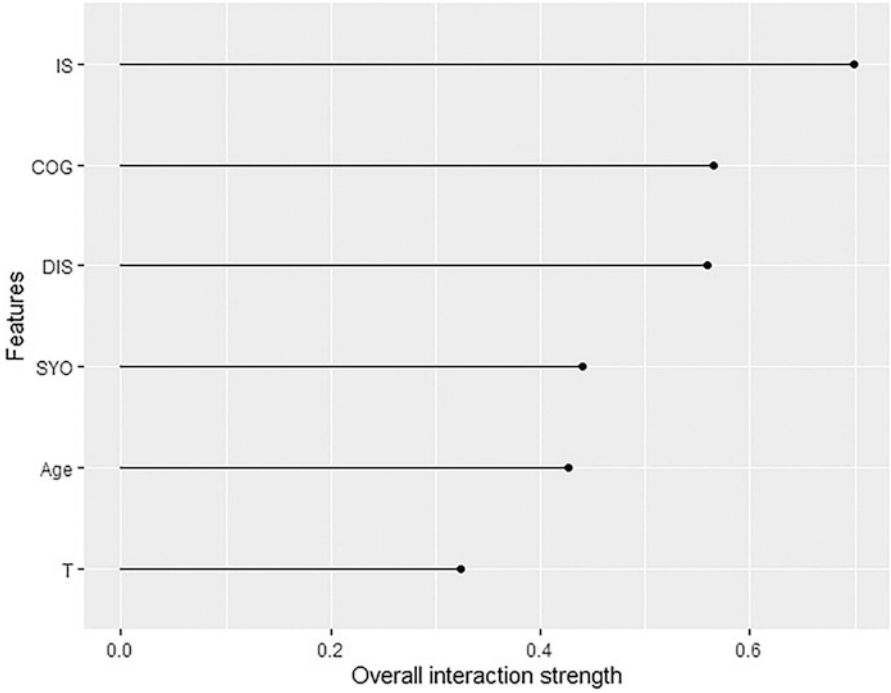


Fig. 6.25 Interaction plot of the variables

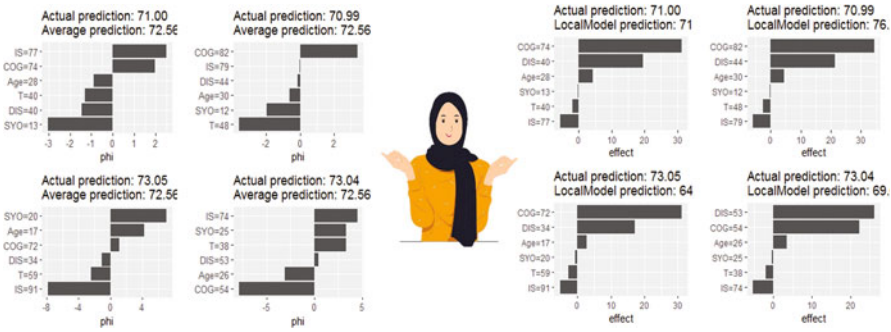


Fig. 6.26 Interaction plot of the variables

varied between 0 and 1. Zero (0) represents no interaction between two variables and 1 (one) is for when all the effect on the observed values is due to the interaction.

As the interaction plot in Fig. 6.26 shows, the two variables IS and COG have the most interaction with other variables and *T* the least.

As discussed, LIME and the Shapley value are used for local interpretation. LIME is not used in this example because the package used in this case does not use LIME, but we will discuss it in the next example. We use the Shapley value in this example.

The artificial psychologist should keep in mind the importance when using black-box ML in Interpretability and explainability of findings and results. The Shapley value was developed in 1953 based on game theory, which actually shows the relative contribution of predictor variables in predictions. In the IML package available in R, this indicator is based on a Monte Carlo sampling approximation by Strumbelj et al. in 2014.

Figure 6.26 shows the average predicted values compared to the actual prediction. Using the Shapley value, it is possible to compare actual and average predictions for a subset of data and even for a single data point.

6.13 Hyper-Parameter Tuning

The model training process involves the use of parameters. Depending on the data and the type of problem, parameters are obtained from different data sets.

However, hyper-parameters are determined by the involvement of a researcher. For example, they refer to the number of hidden layers in a DNN or the number of trees in a random forest. In general, this process aims to optimize models by reducing errors or cost functions and improving accuracy. Two approaches should be considered by artificial psychology when determining hyper-parameters: (1) a manual approach and (2) an automated approach. The former is based on prior knowledge. Sometimes an artificial psychologist knows that better results are obtained when hyper-parameters are determined by the knowledge-based model. However, the second approach does not require prior knowledge. There are various parameters in DNNs. They can be manipulated directly. This changes the model architecture or the learning rate (Yang & Shami, 2020).

Hyper-parameters include the number of neurons in a hidden layer, the number of epochs, the number of hidden layers, the number of mini-batches, the activation function, the cost function, the learning rate, and regularization methods (Koutsoukas et al., 2017).

The first approach consists of some recommendations and approximation techniques that can be used to determine hyper-parameters and observe their impact on the model.

When determining the number of neurons in a hidden layer, one solution is to use half of the input data (Heaton, 2015). A better solution is to employ the hyper-parameter tuning technique. To do this, the training data is split into two halves:

1. Training set.
2. Validation set.

The training set is utilized to train the model while the validation set is only used to check the model accuracy.

In practice, 50% of the data goes into the validation set and the rest goes into the training set, so both sets can contain the same amount of data. Therefore, the

accuracy analysis is performed on the validation set rather than the test set, preventing overfitting. The optimal model is determined by hyper-parameter tuning in the validation set.

The random search cross-validation can be used in an automated approach. This method can be employed to resolve the opacity in parameter regularization when there is no prior knowledge. In this method, a grid of value ranges is determined for hyper-parameters, from which some values are then randomly selected and evaluated. In this method, the most important arguments control the number of iterations (*n*-iter) and the number of different combinations tested (usually, it is considered 100). Undoubtedly, increasing the number of these two arguments will reduce overfitting; however, a long time is required to fit the model.

Therefore, the best post-training hyperparameters can usually be obtained when random search CV is implemented. This is considered a major achievement as it helps access a smaller range of hyperparameter values. All resulting combinations are then analyzed using grid search CV instead of random sampling. The CV random search is used to reduce the space of hyper-parameters, while CV grid search is used to determine the best parameters.

After the model has been trained and validated, a new sample is used to check generalizability. In other words, 30% of the prototype that was not used for model training is now used.

Sometimes the data may not be enough (i.e., the sample is small). In this case, cross-validation can be used. To do this, a single data set is divided into two or more approximately equal data sets. They are called folds and are represented as "k," which is usually taken as 10. In this case, the algorithm trains the model in $k - 1$ folds and then tests it on the remaining fold, called the holdout. The number of folds is usually considered to be 10 in cross-validation.

6.14 Evaluation of DNNs

The confusion matrix is used to evaluate DNNs with the purpose of classifying the target variable. This matrix helps determine the classification and misclassification rates in the model. Conventionally, the effectiveness and performance are determined by its accuracy. There are some components in the confusion matrix (Table 6.3).

Table 6.3 Confusion matrix layout. "1" for positive class

	1	0
0	TN	FP
1	FN	TP

TN true negative, *FP* false positive, *FN* false negative, *TP* true positive

The following four metrics are employed to evaluate the performance of a classification model:

1. Accuracy: This metric indicates the total percentage of individuals classified correctly.

$$\text{Global accuracy} = \frac{TP}{N} \tag{6.13}$$

2. Precision: This metric indicates what percentage of individuals having the predicted feature actually have that feature.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{6.14}$$

3. Recall: This metric indicates what percentage of individuals that actually have that feature possess the predicted feature correctly.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{6.15}$$

4. F-Measure or F-Score: This metric is a combination of precision and recall, and R indicates the importance of recall over precision. If $B = 1$, then they are equally important.

$$F - \text{Measure} = \left(1 + \beta^2\right) \times \frac{\text{precision} * \text{recall}}{\beta 2 \text{ precision} + \text{recall}} \tag{6.16}$$

The receiver operating characteristic (ROC) curve is another tool for analyzing the performance of classification models. The area under the curve (AUC) of ROC is measured for performance analysis. In the ROC curve, the true positive rate (TPR) is drawn against the false positive rate (FPR).

One of the most common metrics is $R2$, the proportion of variance accounted for by the model. Higher $R2$ values signify higher accuracy. When the residual variance (i.e., "sum of squared residuals" in the formula below) is zero, the model makes perfect predictions and $R2 = 1$. If the sum of the residuals is equal to the total variance (in the denominator), the model is useless, predicting the mean is equally accurate, and $R2 = 0$. Other metrics focus on the average residual size (instead of residual proportion as in $R2$), with smaller residuals signifying higher prediction accuracy. The most common is the mean absolute error (MAE); negative signs of residuals are removed and the root means square error (RMSE) residuals are squared.

$$\text{MAE} - \frac{1}{n} \times \sum_{i=1}^{n} |y_i - \widehat{y}_i| \tag{6.17}$$

$$\text{RMSE} - \sqrt{\frac{\sum_{i=1}^{n} (y_i - \widehat{y}_i)}{n}} \tag{6.18}$$

6.15 Interpretability and Explainability in DNNs

First, it is essential to distinguish interpretation from explanation. In fact, interpretation refers to a model's ability to interpret the input–output relationship, while explanation denotes the ability to explain the model output in human language.

As a model becomes more complex, it becomes increasingly difficult to introspect and understand how that model makes a particular prediction. There are many methods to explain black box models. These methods are usually adopted in terms of their scope, model dependency or independence, and runtime.

An artificial psychologist knows that a DNN is a black box model. Although it has accurate inputs and outputs, it must analyze these results to determine where to find the meaning of features and how to interpret and explain the results. LIME is a new modern tool for the model-agnostic technique. It can be applied to any model to make the results interpretable and justifiable. LIME was first introduced in 2016 by Marco Tulio Riberio, Sameer Singh, and Carlos Guestrin and stands for locally interpredictable model-agnostic explanations. They proposed this method in a paper entitled, "Why Should I Trust You? Explaining the Predictions of Any Classifier?"

According to the results reported by the developers of this method,

1. The explanations of predictions should be understandable; they should be explainable in every aspect.
2. It should be possible to make individual predictions, something which is called the local fidelity.
3. The explaining model should be applicable to all models, something which is considered model-agnostic.
4. The model should be able to provide a general explanation, which means the global perspective.

Therefore, it can be concluded that explanation is very important because we need to trust that the model predictions are correct (trust), the model behavior is correctly perceived (control), the signs of explanation are sufficient and data meet the expectations (prediction evaluation), and how it is possible to improve and enhance classification (improvement and enhancement) (Figs. 6.27, 6.28, 6.29, and 6.30).

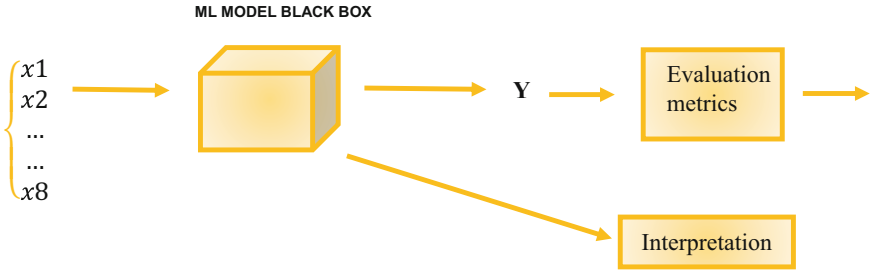
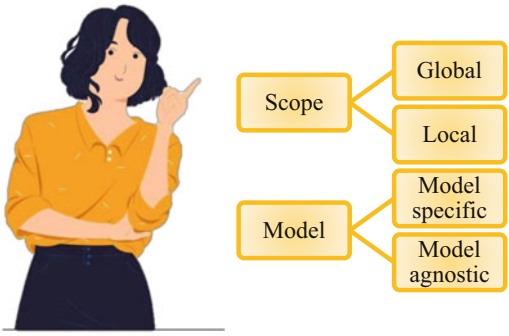


Fig. 6.27 Explainable and interpretable ML models

Fig. 6.28 Two main types of explainable and interpretable approach



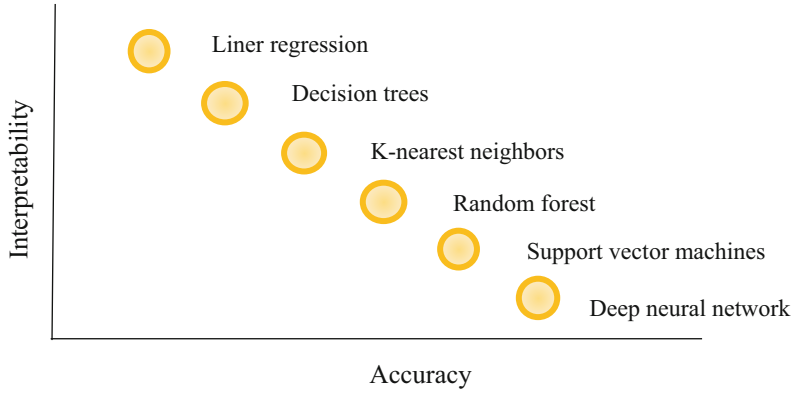


Fig. 6.29 Reverse relationship between interpretability and accuracy of ML models

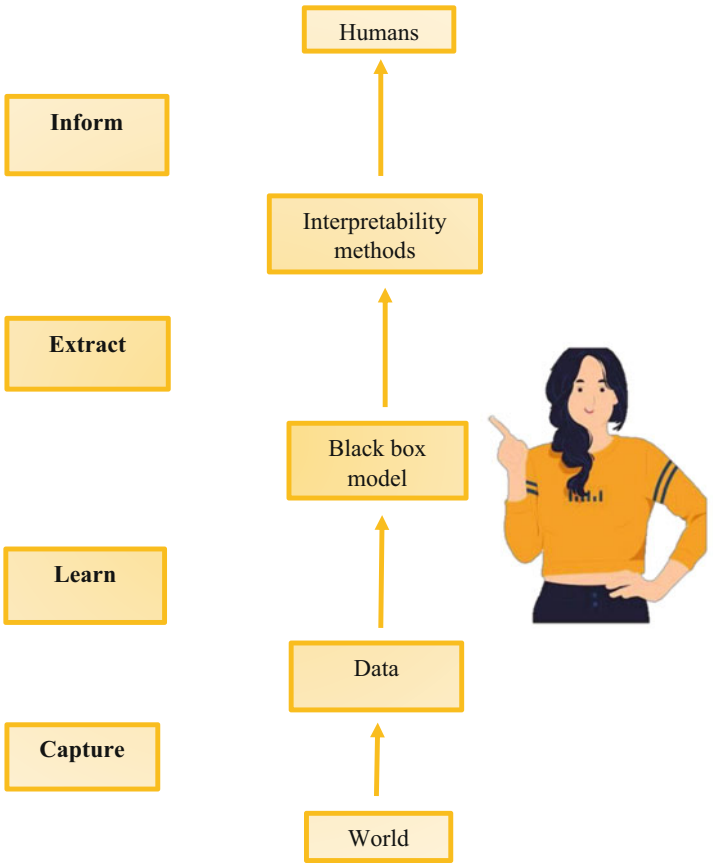


Fig. 6.30 The conceptual framework to capture knowledge from the world using models

In LIME, it is desirable that artificial psychology perceives the relationship between the features of a specific sample and the predictions of that model by training a more explainable model, such as a linear model driven from samples that are slightly different from the original inputs.

An explanation can be extracted from the coefficients of features that exceed a threshold in a linear model. The intuitive logic behind this reasoning is that these features of the linear model have the greatest importance in explaining and predicting that model. As a result, these local examples can be employed to assess the contributions of every feature in the resultant explanation and prediction. In other words, LIME creates new datasets that include perturbed samples. Hence, LIME trains an interpretable model on these new data. This model is then weighted through the proximity of samples taken from the analysis samples.

A trainable model might be a good approximation of local predictions for machine learning; however, it may not necessarily have a global fit.

Another new method of explaining the predictions of a deep neural network is to use Shapley additive explanations abbreviated to SHAP.

In fact, SHAP is a novel method for making individual predictions obtained from a complicated model. The goal of SHAP is to calculate the share of each feature in prediction to determine the effect of each input. The origin of the technique of SHAP is rooted in the cooperative game theory, the principles of which are used to calculate Shapley values.

As the game theory aims to analyze how the coalitions of some players affect the results, SHAP uses the same method to determine how the features contribute to the model outputs. In game theory, specific players have key roles in the results. Likewise, some features play major roles in model prediction in artificial neural networks; therefore, they are more important.

As mentioned previously, Shapley values are calculated in SHAP. This is the average metric for all marginal contributions for all coalitions. In SHAP, the feature-related data act as the members of a coalition in an instance, and the explanations of Shapely values are expressed as a form of an additive approach to a linear model.

Not only do Shapley values determine the relevance of a feature, but they also indicate whether that feature has a positive or negative effect on a prediction.

6.16 Difference between LIME and SHAP

In LIME, a set of important features is obtained. However, the size of the role of those features in the model is not outputted. In other words, LIME cannot, for example, determine the attribution of the exact amount of stress in predicting the pain severity in fibromyalgia. It can merely indicate whether a feature is important in predicting the pain severity of patients with fibromyalgia.

However, an output is attributed to a combination of measured features in SHAP. In fact, SHAP can be considered as the statistical attribution of a model output to a set of inputs. For instance, 4 degrees of the total pain severity are predicted out of a possible 10 degrees and then attributed to a 20-degree stress severity in these patients.

6.16.1 Practical Example Using R

Example 6.2 An artificial psychologist seeks to predict the probability of group membership using four emotional variables ($A1$, $A2$, $A3$, and $A4$). He uses a multilayer perceptron artificial neural network.

Since there is one hidden layer in the model, this example is not classified as deep learning. For this purpose, he examined 194 10–12-year-old elementary school boys, of whom 97 had high hyperactivity symptoms and 97 had low hyperactivity symptoms. He measures a child's stress level = $A1$, child's anxiety = $A2$, mother's

stress = $A3$, and mother's anxiety = $A4$. Then, he considers 80% of the total sample as the training sample and 20% as the test sample group. He also considers the execution times once for convenience. Also, he uses the CV method to validate the model and, if necessary, changes the hyper-parameter of the model to find the best model. He has used the grid search method for tuning his model. Also, size and decay must be changed in the model. Size is the number of hidden layer nodes and the decay parameter controls the amount of decay to change the weights. The weights are changed in backpropagation models. In this case, it is said that the model learns. The decay parameter takes values from 0.001 to 0.1 depending on the model. He has considered the amount of decay in the range of 0–0.1. Weights in an artificial neural network are similar to regression coefficients. Using the Garson algorithm, the relative importance of each variable in the prediction is determined. The process of pooling and scaling of all the weights related to a variable in the model creates a single value that is in the range of 0 to 1 to assess the relative importance of that variable in prediction.

Accordingly, the graph from the Garson algorithm shows in order of decreasing importance that $A4$, $A3$, $A1$, and $A2$ are the most important predictor variables (Fig. 6.31).

Figure 6.32 shows that the inputs in the net plot are $A1$ to $A4$, which have a hidden back layer with two neurons. Fact or factor is the output variable (group membership). He used LIME to explain the model.

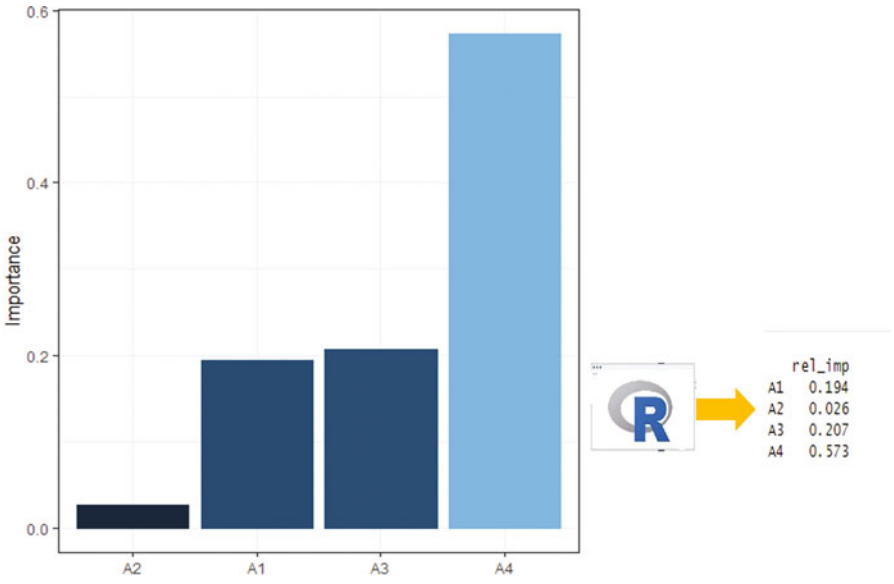


Fig. 6.31 Importance plot of MLP

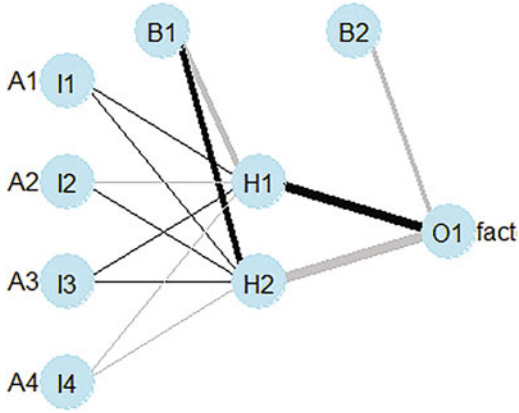


Fig. 6.32 Net plot of MLP

LIME is an interpretability agnostic method. Agnostic, which is related to the black box concept, refers to the fact that it can be used to generate insight into a process that is unknown or cannot be known. This method is used to interpret complex models. This method creates a qualitative link between input and response variables (group membership). In the LIME algorithm, the explain () function is used, which is used for new observations (such as children numbers 21–23). In this diagram, case refers to the person's number in the data set, and label refers to the observed value of the person on the target variable. Probability refers to the predicted probability for that label, and the explanation fit index measures the quality of the model used in the explanation.

The features (variables) marked in blue in the plot are the variables that support their label, and the length of the bar in the diagram indicates the weight ratio of a characteristic. A facetted heat map style plot shows the feature combinations for cases 21–23 (Fig. 6.33).

The individual number can be seen on the horizontal axis (X) and the features that have been categorized can be seen on the vertical axis (Y). The results of running the model based on size and decay are shown in Fig. 6.34.

As can be seen, with a size of 2.0 and decay equal to 0.1, the accuracy of the model is maximized and equal to 0.674.

Figure 6.35 is a facetted heatmap-style visualization of all case-feature combinations for three selected cases. The case numbers are shown on the horizontal axis, and categorized features are shown on the vertical axis.

Example 6.3 An artificial cyber psychologist tries to predict the tendency to cyber addiction (low-high) based on 13 early maladaptive schemas that he thinks are effective based on the background of the research. Since the criterion variable or target is binary, he is trying to predict group membership (categorization). He fits a multilayer perceptron artificial neural network (MLP), which uses two hidden layers. He considers each hidden layer to consist of 10 neurons. The number of hidden layers and their number of neurons are hyper-parameters. By increasing the number of hidden layers and the number of neurons, the accuracy of the prediction model

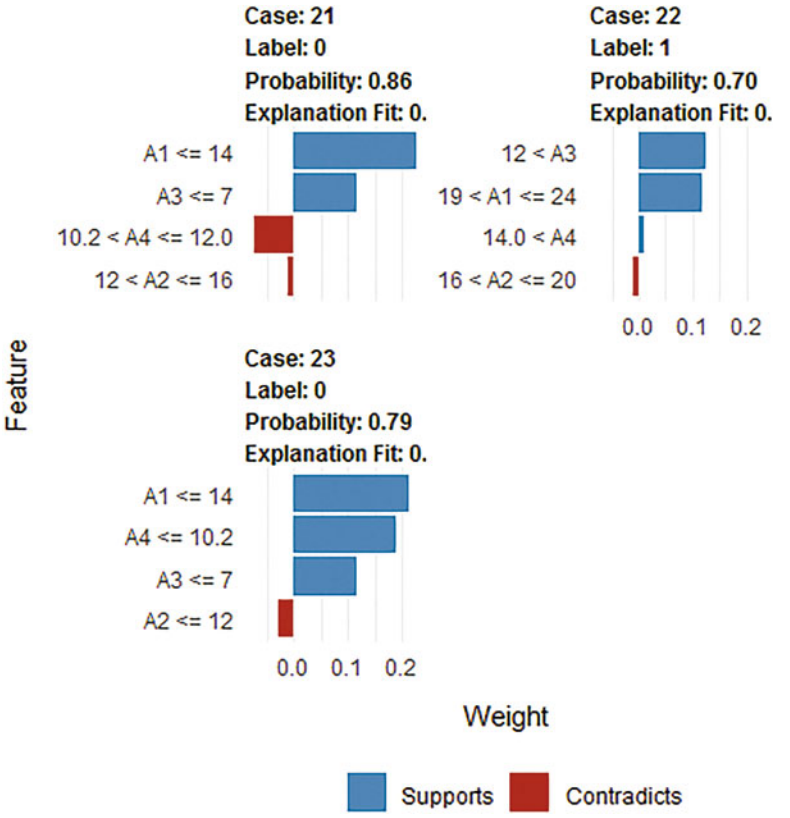


Fig. 6.33 LIME for cases 21–23

	size	decay	Accuracy	Kappa	AccuracySD	KappaSD
1	2	0.000	0.6084211	0.2171782	0.08327145	0.1553832
2	2	0.001	0.6589474	0.3113958	0.08534687	0.1725749
3	2	0.010	0.6686842	0.3326997	0.10544903	0.2130580
4	2	0.100	0.6736842	0.3426910	0.12341329	0.2496476

Fig. 6.34 Size and decay of the model

(Classification Regression) increases, but the model becomes complicated and its execution time becomes longer. The R codes are in the Listing. In this model, high group is considered as code 1. First, the quantitative predictor variables (which are all quantitative in this example) are standardized, and the cyber addiction variable, marked with the AD symbol, is set as the criterion or target variable. To implement the model, the researcher will need the necessary specifications in NE.net. Of course, these specifications can be changed. A confusion matrix is used to check classification accuracy. The R output of this analysis is shown in Fig. 6.36.

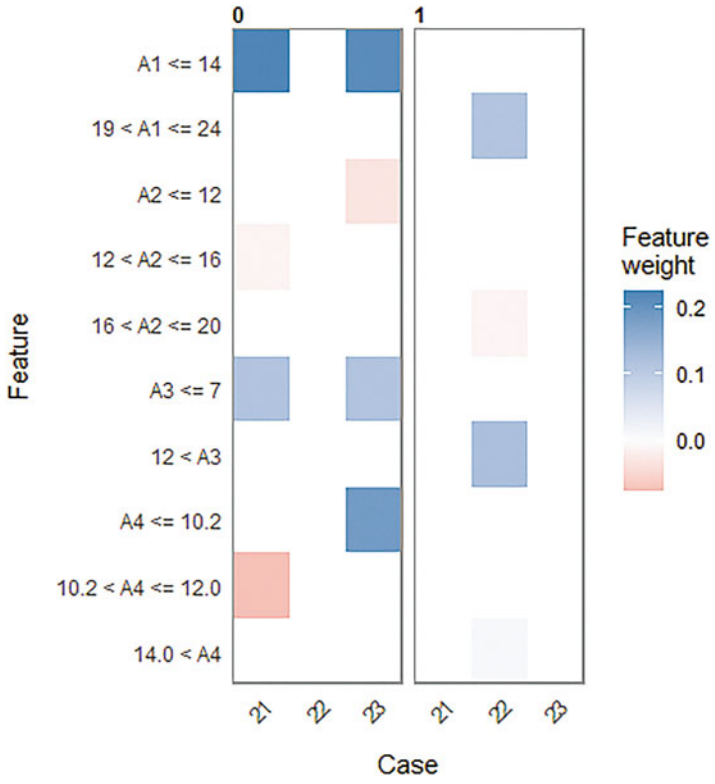


Fig. 6.35 Overall LIME for the cases and feature weight

In this figure, it can be seen that the accuracy of the model is equal to 0.55, the p-value is equal to 0.31, the kappa coefficient is equal to 0.10, and the sensitivity is equal to 0.53 with the specificity equal to 0.58, which is indicative of low model accuracy.

Figure 6.37 shows the network plot. The number of neurons in the input layer is equal to the number of predictor variables, i.e., 12 variables (the same 12 primary maladaptive schemas that the researcher identified among the primary maladaptive schemas most related to cyber addiction). The number of hidden layers is 2 layers each having 10 neurons, and finally we have the output layer, which is the tendency to cyber addiction (high or low).

As discussed in detail in this chapter, interpretability and explainability are two very important features that should be taken into account when using these models. The SHAP algorithm has been used for this purpose. In this figure, it can be seen that each point is based on the test sample. In the next example, we will examine this issue in a clearer way.

As mentioned, the psychologist uses a multilayer perceptron neural network with two hidden layers for classification. For this purpose, out of 194 participants, he selects 155 people (80%) as a training sample and 39 people (around 20%) as a test sample (Listings 6.3 and 6.4).

```
> confusionMatrix(predicted, test$AD)
Confusion Matrix and Statistics

          Reference
Prediction High Low
      High   10   8
      Low     9  11

               Accuracy : 0.5526
                 95% CI : (0.383, 0.7138)
    No Information Rate : 0.5
    P-Value [Acc > NIR] : 0.3136

                  Kappa : 0.1053

 Mcnemar's Test P-Value : 1.0000

            Sensitivity : 0.5263
            Specificity : 0.5789
         Pos Pred Value : 0.5556
         Neg Pred Value : 0.5500
             Prevalence : 0.5000
         Detection Rate : 0.2632
   Detection Prevalence : 0.4737
      Balanced Accuracy : 0.5526

       'Positive' Class : High
```

Fig. 6.36 The R output of the deep MLP classifier

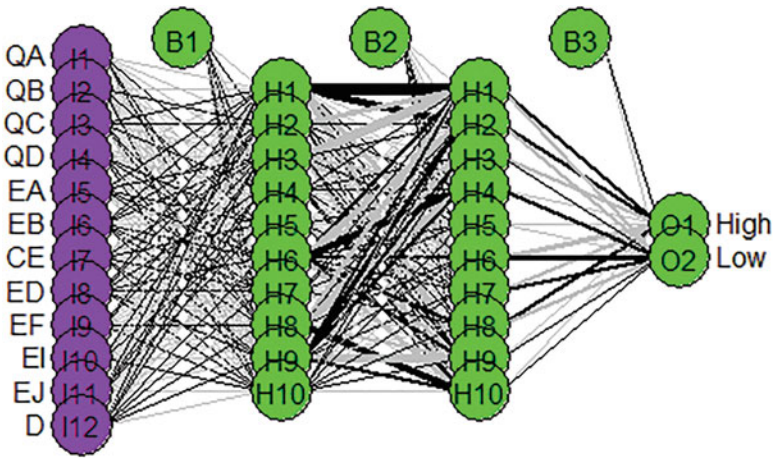
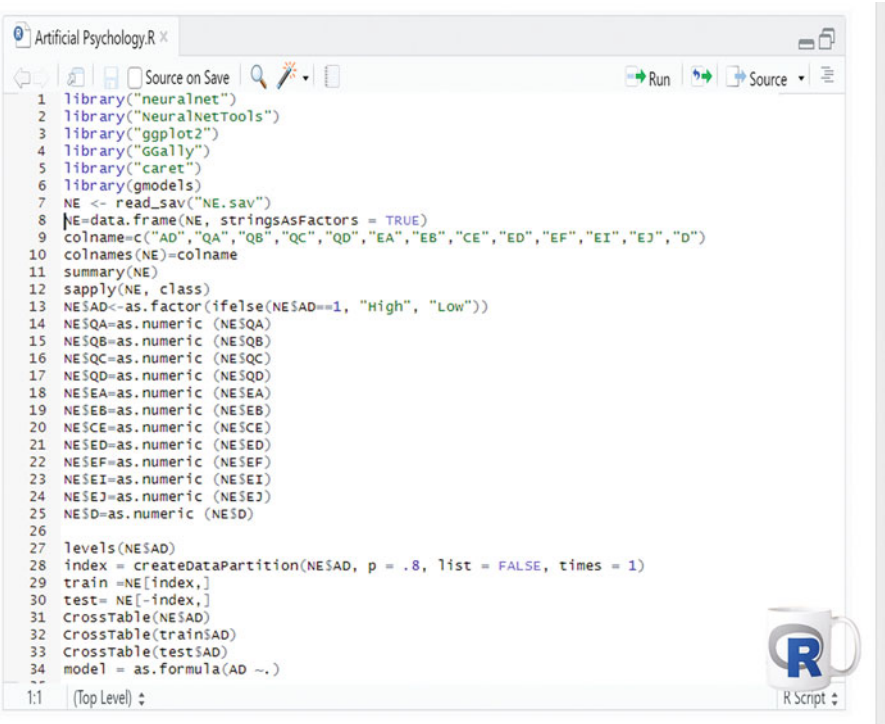


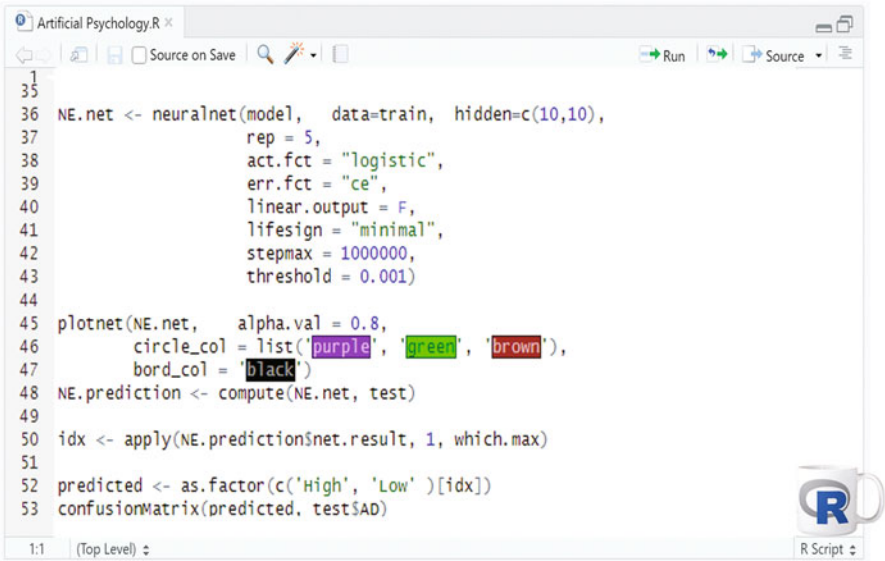
Fig. 6.37 Net plot

Listing 6.3 R codes for deep MLP



```
Artificial Psychology.R
Source on Save | Run | Source
1  library("neuralnet")
2  library("NeuralNetTools")
3  library("ggplot2")
4  library("GGally")
5  library("caret")
6  library(gmodels)
7  NE <- read_sav("NE.sav")
8  NE=data.frame(NE, stringsAsFactors = TRUE)
9  colname=c("AD","QA","QB","QC","QD","EA","EB","CE","ED","EF","EI","EJ","D")
10 colnames(NE)=colname
11 summary(NE)
12 sapply(NE, class)
13 NE$AD<-as.factor(ifelse(NE$AD==1, "High", "Low"))
14 NE$QA=as.numeric (NE$QA)
15 NE$QB=as.numeric (NE$QB)
16 NE$QC=as.numeric (NE$QC)
17 NE$QD=as.numeric (NE$QD)
18 NE$EA=as.numeric (NE$EA)
19 NE$EB=as.numeric (NE$EB)
20 NE$CE=as.numeric (NE$CE)
21 NE$ED=as.numeric (NE$ED)
22 NE$EF=as.numeric (NE$EF)
23 NE$EI=as.numeric (NE$EI)
24 NE$EJ=as.numeric (NE$EJ)
25 NE$D=as.numeric (NE$D)
26
27 levels(NE$AD)
28 index = createDataPartition(NE$AD, p = .8, list = FALSE, times = 1)
29 train =NE[index,]
30 test= NE[-index,]
31 CrossTable(NE$AD)
32 CrossTable(train$AD)
33 CrossTable(test$AD)
34 model = as.formula(AD ~.)
1:1 (Top Level) R Script
```

Listing 6.4 R codes for deep MLP (continued)



```
Artificial Psychology.R
Source on Save | Run | Source
35
36 NE.net <- neuralnet(model,   data=train,  hidden=c(10,10),
37                     rep = 5,
38                     act.fct = "logistic",
39                     err.fct = "ce",
40                     linear.output = F,
41                     lifesign = "minimal",
42                     stepmax = 1000000,
43                     threshold = 0.001)
44
45 plotnet(NE.net,    alpha.val = 0.8,
46        circle_col = list('purple', 'green', 'brown'),
47        bord_col = 'black')
48 NE.prediction <- compute(NE.net, test)
49
50 idx <- apply(NE.prediction$net.result, 1, which.max)
51
52 predicted <- as.factor(c('High', 'Low' )[idx])
53 confusionMatrix(predicted, test$AD)
1:1 (Top Level) R Script
```

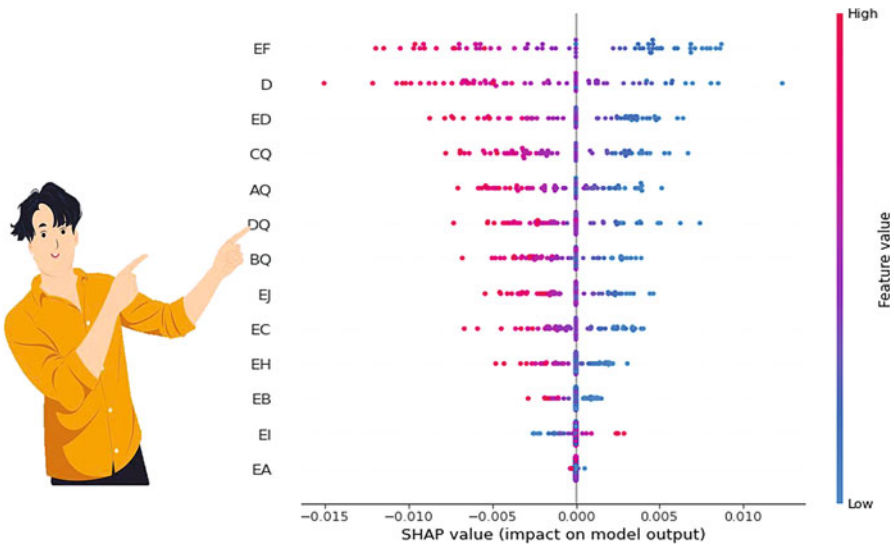


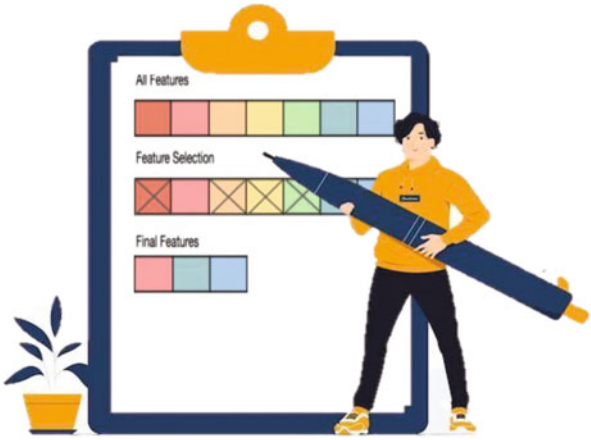
Fig. 6.38 SHAP values plot of deep MLP classifier

Fig. 6.39 The force plot of the model

Each point of every row is a record of the test dataset. The features are sorted from the most important (top) to the less important. We can see that EF is the most important feature. The higher and more positive the SHAP value of this feature, the more influential the impact on the classification accuracy of the target. The more negative this value, the less importance the variables have on classification accuracy (Fig. 6.38).

A very useful plot we can draw is called a force plot. The force plot is an explanation of feature importance based on the SHAP values. The force plot shows the influence of each feature on the current prediction. Values in red can be considered as the values that have a positive influence on the prediction pushing the values higher than the average value across all the cases without any variables in the model (baseline), whereas values in green have a negative influence on the prediction pushing the prediction lower than the baseline prediction (Fig. 6.39).

Chapter 7
Feature Selection in AP



7.1 Feature Selection Problem

In most real-world problems, a large fraction of the sample features are not useful and have a negative impact on model performance. Therefore, in machine learning, it is important to identify good features that affect model accuracy. Feature selection methods aim to find the smallest possible subset of features that is necessary and sufficient for determining the target. In fact, the distribution of the selected features should be close to the distribution of the main class (Dash & Liu, 1997).

Since this is an optimization problem, choosing the right feature selection method is in itself a refinement.

H. Farahani et al., *An Introduction to Artificial Psychology*,
https://doi.org/10.1007/978-3-031-31172-7_7

Advantages of Feature Selection
Feature selection has several advantages:

- Better classifiers: Since most redundant features generate noise and affect model accuracy, the main advantage of feature selection is improved accuracy.
- Better knowledge discovery: Feature selection also provides insights into the data. Separating important and influential features from unusable ones can be very informative.
- Reducing the cost of data collection: Where data collection is costly, such as in medical applications, identifying a minimal set of features for classification saves money.
- Computational costs: Identifying appropriate subsets of features simplifies the model and reduces computational and implementation costs.
- Dimensionality: Existing theories suggest that a higher number of features exponentially increases the volume of data required to build a classifier (Cunningham et al., 2021).

7.2 Feature Categorization

In general, features are divided into three subgroups:

1. Related Features: These features directly affect the output and target and cannot be replaced by other features.
2. Unrelated Features: These features are unrelated to the output and create random values in the sample.
3. Feature Redundancy: In this case, one feature replaces another. Therefore, these features do not provide any additional information.

According to these definitions, feature selection should identify features with high correlation, while redundant features should be removed as much as possible (Dash & Liu, 1997).

7.3 General Procedure of Feature Selection

Typically, there are four stages and functions in the process of feature selection, which are shown in relation to each other in the following diagram.

1. Generation procedure.
2. Evaluation function.
3. Stopping criterion.
4. Validation procedure (Tang et al., 2014) (Fig. 7.1).

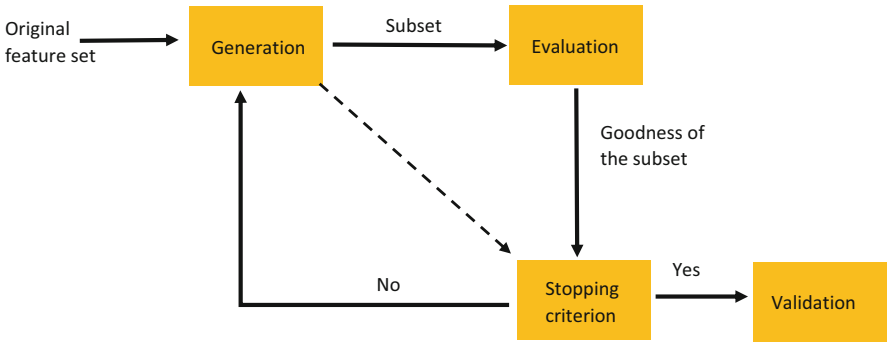


Fig. 7.1 Stages of feature selection with validation

Generation Procedure
The generator function is actually a search function that generates various subsets and is checked by the evaluation function. In this case, it can start from the initial featureless state, with a set of features, or with a random subset.

Evaluation Function
After generation, the subset of features is evaluated using different methods. In this phase, the goodness of a set is examined and the new subset is compared with the old one. If it is better, it replaces the older subset. In this case, finding a suitable subset of features depends directly on the evaluation function used because if the evaluation function does not provide a suitable subset, the subset of features will never find an optimal value.

Stopping Criterion
The criterion needed to end and stop the algorithm can be one of the following:

1. A certain number of features.
2. A certain number of iterations.
3. When adding or removing a feature does not result in a better subset.
4. When the optimal subset is reached based on the evaluation function.

Validation Function
Although this function is not part of feature selection, it checks the validity of the subsets. The results of the data can be evaluated based on prior knowledge (Ansari, 2021).

7.4 Feature Selection Methods

In many studies, feature selection methods are divided into three categories (Ansari, 2021):

1. Wrapper Method: This method is considered a prediction method, and its performance is crucial for subset selection. At each stage, a subset of features is selected, based on which the performance of the machine learning algorithm is evaluated and the performance result is used for feature selection.
2. Embedded Method: This method simultaneously uses machine learning algorithms with model fitting for feature selection.
3. Filter Method: This method measures the importance of features without machine learning algorithms. In this method, features related to the input are selected to obtain the output. The filter method typically employs the ranking method. Different features are ranked in the order of their acceptability. Due to its simplicity, the ranking strategy can be applied to any application. Before the features enter the classification phase, they are ranked, i.e., filtered. Incompatible elements are filtered into compatible elements. Each ranked element must have a unique feature to identify that class. This method is faster than packing methods, but may give poor results if the data is not properly correlated. These methods are also used when the spatial dimensions of the feature are large and the computational cost of the wrapper method is high, so it is more economical to use the filter method (Fig. 7.2).

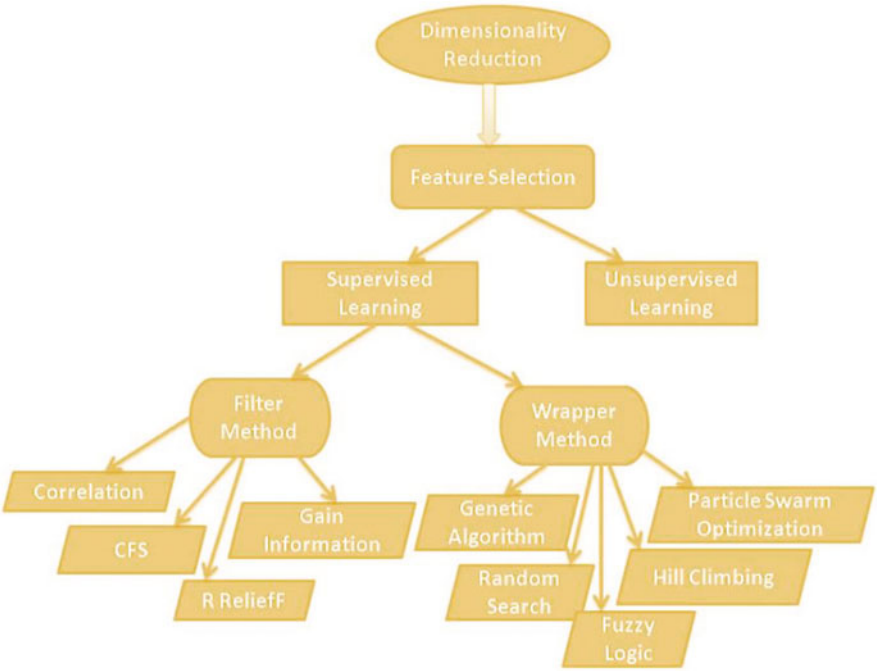


Fig. 7.2 Classification of feature selection methods. (From Ansari, 2021)

Genetic algorithm, a wrapper method, will be discussed in detail below.

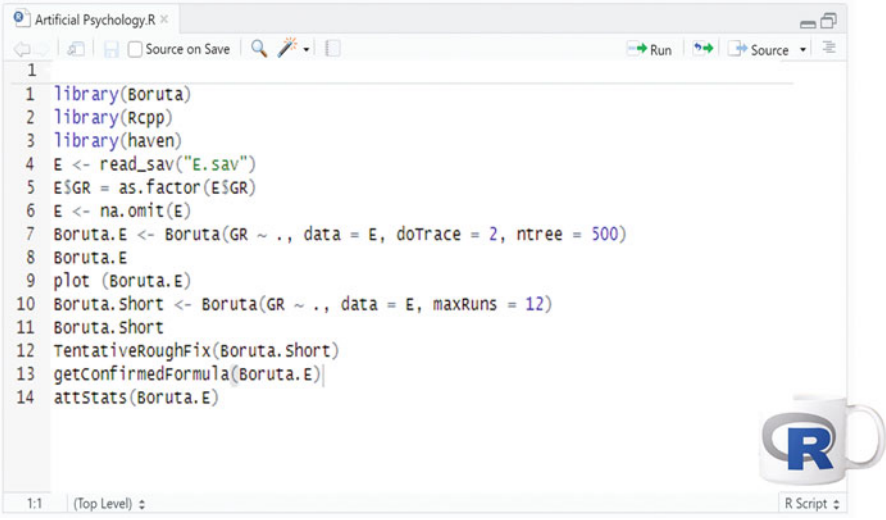
This section will first introduce the basics of the genetic algorithm and its feature selection stages.

7.4.1 Practical Example Using R

An artificial psychologist tries to implement a classification model, but he realizes that the number of predictor variables may decrease the accuracy of the model. It means some variables may be noisy variables; therefore, he decides to select the Feature Selection feature to determine a subset of the most important variables. In this model, he decides to first determine the most important features in predicting group membership with the Feature Selection algorithm. He wants to determine which of the 13 variables A, B, C, D, E, F, G, K, L, M, N, P, and Q are the most important variables. He uses the Boruta method. This is a wrapper method based on a random forest algorithm. The random forest itself is based on the decision tree algorithm. A decision tree is a sequence of steps that are performed in the training phase. Random forest is an ensemble of decision tree algorithms. The random forest algorithm trains hundreds of decision trees, each of which only has access to a random set of columns in the data (variables).

Boruta is a technique that takes this randomization much further. Based on this method, it captures all the features in the data that are related to the target variable (Kursa & Rudnicki, 2010).

Listing 7.1 R codes for feature selection implement



Listing 7.1 shows the R codes for feature selection implemented using the Boruta algorithm. The results are shown in Fig. 7.4. This plot shows that the variables *M, Q,*

	meanImp	medianImp	minImp	maxImp	normHits	decision
A	1.73690394	1.73837641	-0.9757289	3.797220	0.14141414	Rejected
B	5.35887887	5.21626185	1.5729358	9.002829	0.90909091	Confirmed
C	7.53427057	7.62499916	4.6332685	10.645457	0.98989899	Confirmed
D	3.44820270	3.51926105	-0.9915755	6.008792	0.60606061	Tentative
E	0.30376909	0.57438863	-1.9086028	1.437561	0.00000000	Rejected
F	3.09041276	3.13917899	-0.4840725	6.044032	0.52525253	Tentative
G	0.04440417	0.07386071	-1.7392264	1.162139	0.00000000	Rejected
K	0.71892061	0.55106261	-1.1037523	2.680069	0.00000000	Rejected
L	7.84496301	7.84846559	4.8959371	11.189794	0.98989899	Confirmed
M	4.69333423	4.79831467	0.5045418	7.271544	0.81818182	Confirmed
N	6.67107067	6.57505858	3.7858340	9.679785	0.95959596	Confirmed
P	0.59910742	0.44061058	-2.3356909	3.397893	0.03030303	Rejected
Q	5.24877043	5.15757572	2.4849780	7.702762	0.89898990	Confirmed

Fig. 7.3 R output of feature selection using the Boruta algorithm

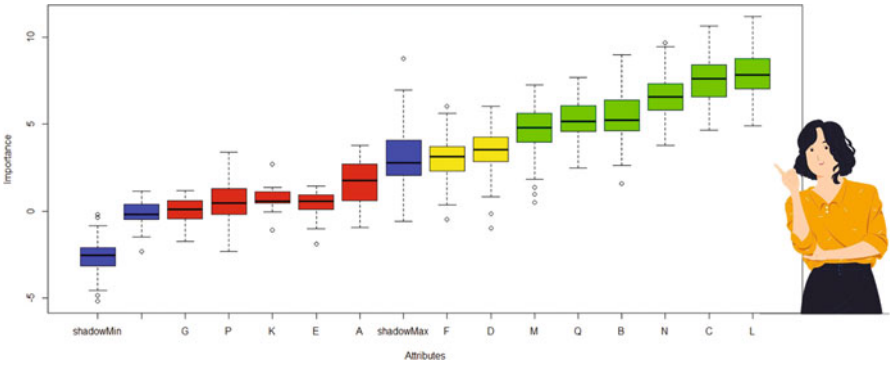


Fig. 7.4 R output of feature selection using the Boruta algorithm

B, N, C, and *L* are the most important variables in determining group membership, respectively.

Figure 7.3 shows the R output, where meanIMP shows the average importance, which is the highest average importance of each predictor to show how it is related to group classification. The last column is about making a decision about which variable remains in the model or is removed. This column has three options: confirmed, rejected, and tentative, which is the last option that includes the variables that need further investigation (Fig. 7.4).

Example 7.1 An artificial psychologist wants to predict the post-corona depression rate (y) based on variables A1 to A10. He performs feature selection based on the random forest algorithm using the DALEX Package in R and prepares the necessary codes (Listing 7.2). This is a powerful package. By using the function variable-dropout (), which determines the importance of a variable based on a dropout loss, how much loss is imposed on the model by removing a variable.

Listing 7.2 R codes for feature selection implement using random forest

```
library(randomForest)
library(DALEX)
rf_mod <- randomForest(Y ~., data=S, ntree=100)
rf_mod
explained_rf <- explain(rf_mod, data=S, y=S$Y)
varimps <- variable_importance(explained_rf, type='raw')
print(varimps)
plot(varimps)
```

As the importance plot and the output of the software in Figs. 7.5 and 7.6 show, variables A5, A7, A4, A2, and A10 are the five most important variables of Feature Selection.

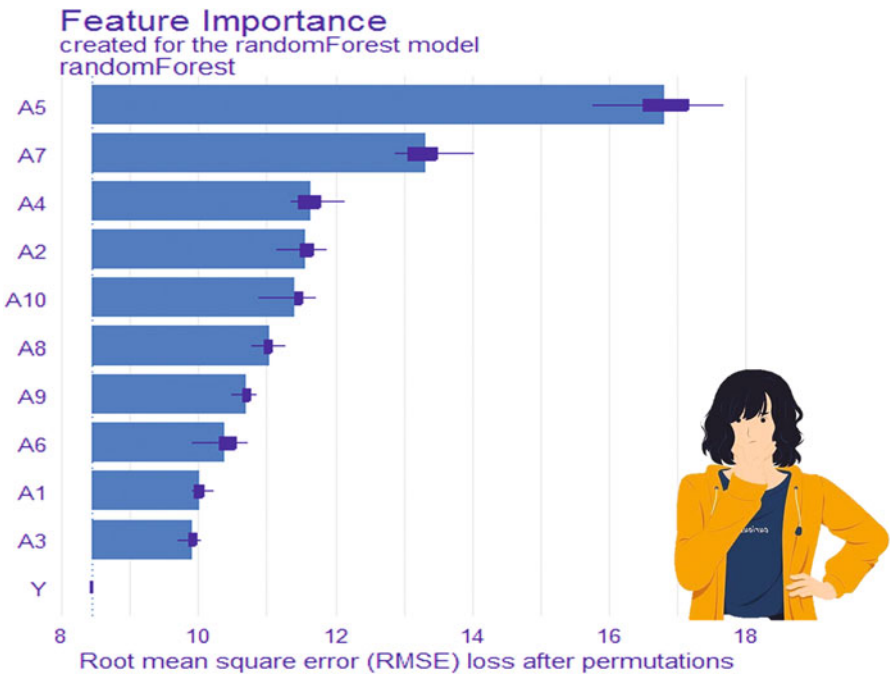


Fig. 7.5 Importance plot

	variable	mean_dropout_loss	label
1	_full_model_	8.432868	randomForest
2	Y	8.432868	randomForest
3	A3	9.900159	randomForest
4	A1	10.016357	randomForest
5	A6	10.376602	randomForest
6	A9	10.691588	randomForest
7	A8	11.028223	randomForest
8	A10	11.411271	randomForest
9	A2	11.553415	randomForest
10	A4	11.654499	randomForest
11	A7	13.317482	randomForest
12	A5	16.815061	randomForest
13	_baseline_	28.663382	randomForest

Fig. 7.6 Mean dropout – loss of each variable

```
-> model label       :  randomForest  (  default  )
-> data              :  250  rows  11  cols
-> data              :  tibble converted into a data.frame
-> target variable   :  250  values
-> predict function  :  yhat.randomForest  will be used (  default  )
-> predicted values  :  No value for predict function target column. (  default  )
-> model_info        :  package randomForest , ver. 4.7.1.1 , task regression (  default  )
-> predicted values  :  numerical, min =  24.79735 , mean =  90.35392 , max =  115.3006
-> residual function :  difference between y and yhat (  default  )
-> residuals         :  numerical, min =  -34.77893 , mean =  0.03848393 , max =  21.69607
```

Fig. 7.7 Model summary

Figure 7.6 shows that the mean loss is maximized if variable A5 is dropped from the model. This dropout loss is the highest value among the 10 predictor variables and is equal to 16.81.

As Fig. 7.7 indicates, the mean prediction error using this random forest model is 0.038.

7.5 Metaheuristic Algorithms

In general, optimization algorithms are divided into two categories: exact algorithms and approximate algorithms. Exact algorithms provide the optimal solution accurately but have exponentially longer resolution times. In approximate algorithms, on

the other hand, the solution is close to the optimum and takes less time for difficult problems.

Therefore, exact algorithms do not seem to be suitable for many problems. There are three types of approximate algorithms, namely heuristic algorithms, metaheuristic algorithms, and hyper-heuristic algorithms. In this book, we will discuss the genetic algorithm, which is a metaheuristic algorithm (Fraser, 1957).

In metaheuristic algorithms, the behavior is random, i.e., the search for the solution and the optimal value starts at a random point (i.e., from absolute ignorance to knowledge). The conceptually simple metaheuristic algorithms can be implemented easily and flexibly. In other words, the algorithms can be easily modified depending on the problem. The main feature of metaheuristic algorithms is their ability to prevent early convergence of algorithms (Golberg, 1989).

These algorithms have been successfully applied to various engineering and scientific problems, e.g., in electrical engineering (to find the optimal solution for power generation), industry (planning, transportation, vehicle routing, location determination), civil engineering (for bridge and building design), communications (radar and network design), data mining (classification, prediction, clustering, system modeling), and psychology and cognitive science (to reduce feature dimensionality in questionnaires and for features obtained from brain imaging) (Engelbrecht, 2007).

Metaheuristic algorithms are divided into the following two main categories:

Single Solution-Based Metaheuristic Algorithms: In these techniques, the problem typically starts with a solution that is updated in iterations. The problem is that one falls into the trap of local optimization and does not fully explore the search space.

Population (Multiple) Solution-Based Metaheuristic Algorithms: These algorithms generate a population of solutions and start the optimization process. The population is updated with the number of generations or iterations. These algorithms do not suffer from the problem of single solution-based algorithms; i.e., they do not fall into the local optima because multiple solutions help each other and there are many searches in the search space. Moreover, population-based algorithms are used for solving most real-world problems.

Due to the aforementioned properties, much attention is paid to metaheuristic algorithms. These algorithms are classified into four categories based on their behavior, namely, evolution-based, swarm intelligence-based, physics-based, and human-related algorithms (Agrawal et al., 2021) (Fig. 7.8).

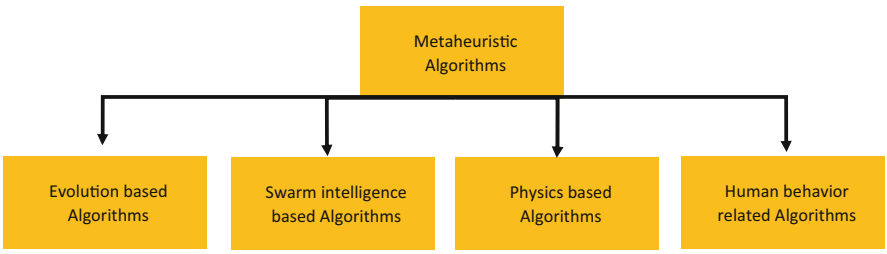


Fig. 7.8 Four categories of metaheuristic algorithms

Evolution-Based Algorithms are inspired by natural evolution and start with a population of random solutions. They combine the best solutions to create new individuals through mutation, combination, and selection of the best population. The genetic algorithm (GA) based on Darwin's evolution is the most popular in this category. This category also includes other algorithms such as evolution strategy, genetic programming, tabu search, and differential evolution (Agrawal et al., 2021).

Swarm Intelligence-Based Algorithms are inspired by the social behavior of insects, animals, fish, or birds, such as the group behavior of birds and how they fly around the search space and find their best location. Other swarm intelligence algorithms include ant colony optimization, the honey bee swarm optimization algorithm, monkey optimization, and more (Agrawal et al., 2021).

Physics-Based Algorithms: These algorithms are inspired by the laws of physics in the world and include simulated annealing (Goldberg & Richardson, 1987), harmony search (Higashi & Iba, 2003), and so on.

Human Behavior-Based Algorithms: These algorithms are purely inspired by human behavior. The approaches humans take to various activities affect their performance. Popular algorithms include the teaching learning-based optimization algorithm (TLBO), the League Championship algorithm, and more (Agrawal et al., 2021).

As mentioned earlier, metaheuristic algorithms mimic the principles of natural evolution to search for the optimal solution. The genetic algorithm is a metaheuristic algorithm rooted in Darwin's theory of evolution. In this theory, a better generation typically emerges from the combination of chromosomes. In the meantime, possible mutations can improve or worsen the generation. The genetic algorithm is explained in detail below.

7.6 An Introduction to the Genetic Algorithm

Many human inventions are inspired by nature, as its evolution over a long period of time means that it always provides the best and most optimal solutions and processes. Given the ever-changing environment around us, the ability to change and adapt to conditions is necessary for any living thing to survive. Phenomena such as natural selection, reproduction, mutation, and symbiosis are involved in this adaptation and evolution. Living organisms also rely on chemical elements to store information about each organism. DNA contains all the necessary information for the reproduction of a living organism and is the known regulator of genetic information. In fact, genetics is the alphabet of nature. Hereditary factors and individual characteristics exist along the chromosome. The human body usually has four bases

Fig. 7.9 Illustration of crossover child and mutation child

for making chromosomes from DNA, namely bases A[1], C[2], G[3], and T[4]. DNA also has four combinations in DNA, namely TA, AT, GC, and CG, which can produce different phenotypes. Consequently, there is a one-to-one correspondence between each living organism and its gene sequence, which is called coding.

The genetic algorithm is a family of "computing models" that are conceptually based on evolution. This field in artificial intelligence is inspired by the evolution of natural living organisms. The genetic algorithm is based on the principle of survival of the fittest and reproduction of the superior. It is often considered a "function optimizer"; that is, it is used to optimize the objective functions in problems. In fact, however, it has other applications (Fig. 7.9).

7.7 Basics of the Genetic Algorithm

Definition of Gene: In the genetic algorithm, a gene is a parameter of the problem and the genotype is considered as a bit string with a fixed length. The length, which must be specified before optimization, shows the dimensions of the problem and can sometimes be considered as a variable or vector (Wang & Jiang, 1994).

Definition of Chromosome: A set contains all genes that have been valued and represent a possible solution to the problem. Answers, suitable or unsuitable, are actually a string or sequence of bits that act as natural genes in chromosomes. The coding in bits can be binary or n. In nature, a chromosome is a long and complex string called "deoxyribonucleic acid" or DNA.

Genetic Population: A genetic population is a set of chromosomes. Instead of focusing on one point or one chromosome, the genetic algorithm operates on a population of chromosomes that has more desired features than those of the previous stage. The population size indicates the number of chromosomes within a generation. If the population size is too small, the algorithm performs poorly; if it is too large, it slows down. The population size is usually between 20 and

[1] Adenine
[2] Cytosine
[3] Guanine
[4] Thymine

30 chromosomes and can be as large as 50–100 depending on the problem. Note that increasing the population size beyond a certain limit does not help solve the problem faster (Wang & Jiang, 1994).

Fitness Function: The genetic algorithm has a stage to evaluate the obtained answers and determine their value. Namely, the suitability of the answer is checked with the fitness function, with more suitable answers having a higher fitness value. Suitable solutions with a higher chance of survival have higher fitness values. They have a greater probability of producing children and more sequences. Therefore, optimal chromosomes have a greater chance of joining with other chromosomes.

Evolutionary Operators

In genetics, evolutionary operators include reproduction, mutation, and selection.

Reproduction: This is the reproduction of genetic information between chromosomes, which means sharing information between chromosomes. In this operator, children inherit the characteristics of their parents, which are then reproduced to create a better generation. Reproduction usually occurs between a pair but can also occur in several parents. In this case, the traits of more than two parents are passed to the child. This operator improves fitness and its implementation is related to chromosome coding.

Mutation: The mutation operator generates and introduces new genetic material and contributes to the diversity of the population, which can be achieved by adding a random value. In nature, mutation does not usually produce good results, but it is necessary for continuous evolution, and the results may eventually include a favorable outcome.

Selection: Selection is the stage of choosing the parents and ideal promotion of the good gene. In selecting the best individual, all individuals are included in the population based on merit, and individuals with greater fitness are marked and selected for the next generation.

7.8 The Initial Design of the Genetic Algorithm

Now that we have explained the basics, it is time to implement the genetic algorithm. First, the variables should be introduced according to the problem, i.e., the encoded chromosomes. Then, a fitness function for the chromosomes is considered based on the objective function. This algorithm first generates a population of chromosomes (the initial population is randomly selected). Then, the fitness of each chromosome is checked and selected according to the following stages.

Stage One: An appropriate number of chromosome pairs are selected for the next stages based on their fitness.

Stage Two: Parents are selected and reproduced to produce a population of children. In this stage, the reproduction operator is applied. The chromosomes change through the reproduction process. The parent chromosomes are randomly exchanged through the crossover process. Therefore, the children have some of the characteristics of their parents.

Stage Three: The members of the population are selected to apply the mutation and create the mutation population. As mentioned earlier, there is also the rare process of mutation that changes the characteristics of living organisms. For example, there could be an error in the chromosome copying process, mitosis. Often mutations destroy living organisms, but in the long run they create new and better species. Mutation often creates traits that are certainly not possible in reproduction. In other words, this is the way of entering new information.

Stage Four: The original population is merged with the population of children and mutants to create a new original population. Now the fitness of the new children is calculated.

In stage five, if the stopping conditions are not met, the process is repeated from stage two and the new population enters the next stages as the initial population. In this case, the generated data structure (chromosomes) is evaluated. The chromosomes that can better represent the optimal solution or goal of the problem have a greater chance of reproduction than others; i.e., they are given more opportunities.

Stopping Conditions

1. A threshold or optimal value is considered for the cost function.
2. There is also a criterion for the number of iterations or one relative to time.

The stall iteration is stopped if it exceeds a limit. In this case, when changes are less than a specified level and the result is not better, the process stops and starts from another point (Sahdra et al., 2016) (Figs. 7.10 and 7.11).

7.9 Feature Selection Using the Genetic Algorithm

The genetic algorithm is implemented to select optimal features. This is a robust machine learning method that limits the number of features without significant loss of information.

Conventional scale abbreviation methods must manually check multiple criteria to select elements. However, the genetic algorithm is a fully automatic and complex optimization tool. Fortunately, its software implementation is relatively simple (e.g., the Precis tool in Python and the GAabbreviate tool in R) (Sahdra, et al., 2016).

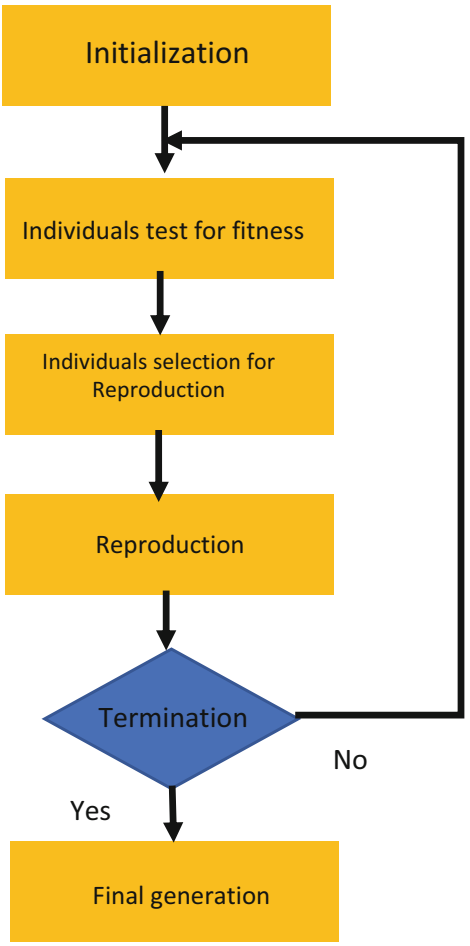


Fig. 7.10 Diagram of genetic algorithm implementation

```
Iter = 334   | Mean = 0.275287   | Best = 0.2438558
Iter = 335   | Mean = 0.2719858  | Best = 0.2438558
Iter = 336   | Mean = 0.2702471  | Best = 0.2438558
Iter = 337   | Mean = 0.2716116  | Best = 0.2438558
Iter = 338   | Mean = 0.2786182  | Best = 0.2438558
Iter = 339   | Mean = 0.2914065  | Best = 0.2438558
Iter = 340   | Mean = 0.2977796  | Best = 0.2438558
Iter = 341   | Mean = 0.2917912  | Best = 0.2438558
Iter = 342   | Mean = 0.2822553  | Best = 0.2438558
Iter = 343   | Mean = 0.2734095  | Best = 0.2438558
Iter = 344   | Mean = 0.2689914  | Best = 0.2438558
Iter = 345   | Mean = 0.2651342  | Best = 0.2438558
```

Fig. 7.11 Output of genetic algorithm implementation

7.10 The Genetic Algorithm's Application in Artificial Psychology

Self-report criteria are a common form of assessment in psychology. Although not entirely accurate and appropriate because of substantial bias due to social desirability, arbitrary criteria, and interference with these reports, these reports are very common, simple, and inexpensive to use. At the same time, many questionnaires are long and time-consuming, which leads to people not completing the questionnaires accurately and faithfully. In contrast, with short questionnaires, people are more likely to fill them out honestly. In one meta-analysis, the same short and long questionnaires were randomly distributed to people, and researchers found that participants who received a shorter questionnaire had a greater return rate (Rolstad et al., 2011). Another study found an inverse correlation between questionnaire length and response rate (Fan & Yan, 2010).

In addition, shorter questionnaires can help develop science through description, prediction, and causal inference. In fact, lower numbers of variables can be useful in predicting theories and examining cause and effect. Thus, although short questionnaires are more interesting, in many cases they do not meet psychometric standards. In fact, traditional methods of shortening questionnaires require careful consideration of competing factors. This approach requires finding items with high autocorrelation, low cross-loading, low correlated uniqueness, low chance of missingness, high face validity of construct coverage, and high internal consistency of the resulting scale.

Reducing the 15-item questionnaire to 6 items results in over 2500 possible permutations that must be considered. Because of this complexity, researchers must rely on unwritten heuristics that often result in poor performance.

One efficient method is to use machine learning to find items that explain the greatest diversity. The genetic algorithm has recently been used to shorten questionnaires in several domains, including personality (Yarkoni, 2010), values (Sandy et al., 2014), psychopathy (Eisenbarth et al., 2015), experiential avoidance (Sahdra et al., 2016), and body image-acceptance (Basarkod et al., 2018).

In the algorithm, the items are like genes and a set of items that make up a scale are like chromosomes. As explained above, the algorithm first generates a sample of chromosomes that forms a random set of subscales from the whole. Then, the chromosomes compete for fitness. The fitness scale in shortening a questionnaire includes chromosomes that explain more variability in the full questionnaire. Then, the selection operator is used to remove the chromosomes with the lowest fitness from the gene pool and reproduce those with the highest fitness (usually the top 5%) to create a new set of chromosomes (Yarkoni, 2010).

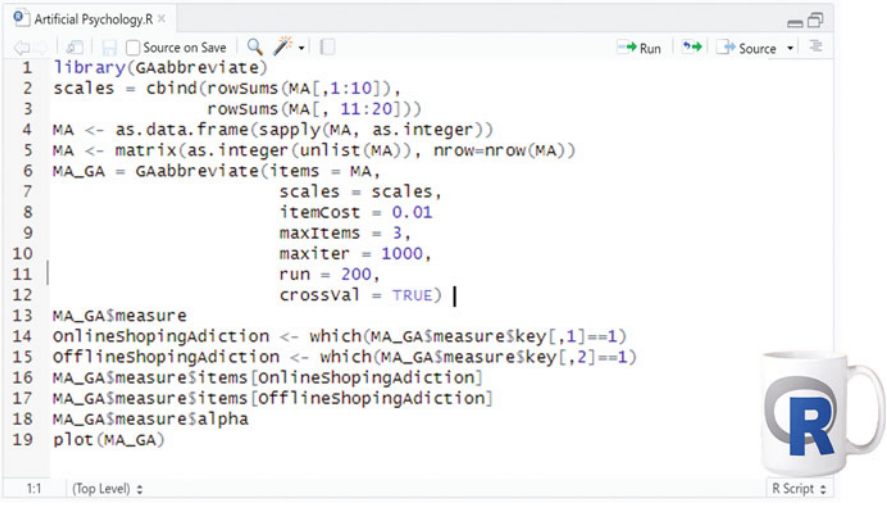
As mentioned in the stages of the genetic algorithm, chromosomes also mutate and genes crossover. Mutation replaces random elements in the chromosome. In reproduction, the genes of one fit item are exchanged with the genes of another. After mutation and reproduction, the new set of chromosomes is generated in terms of fitness and examined according to the fitness function. This process is repeated until

the algorithm finds a stable solution. Items shortened by the genetic algorithm often yield scores that are as valid as those derived from traditional psychometric approaches (Sandy et al., 2014; Noetel et al., 2019).

7.10.1 Practical Example Using R

An artificial psychologist tries to select the best items based on the genetic algorithm from a psychological scale for measuring sexual pain that has 20 items with two 10-item subscales. A classical psychologist with psychometric techniques can select the best items based on corrected item-total correlations, factor loadings, and the correlation of each item with a criterion or using the discriminant coefficient in Item Response Theory (IRT). Although this works, the artificial psychologist seeks to select the important items with the genetic algorithm in such a way as to keep the cost function to a minimum. The cost function in this case was proposed by Yarkoni (2010) and explained in the previous sections of this chapter. The genetic algorithm iteratively tries to select and mate strong individuals (solutions) who are more likely to survive.

Listing 7.3 R codes to implement the genetic algorithm



Listing 7.3 shows the R codes to run the genetic algorithm. Item cost indicates the cost of each item, which is equal to 0.05 by default. Max Items is the maximum number of items that we want to remain in each subscale or factor in the end. In this example, the artificial psychologist considers it to be 3 items, maxIter indicates the maximum number of iterations, which in this case is set to 1000 items and if this function is not able to find a solution, its number can be increased. Cross-validation

```
$nItems
[1] 6

$key
     [,1] [,2]
[1,]    1    0
[2,]    1    0
[3,]    1    0
[4,]    0    1
[5,]    0    1
[6,]    0    1
```

```
> MA_GA$measure$items[OnlineShopingAdiction]
x2 x5 x7
 2  5  7
> MA_GA$measure$items[OfflineShopingAdiction]
x14 x15 x19
 14  15  19

          Scale1    Scale2
alpha 0.7677823 0.7142635
```

Fig. 7.12 Output of genetic algorithm implementation

is also done for unbiasedness. If the GA abbreviate function finds the optimal solution before the maximum number of iterations, it will stop automatically; otherwise, the iteration continues until it reaches an optimal solution. Figure 7.12 shows which items belong to which dimension of the factor or scale. It is marked with $ key in the output.

As Fig. 7.12 shows, this algorithm has reached the optimal solution in the iteration equal to 0.24, and the GA abbreviate function terminates.

The output of this method shows that items 2, 5,7, 14, 15, and 19 are selected (3 items for each subscale) (Fig. 7.12).

Figure 7.13 shows a visual summary of the search process using the plot function. Diagnostic plots, which are the three graphs on the left, show how the total cost and the length of the scale and the average value of R^2 change during the search process. The middle graph of this figure shows the percentage of explained variance, that is, R^2 for the best solution, which in this example is excellent and close to 0.9 for each subscale.

And finally, the plot on the right side of the figure shows which items are selected during the search process. Items 2, 5, 7, 14, 15, and 19 are the items that were finally selected. You can easily see them in this plot (Kursa & Rudnicki, 2010).

7.11 The Genetic Algorithm's Application in Neural Network Sciences

In a new approach, there are studies for identifying biomarkers from the functional matrix formation strategy (see Chap. 5). Functional matrices are considered a good method for diagnosing many mental disorders, such as schizophrenia, autism, and MCI. After obtaining the functional matrix, it is necessary to use feature selection techniques, one of which is the genetic algorithm discussed here. In fact, rs-fmri data are used and a functional matrix is provided based on time series to detect abnormal patterns even from a younger age, which can be very beneficial for early diagnosis and treatment.

One of the main challenges in fMRI analysis is the high dimensionality of data. Although data from the functional connectivity matrix (FCM) provide

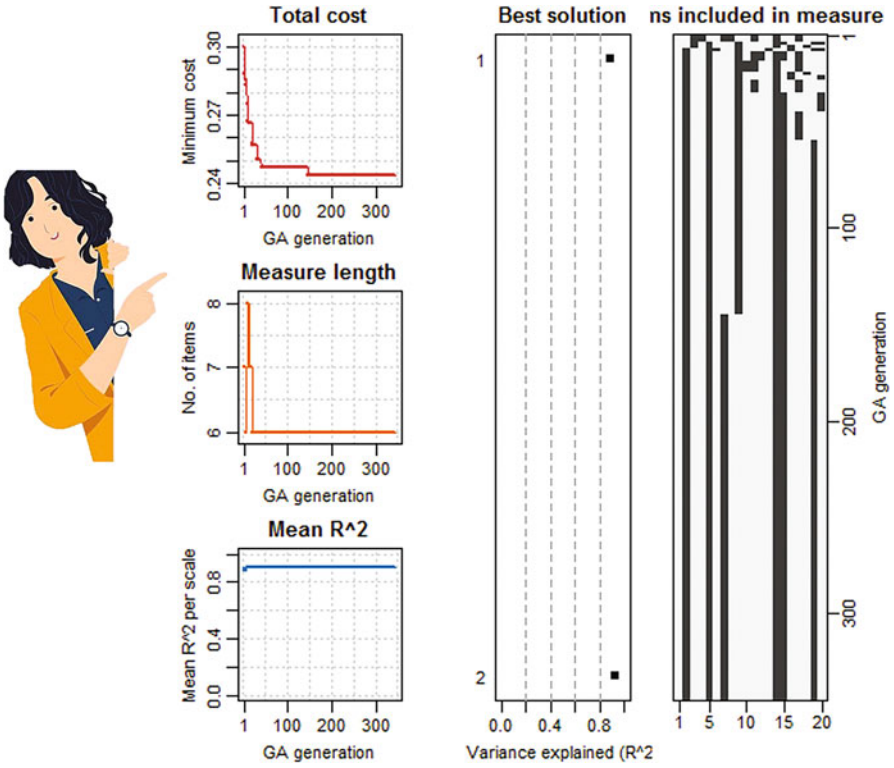


Fig. 7.13 Plot of the search process

comprehensive information about connections between different regions, the high dimensionality of feature space poses several challenges in classification and analysis. Feature selection is hence considered a critical issue in biomarker detection. Feature selection aims to reduce dimensionality, accelerate operation, increase the accuracy of classification algorithms, and better understand the results.

Mapping the brain's functional connections is used for determining the function and correlation of time series between different brain regions. Brain FCM is determined based on correlation coefficients between regions. Here, for instance, the brain is divided into 116 regions. As a metaheuristic optimization algorithm, the genetic algorithm considers a set of feature subsets in each iteration and can find an optimal or near-optimal feature subset.

As mentioned, the reduction of data dimensionality using a genetic algorithm includes chromosome coding, selection, reproduction, and mutation. Each chromosome represents an array of zeros and ones, i.e., the total number of features (length 1×116). The values 0 and 1, respectively, denote the absence and presence of the respective region in the FCM calculation (Sadeghian et al., 2021) (Fig. 7.14).

A set of the initial population is created and the population's fitness is evaluated based on the fitness function. For example, here, the KNN classifier's accuracy

Fig. 7.14 Binary coding of chromosomes. (Sadeghian et al., 2021)

based on the selected regions can be considered a fitness function. Reproduction is applied to a pair of parents. After producing a member in a new population, some genes mutate randomly, which allows for searching the entire feature space and escaping the local optima. Finally, the genetic algorithm stops after a specific number of iterations, and the solution with the highest classification accuracy determines the final feature subset. This method has been implemented for distinguishing autism patients from healthy individuals, removing 49 regions to improve the genetic algorithm's diagnostic accuracy by 9%.

The GAabbreviate uses the GA package (Scrucca, 2013) to efficiently implement Yarkoni's (2010) scale abbreviation cost function:

$$\text{cost} = Ik + \sum_{i=1}^{s} w_i\left(1 - R_i^2\right) \tag{7.1}$$

where I represents a user-specified fixed item cost, k represents the number of items retained by the GA (in any given iteration), s is the number of subscales in the measure, w_i are the weights (by default $w_i = 1$ for any i) associated with each subscale (if there are any subsets to be retained), and R_i^2 is the amount of variance in the ith subscale that can be explained by a linear combination of individual item scores. Decreasing or increasing the value of yields longer or shorter measures, respectively. When the cost of each individual item retained in each generation outweighs the cost of a loss in explained variance, the GA yields a relatively brief measure. When the cost of each individual i term is low, the GA yields a relatively longer measure maximizing explained variance (Yarkoni, 2010).

Chapter 8
Bayesian Inference and Models in AP

8.1 Bayesian Inference and Models in Artificial Psychology

Bayesian inference includes Bayesian parameter estimation and Bayesian hypothesis testing. In recent years, this approach has been proposed as an attractive alternative to estimation and hypothesis testing in classical statistics.

In classic Fisherian or Frequentist statistics, confidence interval and p-value are used for estimation and hypothesis testing (Wagenmakers et al., 2018).

H. Farahani et al., *An Introduction to Artificial Psychology*,
https://doi.org/10.1007/978-3-031-31172-7_8

Fig. 8.1 A boxing analogy of the p-value. (From Wagenmakers et al., 2018)

The background of psychological research is full of p-value reports (Farahani & Azadfallah, 2020). It can be said that the use of p-values causes a crisis of confidence in the results of psychological researches that are mainly conducted using classical statistics; in other words, p-value hacks the results of psychological research. The frequent use of p-value in Null Hypothesis Statistical Testing (NHST) has been seriously criticized by a large number of researchers repeatedly from different points of view (Wagenmakers et al., 2018) (Fig. 8.1).

$$p(c|X) = p(x_1|c) \times p(x_2|c) \times \ldots \times p(x_n|c) \times p(c) \tag{8.1}$$

8.2 Bayesian Statistics in a Nutshell

A quick look at Bayesian statistics may help ease the concepts of this section and other sections of this chapter. Suppose an artificial psychologist examines 100 boys with ADHD and finds that 43 of them are the first child in the family. Therefore, the probability that a boy is the first child in the family is equal to $p(\text{first order}) = \frac{43}{100} = 43\%$.

Here we can talk about two other terms in probability, and they are dependent and independent probabilities. If the occurrence or nonoccurrence of a phenomenon has no effect on the occurrence or nonoccurrence of another phenomenon, then the two phenomena are independent and $p(A \cap B) = 0$, for example, the probability that a child with ADHD is the first child in the family and the probability of catching a

Table 8.1 Patients' memory problems based on sex in MS patients

	Female	Male
Severe	8	12
Memory problems		
Mild	32	48

goldfish in a big river in Hawaii. These two phenomena are completely independent of each other because catching a large goldfish in the Hanalei River in Hawaii has no effect on the probability of a child having ADHD as a first child, and vice versa. If the probability of catching a large goldfish in that river in Hawaii is 10%, then the two probabilities must be multiplied together to calculate the coincidence.

$$P(A) \times P(B) = (\text{fishing (B) and ADHD child becoming the first child (A)})$$
$$4.3\% = 10\% \times 43\% \tag{8.2}$$

If two phenomena are supposed to be dependent, then a different approach is used to calculate the probability. Pay attention to this example: suppose that the artificial psychologist examines 100 patients with MS, he also examines the memory and gender of these 100 people. The results are shown in Table 8.1.

Based on this table, $P = 0.2$ 12/60 $(A\big|B) =$ in which ($A =$ severe, $B =$ male) can be divided into the intersection of problems, having severe memory of A and the gender of the patient being B being male.

$$P(A|B) = \frac{p(A \cap B)}{p(B)} \tag{8.3}$$

Similarly,

$$P(B|A) = \frac{p(A \cap B)}{p(A)} \tag{8.4}$$

Bayes' rule, which is used for NB, can be considered as the result of dividing these two expressions. Therefore, Bayes' rule can be considered as follows:

$$(A|B) = \frac{p(A \cap B)}{p(B)} \tag{8.5}$$

In this formula, $P(\text{Evidence}\big|\text{outcome})$ is obtained from training data.

$$P(B|A) = \frac{p(A \cap B)}{p(A)} \tag{8.6}$$

In this formula, $P(\text{outcome}|\text{Evidence})$ is checked, which is used to predict test data.

Bayes Rule

$$P(B|A) = \frac{p(A|B) * p(B)}{p(A)} \tag{8.7}$$

$$(\text{Outcome}|\text{Evidence}) = \frac{\text{Likelihood of evidence} \times \text{Prior}}{\text{Probability of evidence}} \tag{8.8}$$

In the above formula, the left part is called the Posterior Probability or simply posterior, and the first term is called the likelihood of evidence, which is actually the conditional probability for a particular class, and provided that the predicted variables are independent, all of them can be multiplied. The likelihood value is determined from the training data for $Y = c$ (a specific group) and the second part, the prior, shows the overall probability of $y = c$ where c is a class of Y.

$$\text{Prior} = \frac{\text{frequency } (y = c)}{n} \tag{8.9}$$

With this introduction, you can get a general understanding of the necessary terms in NB, such as conditional probability, Bayes' rule, independent and dependent probability, and posterior, prior, and likelihood.

The most important design parameter in NB is the smoothing method. The idea of smoothing goes back to the efforts called Cromwell's rule, and based on that, if the estimate of a probability is equal to zero, it should not be used in probabilistic reasoning because as discussed, to combine the probabilities, we multiply them together and, so, if one of them is zero, regardless of the probabilities of the other variables, it will be zero. The most common form of smoothing is called Laplace smoothing, in which the number of desirable cases (K) out of n trial attempts is considered as the desirable ratio $(k + 1)/(n + 1)$ and not as k/n.

Classical statisticians consider smoothing as a form of regularization and Bayesian statisticians consider smoothing as a prior.

8.3 A Critique on the Use of p-value

As stated, the classic or Fisherian statistic relies on the central core of the p-value in the test hypothesis. p-values are easy to obtain using routine software such as SPSS. Interpreting the p-value, however, is challenging. Concluding that $p < 0.05$ guarantees the rejection of the null hypothesis (H_0) and thus supports the acceptance of the alternative hypothesis (H_1) is a misinterpretation. Let's get a little more specific. The p-value indicates the probability of obtaining a result at least as large as the

observed result, provided that the null hypothesis is correct. Therefore, the p-value cannot recognize the fact that the data that are unusual under H_0 can also be unusual under H_1. However, the p-value is still of interest to psychologists. Some of the reasons for this interest is that most psychologists, like other people, are addicted to their own beliefs, so they tend to teach others what they have learned and do not take steps to change their statistical knowledge. In addition, it seems incorrect that the interpretation of p-value and $p < 0.05$ is enough to reject H_0 and confirm H_1. Psychologists may also worry about reducing their chances of publishing their research articles, if they use new methods. Perhaps these reasons are the general reasons why there is resistance to new statistical methods alternatives to the p-value (Sharpe, 2013).

This book aims to break down this resistance. To overcome the weaknesses of the p-value, researchers have made efforts to replace it. One of these attempts is to replace the confidence interval (CI) with the p-value. The confidence interval has also been criticized, such as the fact that it considers the real value to be estimated as a fixed value.

The 95% confidence interval (CI) for an effect shows that if the confidence interval is calculated repeatedly from the data, there is a 95% probability that the desired effect or parameter is in a given range. This interpretation is somewhat counter-intuitive. In Bayesian statistics, similar to the confidence interval, there is a credible interval, which shows that according to the observed data, there is a 95% probability that the desired effect falls within this domain. By examining the background of the research done on the benefits of Bayesian inference, it can be said that this method can be more useful with high dimensional data than the classical statistical method, and the information is more reliable. Bayesian approaches are more accurate in conditions where there is noisy data and there are small samples (Kruschke et al., 2012). Bayesian inference provides two possibilities for combining prior knowledge in the final analysis (Andrews & Baguley, 2013; Kruschke et al., 2012). Bayesian analysis also gives straightforward, intuitive results (Kruschke, 2014; Wagenmakers et al., 2018). Bayesian inference includes a measure of evidence that the data additionally provides in favor of H_0 versus H_1, the Bayes factor, which, unlike the p-value, does not have a serious bias against H_0 (Edwards, 1965; Sellke et al., 2001).

In summary, it can be said that the main focus of classical statistics compared to Bayesian statistics is that classical statistics is strongly focused on the statistical test of the null hypothesis (NHST) and the misinterpretation of the p-value, and this extreme focus causes serious criticisms and accompanying lack of trust in the results of psychological research.

The theoretical framework of Bayesian statistics is based on Bayes theory. In this book, our goal is not to focus on the theoretical basis that the reader can follow and study elsewhere. What is considered in the theoretical framework of Bayesian statistics is different from what is considered in the theoretical framework of classical statistics. In the theoretical framework of classical statistics, the focus is on hypothesis testing and the p-value, which assumes that the effects are fixed and unknown and that the data are random. That is, it is assumed that the unknown

parameter is a unique value that the researcher tries to estimate with statistical methods using the data obtained from the sample. In the framework of Bayesian statistics, the true effect is not estimated, but instead the probability of different effects is calculated according to the data obtained from the sample, which itself leads to the posterior distribution, a distribution of possible values for the parameters. In Bayesian statistics, indicators such as the median of the posterior distribution and the range of values of that distribution that includes 95% of the most probable values, the 95% credible interval, are calculated to model uncertainty in an estimated parameter. In classical statistics, point estimation and confidence interval are calculated.

In Bayesian statistics, using Bayesian sampling algorithms, which will be explained later, a possible (posterior) distribution is obtained from an effect that is compatible with the observed data.

It follows that, in Bayesian analysis, based on the resulting data and sometimes the prior belief or distribution about the results using Bayesian sampling algorithms, a possible distribution is produced called the posterior. For example, suppose we assume that the correlation between affective metallization and mental health is equal to 0.54 in a sample of 100 people. Based on this distribution, Bayesian analysis tells us that the most probable effect (correlation) is 0.54, but the data is consistent with the correlations of 0.74–0.85, each of which has specific probabilities. To determine the significance of an effect in Bayesian analysis, a p-value is not needed, despite it being commonly used in classical statistics, with it, instead, being sufficient to describe the posterior distribution of that effect. One of these very important indicators is the credible interval. The credible interval is a key concept in Bayesian inference and analysis, whose purpose is to provide a summary of the uncertainty related to the estimated parameter. A credible interval in Bayesian statistics is a range of the posterior distribution that includes the possible magnitudes of the investigated effect with certain probabilities. Instead of a 95% confidence interval that is common in classical statistics in Bayesian statistics, McElreath (2020) suggested using a threshold of 89% when specifying ranges of a credible region.

Kruschke in 2014 states that a credible interval with 89% coverage is more stable than one with 95% and that to calculate a credible interval with a 95% interval, it is necessary that the number of samples in Bayesian sampling is at least 10,000, which is routinely the default number. Posterior samples that are used in most Bayesian statistics software packages, on the other hand, usually only use 4000 samples.

In Bayesian analysis, just like classical statistical analysis, there are descriptive and analytical indicators that should be taken into account when reporting the results.

The median, in general, is a more robust index compared to the mean.

MAP is the maximum posterior probability estimate (MAP). The MAP index in a posterior distribution indicates the value that has the highest probability. The peak of the posterior distribution can be considered the mode of the posterior distribution. The median is more robust compared to this index, but if the distribution has extreme skewness, MAP is more appropriate than the median.

95% or 98% credible intervals (CI) also indicate uncertainty. Generally, a CI is calculated based on the highest density interval (HDI). The highest density interval

(HDI) produces an interval containing values that have the highest probability density and always contain the most likely value of the parameter value corresponding to the mode.

8.3.1 Significance or Existence of a Network

In Bayesian analysis, there is no p-value index, but there is a more interpretable index that is more straightforward. It is interpreted as describing the existence of an effect. This probability index is called the probability of direction (Pd) and indicates the most likely direction (positive or negative) of an effect. In the interpretation of this index, it is possible to have a cutoff point like a p-value. A Pd of $>97\%$ is indicative of a likely effect, a Pd of $>99\%$ suggests an effect probably exists, and a Pd of $>99.9\%$ indicates the effect exists with certainty. Region of Practical Equivalence (Rope) is a region that signifies values of a parameter estimate, such as the amount of change in an outcome, corresponding to denoting practically no effect. This indicator shows whether or not a parameter is related to, for example, a non-negligible change in the outcome.

This index is a continuous index of significance. It can be said that if Rope covers 99% of the highest density region (HDI), i.e., Rope covers most of the credible values, then H0 can be accepted. If Rope covers 97.5% of the HDI, then the chance of the null hypothesis being rejected is probably negligible. If Rope covers between 2.5 and 97.5 of HDI, no conclusion can be made about significance., If Rope covers less than 2.5% of the HDI, it is probably safe to reject the null hypothesis, whereas if Rope covers less than 1% of HDI, H0 can be rejected.

The Bayes factor (BF) index is a multipurpose index that can be used to compare different models. BF is a ratio that gives us information about the probability of the observed data under the two compared models (model with effect) versus (model without effect). Its interpretation depends upon whether the BF is the ratio of the posterior probability of the model with the effect to the model without the effect or vice-versa.

BF can be used both in the context of significance and in the context of the existence of an effect. To interpret BF based on Jeffreys (1998) criterion, which is used by default in R packages, it can be said that $3 < \text{BF} \leq 10$ indicates moderate evidence, $10 < \text{BF} \leq 30$ strong evidence, $30 < \text{BF} \leq 100$ very strong evidence, and $\text{BF} > 100$ represents extreme evidence.

8.3.2 Practical Example Using R

An artificial school psychologist wants to know whether the level of calm continuity of introverted (Class = 1) and extroverted (Class = 2) students is different. In a test, he measures the amount of calm persistence (the patience of students when they are

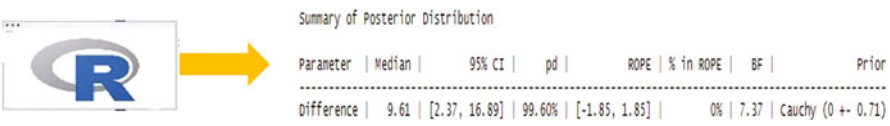


Fig. 8.2 Summary of the Bayesian *t*-test

faced with very difficult questions in an exam). By using a t-test, two independent samples are compared with Bayesian statistics. The relevant codes are shown in Fig. 8.2.

The results of the Bayesian t analysis are given in the table. As can be seen, the median as the central index of the posterior distribution is equal to 12.5. Group membership (extroversion–introversion) shows the median association is positive with a probability of 1 (Pd = 100) (Listing 8.1).

Listing 8.1
R codes of the Bayesian t-test.

```
#t test
library(rstanarm)
library(bayestestR)
library(rstanarm)
library(bayestestR)
library(insight)
library(ggplot2)
library(datawizard)
library(ggplot2)
library(haven)
library(see)
BT <- read_sav("BT.sav")
resultt <- BayesFactor::ttestBF(formula = C~ group, data = BT)
describe_posterior(resultt)
```

Rope supports this with the region corresponding to values associated with a median of zero, showing that it covers a negligibly small amount of possible values. Observation of BF shows moderate evidence in favor of introversion being associated with calm continuity (BF = 7.37).

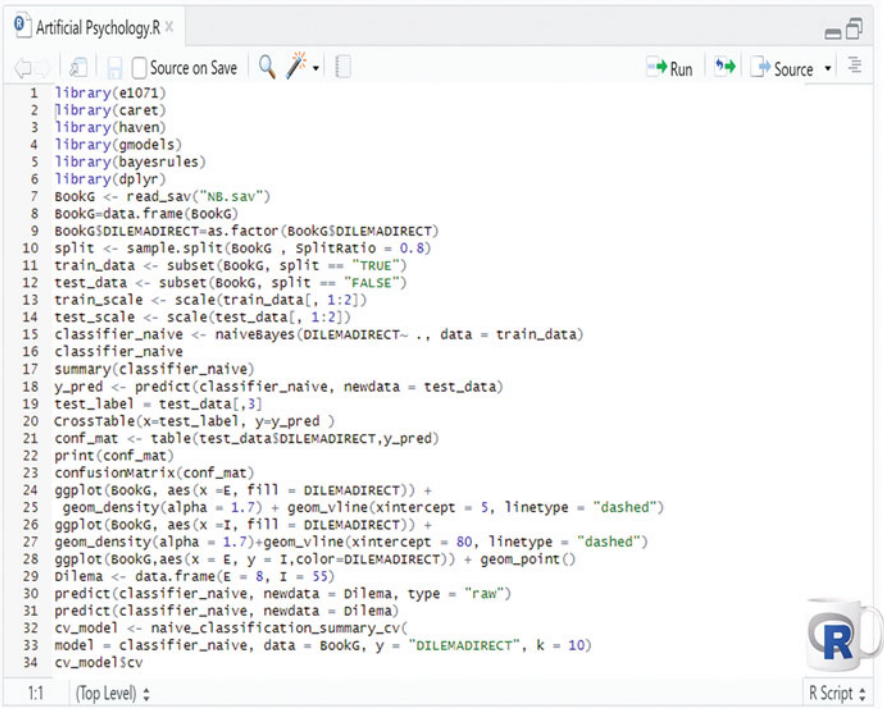
8.4 Naïve Bayes Classifier

Naïve Bayes models are a family of machine learning models, which attempt to classify data into groups allocating to the group with the higher or highest posterior probability of group membership based on a set of characteristics. These models utilize distributions such as the Bernoulli (classification into two groups) and multinomial (classification into more than two groups).

We use a cross-validation approach using 80% of the data to obtain the classification model and the remaining 20% to assess how well it can assign new data into groups.

An artificial cognitive psychologist tries to check the probability of utilitarian-duty-oriented group membership with NB based on emotion regulation (E) and (I) introversion and to know how likely it is that an individual, assuming utilitarianism (1) has high emotion regulation and high introversion. It was already mentioned that the number of predictor variables should be large and independent. Here, for illustrative purposes, the number of predictor variables is 2 and they are assumed to be independent of each other. In this research, he examines 125 people and, based on a cognitive task, separates 94 people as utilitarian and 31 people as task oriented. He also uses the Laplace method for smoothing (R codes are shown in the figure). Also, 80% of the sample is selected as the training sample and 20% as the test sample. The prior for this analysis was equal to 75% (utility oriented) and 25% (task oriented) based on the original data (Listing 8.2).

Listing 8.2
R codes of the naïve Bayesian classifier.



```
Naive Bayes Classifier for Discrete Predictors

Call:
naiveBayes.default(x = X, y = Y, laplace = laplace)

A-priori probabilities:
Y
       1        2
0.746988 0.253012

Conditional probabilities:
   E
Y       [,1]     [,2]
  1 3.870968 3.021691
  2 6.238095 2.791142

   I
Y       [,1]     [,2]
  1 98.62903 17.95918
  2 85.04762 14.00170
```

```
               Accuracy : 0.7143
                 95% CI : (0.5542, 0.8428)
    No Information Rate : 0.8095
    P-Value [Acc > NIR] : 0.9557

                  Kappa : 0.1544

 Mcnemar's Test P-Value : 0.7728

            Sensitivity : 0.7941
            Specificity : 0.3750
         Pos Pred Value : 0.8438
         Neg Pred Value : 0.3000
             Prevalence : 0.8095
         Detection Rate : 0.6429
   Detection Prevalence : 0.7619
      Balanced Accuracy : 0.5846

       'Positive' Class : 1
                    1         2
[1,] 0.3479066 0.6520934
```

Fig. 8.3 Summary of the naïve Bayesian classifier

The results showed that the accuracy of this analysis is 71% and the Kappa coefficient is equal to 0.15, which is a small value. Of course, this is just an example. This analysis has a sensitivity and specificity of 0.79 and 0.38, respectively. Based on the resulting model, which is also validated with ten-fold cross-validation, for a person whose emotional regulation is 8 and whose introversion is 55, the predicted probability of his group membership equals 0.66 for the utilitarian group and 0.34 for the duty-oriented group (Fig. 8.3).

Plots in Fig. 8.4 show the density of variables *E* and *I* in two utilitarian and task-oriented groups based on Posterior distribution.

The last diagram shows the shape of the relationship between *E* and *I*, separated by the duty-oriented and utilitarian groups. The utilitarian group is marked with small red circles and the duty-oriented group with small blue circles. It should be noted that here the aim is only to use NB in classification. Obviously, the sample size is not large enough, and more data is needed (Fig. 8.5).

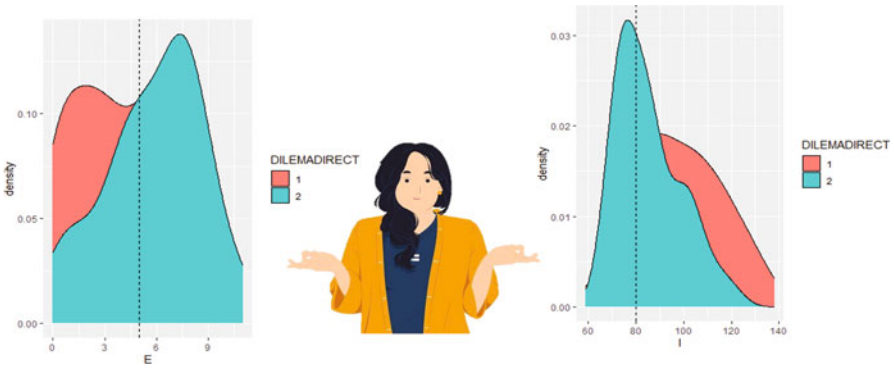


Fig. 8.4 Density of posterior of *E* and *I* variables

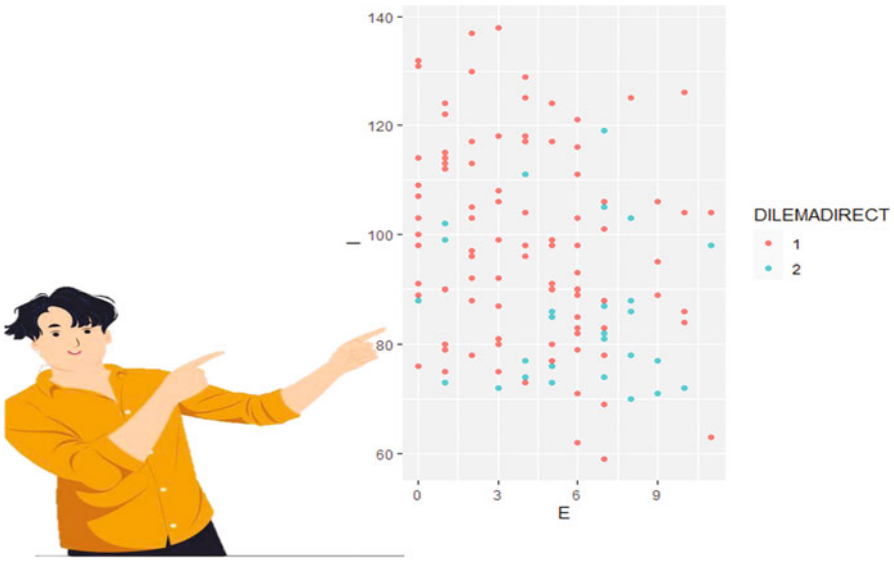


Fig. 8.5 Relationships of *E* and *I*

8.5 Cross-validation

The accuracy of the model is 71.14%, the kappa, comparing predicted group membership to actual group membership, is weak, and McNemar's test is not significant, suggesting a poor rate of classification.

Although the hold-out method is very important in obtaining a valid model, it may lead to over-fitting or under-fitting; therefore, we should use some other methods for this purpose.

Repeated *k*-Fold Cross-validation

The process of splitting the data into k folds can be repeated a number of times. This is when the data is split into *k* groups, with *k*-1 groups used to estimate the machine learning model and the remaining group used to assess its predictive accuracy. This is repeated for each of the *k* groups and called repeated *k*-fold cross-validation. The final model accuracy is the mean of the number of repeats (Listing 8.3).

8.5.1 A Practical Example Using R

Listing 8.3
R codes of the naïve Bayesian classifier (NBC) using CV and LOOCV.

```
library(klaR)
control1 <- trainControl(method="repeatedcv", number=10, repeats=3)
model1 <- train(DILEMADIRECT~ ., data=BookG , trControl=control1, method="nb")
print(model1)
control2 <- trainControl(method="LOOCV")
model2 <- train(DILEMADIRECT~ ., data=BookG, trControl=control2, method="nb")
print(model2)
```

The mean accuracy for this model is 76.84%. The kappa coefficient comparing predicted group with the actual group is low (Fig. 8.6).

Leave-One-Out Cross-validation

Leave-one-out cross-validation, or LOOCV, is the cross-validation technique in which the size of the fold is "1" with "k" being set to the number of observations in the data. This validation is useful when the training data is of limited size, and the number of parameters to be tested is not high (Fig. 8.7).

```
usekernel  Accuracy   Kappa
FALSE      0.7683425  0.3047329
 TRUE      0.8147656  0.4280415
```

Fig. 8.6 The R output of NBC using CV

```
usekernel  Accuracy   Kappa
FALSE      0.7730956  0.3244702
 TRUE      0.8136143  0.4243703
```

Fig. 8.7 The R output of NBC using LOOCV

The mean accuracy for this model is 77.31%. This accuracy in the hold-out validation approach is 71.14% (Listing 8.4).

Listing 8.4
R codes of the ROC curve of NBC.

```
library(ROCR)
nb <- NaiveBayes(DILEMADIRECT~ ., data = train_data)
nbprediction = predict(nb, test_data, type='raw')
#Plot Roc Curve
score = nbprediction$posterior[, 2]
actual.class = test_data$DILEMADIRECT
pred = prediction(score, actual.class)
nb.prff = performance(pred, "tpr", "fpr")
plot(nb.prff,main ="ROC curve -- Naive Bayes Classifier")
```

Fig. 8.8 shows that the curve is above 50% and it means the classification accuracy is more than chance. Good classification is shown in a ROC curve that goes up to the top left hand corner, which is not the case in this example.

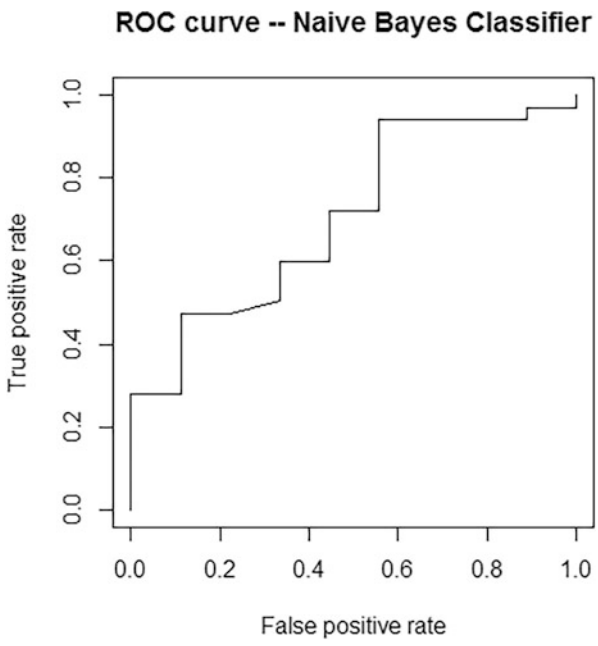


Fig. 8.8 The ROC curve for NBC

8.6 Bayesian Binary Logistic Regression

We can use prior information for model parameters in a logistic regression specified in the example below using the user.prior.density subcommand in the Bayesian logistic regression program, MCMClogit, in R. These prior distributions are then incorporated with information from the raw data to yield posterior distributions for model parameters. The fitting of these models in R is performed using sampling from posterior distributions (Monte Carlo) based upon Markov Chains, which can be viewed as a network of paths where each path relates the parameters (nodes) in the model to one another and to the outcome variable.

Bayesian regression and Bayesian logistic regression are based on Bayes theory, a theory that contains Conditional Probability in its core. Regardless of the mathematical basis and attention to statistical details, the most important distinction between traditional Maximum Likelihood models such as logistic regression and models based on Bayes theory such as Bayesian regression can be summarized as follows:

1. The regression models have slope, intercept, and sigma parameters and each parameter has an associated prior.
2. The estimated parameters have a normal distribution, while in the classic or frequentist-based models, the estimated parameters are fixed and have a probability distribution function based on the same probability distribution function.
3. In logistic regression, each parameter is separate and described by a different distribution.
4. In Bayesian regression analysis, the posterior distribution is made from the prior distribution and likelihood. The mean or other central indicators of a posterior distribution are considered the coefficient of interest of the variable under consideration.
5. In logistic regression analysis, when the posterior distribution is highly skewed or bimodal or multimodal, equal tailed credible sets are used, which are defined as the outer 0.025 quantiles of the posterior distribution. Credible confidence indicates that there is a 95% probability that this Credible confidence interval includes the posterior mean or the true posterior mean of the posterior distribution.
6. The last main difference is the existence of additional or prior information. The posterior distribution that is defined to estimate the parameters in Bayesian analysis can be combined with additional or initial information, which is our initial knowledge about the variable or parameter, which is separate from the data used in the analysis.

The main formula used in Bayesian analysis is as follows:

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)} \tag{8.10}$$

In this formula, $p(x|\theta)$ is the likelihood function and $p(\theta)$ represents the prior distribution. The denominator of the fraction means that $p(x)$ is the probability of x in all x.

As can be seen in the case of subtraction, the probability and initial distributions are multiplied together. Usually, the denominator of the fraction, which is the normalization term, is left out of the calculations in such a way that the posterior distribution or a predictive model includes the product of likelihood and prior. As mentioned, each posterior variable between predictors can have its own posterior distribution. If the artificial psychologist finds that there is no important information outside the obtained data that has an effect on the prior variables, then a uniform prior is usually considered and in this case it is said that the prior in the analysis is noninformative.

It should be noted that if we consider the prior distribution as having a normal distribution with a mean of zero and a very high variance, Bayesian analysis is noninformative, and if all predictor variables in the model are noninformative, then the result is traditional logistic regression based on maximum likelihood which will give the same, or very nearly the same, results as using Bayesian regression. In fact, the purpose of the prior distribution is to reflect the information that does not exist in the existing data from the sample, and therefore, if the prior information is weak, this does not have much effect on the Bayesian analysis.

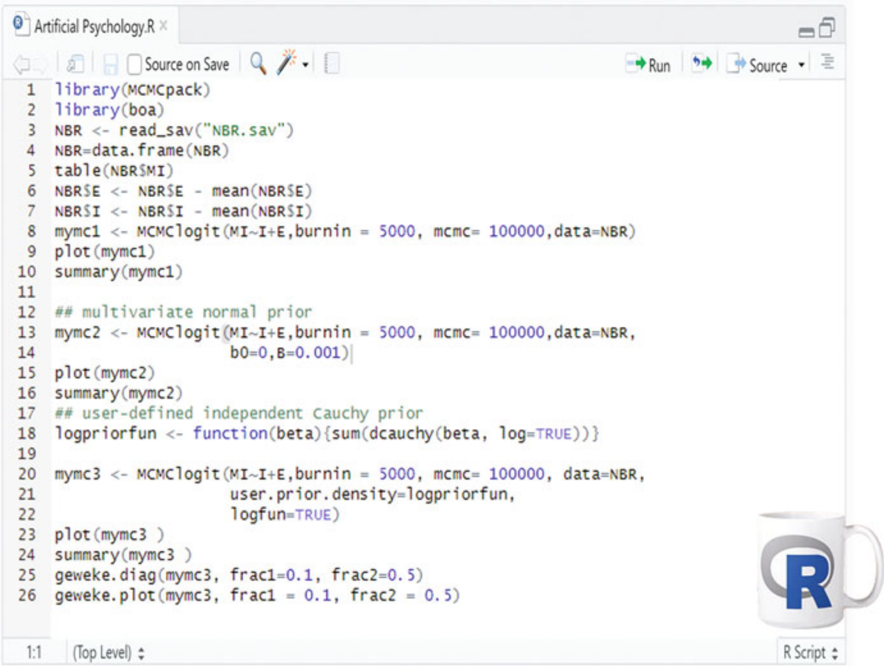
It should be noted that the prior information can not only provide significant quantitative information but also represents a distribution with parameters that is combined with the posterior probability distribution. Markov Chain Monte Carlo (MCMC) methods are used to obtain posterior results. MCMC is a set of algorithms that are used for various purposes such as optimization, dynamic simulation, and sampling. Due to ease of implementation and numerical stability, statisticians prefer the MCMC method, although some consider it a black box of sampling and posterior estimation (Brooks et al., 2011). After the popularization of Bayesian methods in applied problems in the 1990s, the main idea of creating approximate samples from the posterior distribution of interest was expanded by the Markov chain.

MCMC has two traditional approaches, which are called Gibbs sampling (GS) and the Metropolis–Hastings (MH) algorithm. It can be said that the Gibbs sampling algorithm is a special case of MH and can only be used in some conditions such as when the discrete distribution is discrete or normal, while the MH algorithm is used in a wide range of distributions and is based on the possible candidate values of the proposed sample distribution. To achieve a valid inference from the posterior distribution using MCMC, the MCMC chain must converge.

8.6.1 A Practical Example Using R

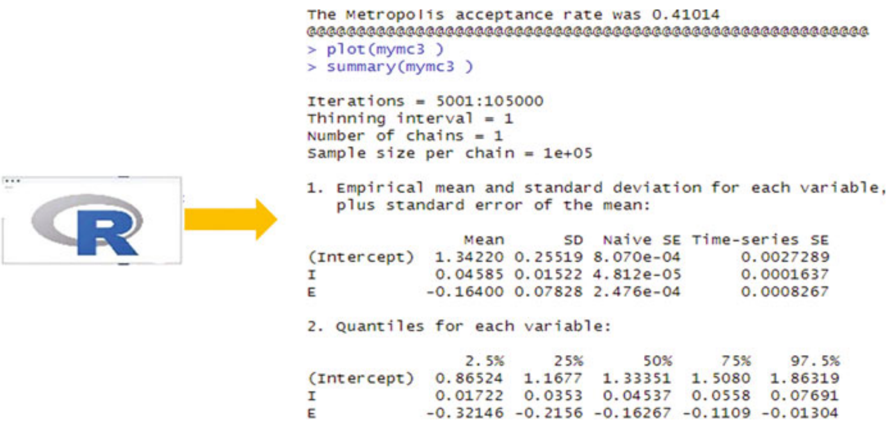
An artificial psychologist tries to predict the probability of marital infidelity based on emotion regulation (E) and cyber introversion (I). He selects a sample of 125 people including two groups (a group with experience of marital infidelity (31 people) and a group of 94 people without experience) and questionnaires on emotion regulation (E) and virtual introversion (I) implemented on them, then he decides to determine the prior distribution by examining the distribution of the variables. He concludes that Cauchy considers the prior distribution of these two variables. This distribution is similar to the normal distribution (Listing 8.5).

Listing 8.5
R codes of Bayesian logistic regression.



```
library(MCMCpack)
library(boa)
NBR <- read_sav("NBR.sav")
NBR=data.frame(NBR)
table(NBR$MI)
NBR$E <- NBR$E - mean(NBR$E)
NBR$I <- NBR$I - mean(NBR$I)
mymc1 <- MCMClogit(MI~I+E,burnin = 5000, mcmc= 100000,data=NBR)
plot(mymc1)
summary(mymc1)

## multivariate normal prior
mymc2 <- MCMClogit(MI~I+E,burnin = 5000, mcmc= 100000,data=NBR,
                   b0=0,B=0.001)
plot(mymc2)
summary(mymc2)
## user-defined independent Cauchy prior
logpriorfun <- function(beta){sum(dcauchy(beta, log=TRUE))}

mymc3 <- MCMClogit(MI~I+E,burnin = 5000, mcmc= 100000, data=NBR,
                   user.prior.density=logpriorfun,
                   logfun=TRUE)
plot(mymc3 )
summary(mymc3 )
geweke.diag(mymc3, frac1=0.1, frac2=0.5)
geweke.plot(mymc3, frac1 = 0.1, frac2 = 0.5)
```

In this example, the result is interpreted with Cauchy's prior distribution. The prior distribution can be considered in a similar way to a normal or t distribution. The result showed that the two variables of emotion regulation and cyber introversion significantly predict group membership and 95% internal credibility is significant for both variables (Fig. 8.9). The R codes are in Listing 8.2.



```
The Metropolis acceptance rate was 0.41014
@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
> plot(mymc3 )
> summary(mymc3 )

Iterations = 5001:105000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 1e+05

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

                Mean      SD  Naive SE Time-series SE
(Intercept)  1.34220 0.25519 8.070e-04      0.0027289
I            0.04585 0.01522 4.812e-05      0.0001637
E           -0.16400 0.07828 2.476e-04      0.0008267

2. Quantiles for each variable:

                2.5%     25%      50%     75%    97.5%
(Intercept)  0.86524  1.1677  1.33351  1.5080  1.86319
I            0.01722  0.0353  0.04537  0.0558  0.07691
E           -0.32146 -0.2156 -0.16267 -0.1109 -0.01304
```

Fig. 8.9 The R output of Bayesian logistic regression

The above figure gives the estimated parameter values from the 2.5% to 97.5% quantiles of their posterior distributions providing a 95% credibility interval for each variable. Parameters of Introversion (I) and Extroversion (E) are 0.04585 and − 0.16400, respectively (Fig. 8.9), which are significant at the 5% significance level in that their 95% credible regions do not contain zero, which would indicate no relationship.

This index is a proportion of the posterior distribution that has the same sign as its median. This index is very similar to a p-value in classical statistics and is known as the maximum probability of effect (MPE). This method is especially useful for models in which there are a large number of variables and categorical input variables have a large number of states and values. This is a simple and interesting method that memorizes how each variable in the training phase is related to the outcome and then makes a prediction by multiplying the effects of each variable.

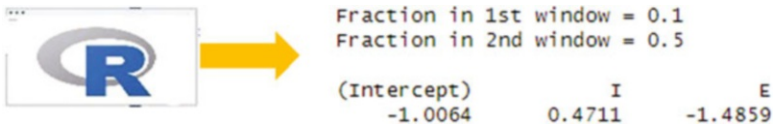
To easily understand this method, let's consider a nonstatistical and intuitive example; suppose we want to predict whether a person is psychologically healthy based on his education level, attachment style, and affective mentalization. In Naive Bayes, this logic is reversed and a question is asked, if a person is psychologically healthy, what is the probability that his emotional metallization is healthy and his attachment is secure?

In short, the question here is that based on the features we have, what is the probability that a person with these features is in a certain group? Naive Bayes answers this question under a very bold naïve assumption, and that assumption is the one that says that all predictor variables are independent of each other. Of course, in the real world, such an assumption is difficult to verify; using NB significantly reduces the complexity of the model. As we know, NB is also based on Bayes law. Bayesian probability is discussed in different parts of this chapter. In short, Bayesian statistics is based on the product of probabilities.

It should be noted that the choice of statistical inference method is not a matter of taste, but a direct result of the problem that the artificial psychologist seeks to answer. If he is worried about the sensitivity of his statistical results to variation in data and modeling procedures, he should use frequentist statistics, but if he is worried about the sensitivity of his statistical findings to possible variation in the unknown quantity that needs to be modeled, he should take steps in the framework of Bayesian statistics.

The artificial psychologist used the dCauchy function. d is the density indicator. The Geweke diagnostic can be used to assess if the parameter estimates have converged, and it compares the values at the start of the repeated sampling from the posterior distributions with those at the end with a z value less than 2 in absolute value indicating no difference. If these values do not differ, the parameter estimates are stable and are deemed to have converged. The output shows that the regression slopes for both the variables of Introversion (I) and Extroversion (E) have $|z| < 2$ (Fig. 8.10). Z values for Geweke are less than 2, indicating that the MCMC posterior distributions of these estimates have converged (Fig. 8.11).

The probability direction (pd) for cyber introversion is 99.92% and for emotion regulation is 98.35% (Fig. 8.12).



```
Fraction in 1st window = 0.1
Fraction in 2nd window = 0.5

(Intercept)           I           E
    -1.0064      0.4711     -1.4859
```

Fig. 8.10 *Z* values of the Geweke diagnostic comparing two fractions of the data

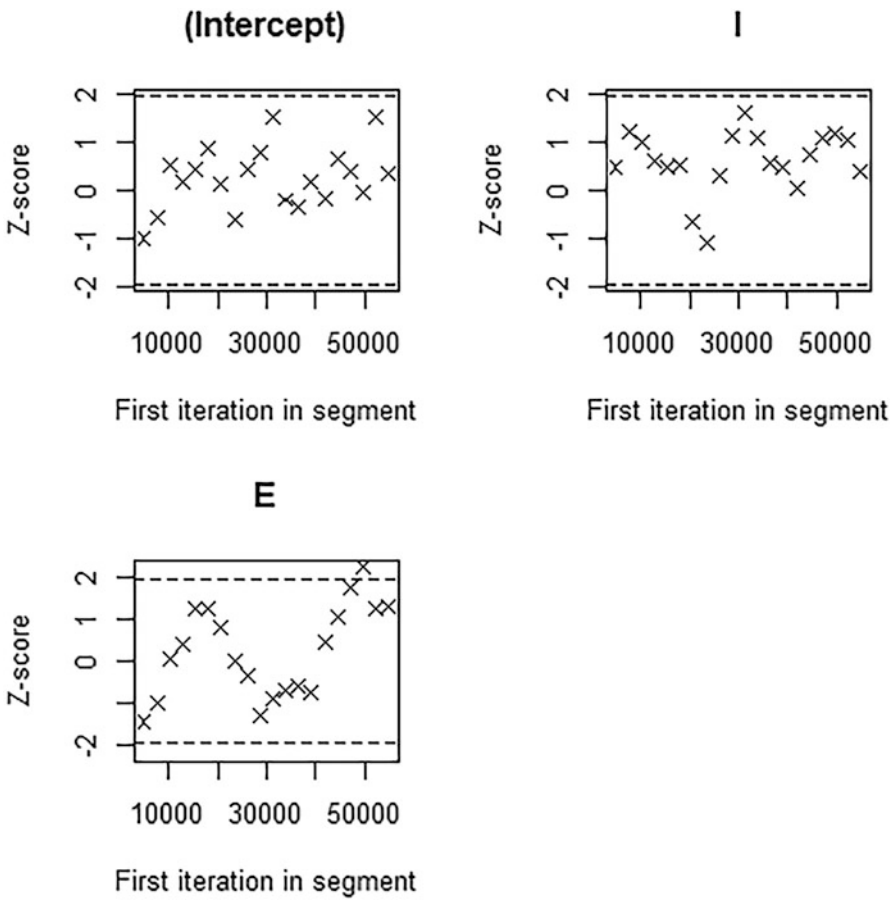
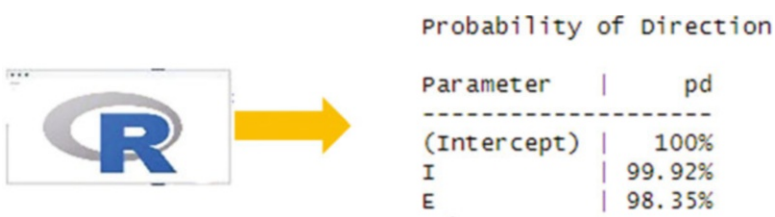


Fig. 8.11 Segment plot of iterations



```
Probability of Direction

Parameter   |     pd
--------------------
(Intercept) |   100%
I           | 99.92%
E           | 98.35%
```

Fig. 8.12 PD of the predictors

8.7 Bayesian Network Analysis

A Bayesian network model is a probabilistic graphical model that represents a set of variables and their conditional dependencies via a graph. The edges in the graph represent the direction and degree of relationship between pairs of variables (nodes). The outputted network is averaged over all possible networks linking the variables specified as being in the network (Figs. 8.13 and 8.14).

The most important concepts of network analysis as an advanced method are presented in the chapter on network analysis in this book. In this chapter, network analysis using Bayesian statistics and its application is presented.

A Bayesian network is a combination of a network structure, specifically a directed acyclic graph (DAG) and a probability distribution associated with it. Therefore, Bayesian networks are graphical methods or models that examine nodes with edges in a graphical structure.

Nodes can be considered psychological variables, symptoms of mental disorders, etc., linked by edges or arrows indicating probabilistic dependencies.

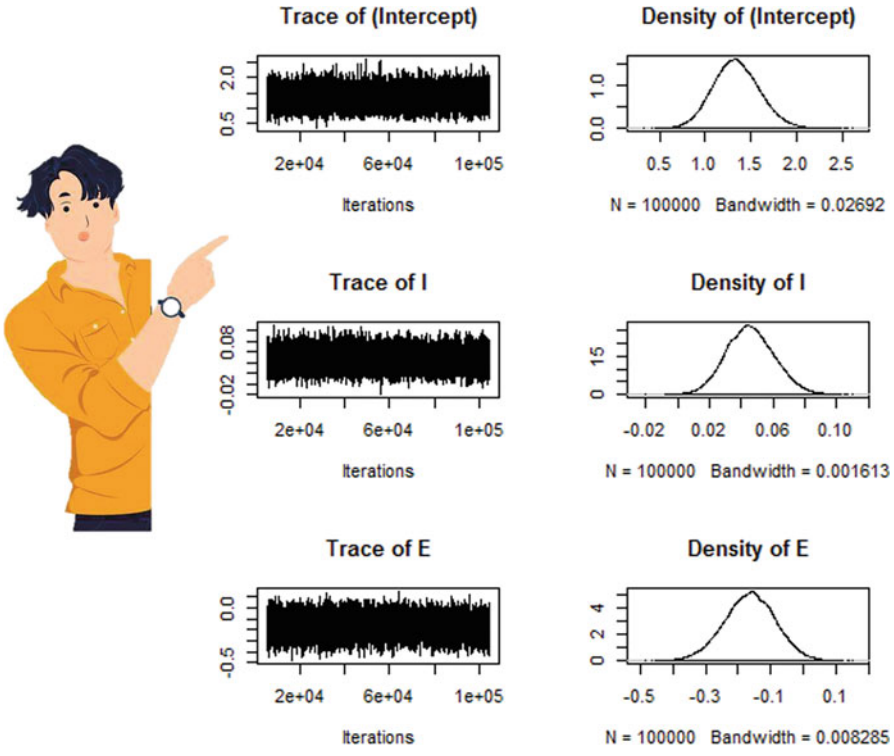


Fig. 8.13 Trace plots of the corresponding posterior estimates of the Intercept, variables via Bayesian network

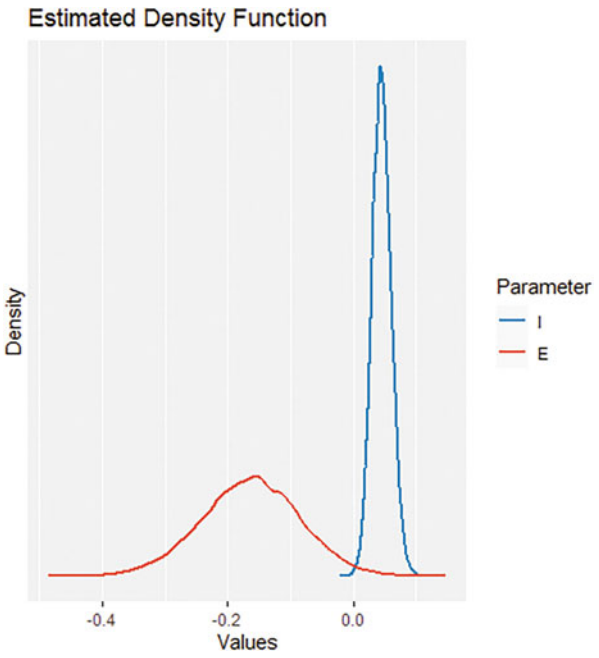


Fig. 8.14 Posterior plot of *I* and *E*

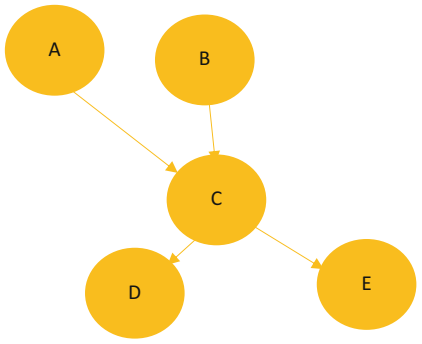


Fig. 8.15 A simple Bayesian network

The graphic structure G = (V, A) is a Bayesian network of a DAG, where V is a node and A is an arc (or edge). DAG can be considered as a factorization of a joint probability distribution (nodes). In simpler terms, the Bayesian network is characterized by three components: (1) nodes (which can be an infinite state), (2) a directed edge (which links two nodes), and (3) a conditional probability for each variable or node.

Nodes and edges define the structure of a network and here the structure of the Bayesian network. The direction of the edges in the form of an arrow (A → B) indicates the causal relationship of two variables. The nodes that are placed immediately before a node are called parent, and the nodes that are placed after a node are called child.

In Fig. 8.15, A and B are parents of C, and D and E are children of C.

In Fig. 8.15, A and B are parents of C, and D and E are children of C.

Prior information and expert information can be used to build a Bayesian network. In Bayesian network analysis, the first step is to determine the network structure, which we call structure learning, and the next step is to determine the parameters of the Bayesian network structure, which we call parameter learning. Methods of structural learning in a Bayesian network may be either constraint-based or score-based. In the following, a more detailed examination of these two categories of algorithms will be carried out. As already mentioned, the main theoretical core of Bayesian networks is based on Bayes' law, which was presented by Tomas Bayes in 1720, and that is why it is called a Bayesian network. The term Bayesian network was coined by Pearl in 1988, and it has been widely used in various fields ever since. In Bayesian networks, if there is a connection between each node and all other nodes in the network, it is called a full Bayesian network, and the main and important feature of Bayesian networks is that each child node matches its parents from the set of nodes and that Non-child and Child nodes are independent.

The goal of structural learning in the Bayesian network is to find the best structure in a way that matches the available data and is optimal in terms of complexity. As mentioned, this learning can be constraint-based or score-based.

Verma and Pearl in 1991 introduced constraint-based algorithms. This algorithm provides a theoretical framework for structural learning of causal models. Three steps are required:

Step 1

In this step, the body of the network, which includes the undirected graph, is learned, generally for ease of implementation of the Markov Blanket, a set of nodes that contain complete information about one another. It is used for each node, including parents, children, and co-nodes that have a common child with that particular node. This structure of a Bayesian network is called the Skeleton of a DAG. Usually, at this stage, to determine the location of edges, conditional independence tests such as X^2 and Fisher tests of independence are used.

Step 2

In the next step, the edges obtained from the previous step are oriented. For this purpose, different algorithms are used such as the Fast Incremental Association (fast.iamb), Incremental Association (iamb), Grow-Shrink (gs), and Interleaved Incremental Association (inter.iam). And finally, the max–min parents and children algorithm (MMPC) is used to identify parents and children in the network.

In this book, the last algorithm, MMPC, is explained. This algorithm is a forward selection method for neighborhood deletion based on the maximization of the minimum association value obtained in each subset of nodes selected in previous iterations. This algorithm teaches the basic structure of the Bayesian network Learn. All rcs are nondirectional, and there is no attempt to direct them.

It should be noted that if the samples are small or there is a lot of missing data, the use of limit-oriented algorithms will cause a lot of errors and these methods cannot

direct some edges. In these cases, the use of score-based algorithms is suggested. In these methods, a measure is used to determine the matching of the networks with the available data searching for a network that matches the data the most. This method is also done in two steps.

In the first step, a search method is specified to make DAG so that all possible structures of DAG are known, and then in the next step, the matching of each structure with the existing data is determined and evaluated with a suitable measure. These two steps continue until there is no possible structure that has a better match.

An important algorithm from the score-based category is the greedy search algorithm, called Hill-climbing (HC), which greedily searches the space of directed graphs (Daly & Shen, 2007).

The PC structure learning algorithm is one of the earliest and the most popular algorithms, introduced by Spirtes et al. (2000). It uses independences observed in data (established by means of classical independence tests) to infer the structure that has generated them.

Another algorithm is Tabu Search, which Glover introduced in 1990. This algorithm is a modified HC algorithm that does not stop after the first DAG but continues until any addition, deletion, and reversal does not improve the score.

The structure of the Bayesian network may change during several executions of the learning algorithm. Therefore, in cross-sectional data, it is necessary to check the stability of the network structure, that is, to obtain a robust set of edges and directions.

For partial correlation networks, bootstrapping methods can be used to evaluate the stability of network estimates (see the network analysis chapter). Such a method can be used to check the stability of structural learning in Bayesian networks. For this reason, the Bayesian network learns the structure from a large set of samples obtained from bootstrapping and based on the criteria (Briganti et al., 2022a, b) that the edges that can be seen in more than 85% of the networks, that their direction appears in more than 50% of networks and that they remain in the network.

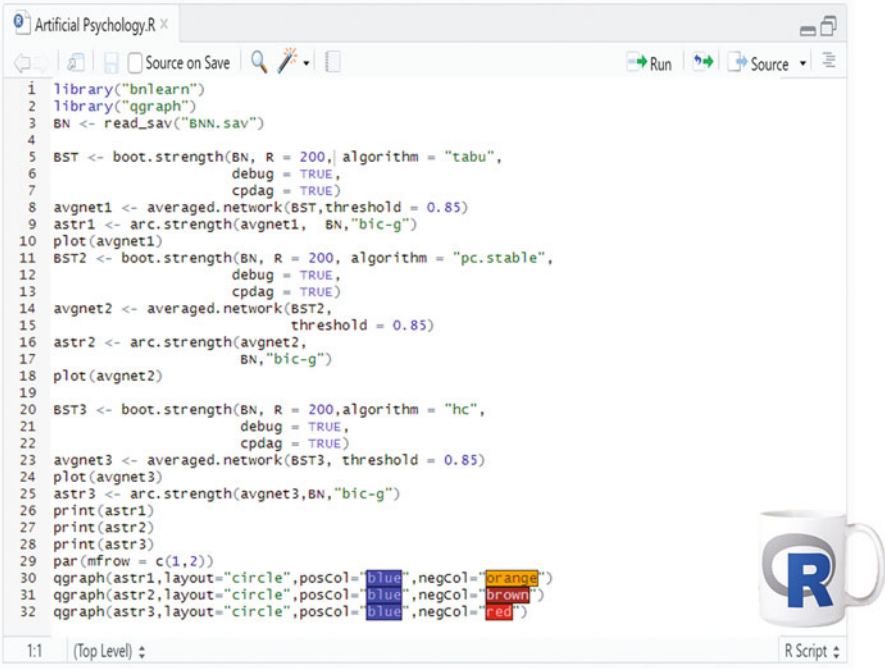
The first criterion is called strength and the second criterion is called minimum direction. It is suggested that the number of samples in the bootstrap method be 100–200 samples. And the proportion of the number of times that the edges are entered in the Bayesian network can also be reported (Briganti et al., 2022a, b). It should be remembered that after learning the structure, it is necessary to determine the intensity of the edges by parameter estimation. The purpose of parameter learning is to estimate network parameters. If the structure of the network is known and all the variables are observable, Maximum likelihood (ML), Maximum a posteriori (MAP), and posterior mean (PM) estimation can be used to estimate the network.

8.7.1 A Practical Example Using R

The researcher tries to model the exploratory Bayesian network based on variables related to marital satisfaction. He measures security (AAIS), avoidance (AAIV), ambivalence (AAIAM), positive affect (PANASP), negative affect (PANASN), and marital satisfaction (GRIMS) with valid and reliable questionnaires. His sample was a sample of 450 married men and women. In this research, he first builds a stable network by using bootstrapping with 200 samples, then analyzes the Bayesian network by using the structural learning PC algorithm, Tabu and HC, and finally estimates the parameters using the average strength of the edges. Listing 33 and the figures indicate the R codes for implementing the Bayesian network and the last step of bootstrapping with network specifications and finally the Bayesian network diagram and intensity of edges based on different methods of network learning.

The R code for implementing the Bayesian network is in Listing 8.6. Table 8.2 shows the variable names and labels for running the Bayesian network.

Listing 8.6
R codes for Bayesian network implementation.



Bootstrap re-sampling can also be used to estimate a level of confidence on the learned edges. In this case, one network can be learned from each bootstrap sample and the resulting PDAGs can be aggregated in a weighted PDAG (WPDAG), where the confidence on each edge is estimated as the fraction of bootstrap samples from

Table 8.2 Variables' names and labels

Variable name	Variable label
AAIS	Secure attachment style
AAIAv	Avoidance attachment style
AAIAm	Ambivalence attachment style
PANASN	Negative affect
PANASP	Positive affect
GRIMS	Marital satisfaction

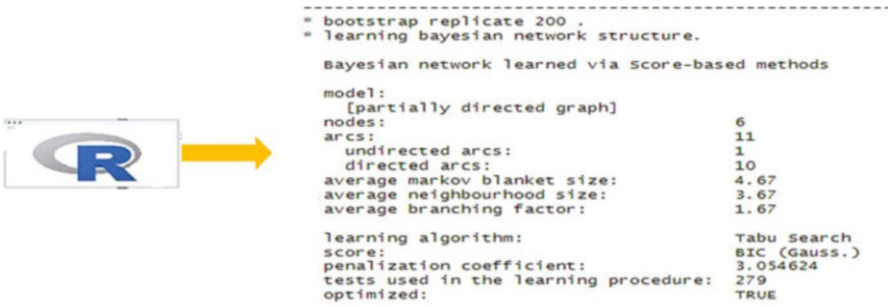


Fig. 8.16 Learning the Bayesian network structure using the score-based method (Tabu Search)

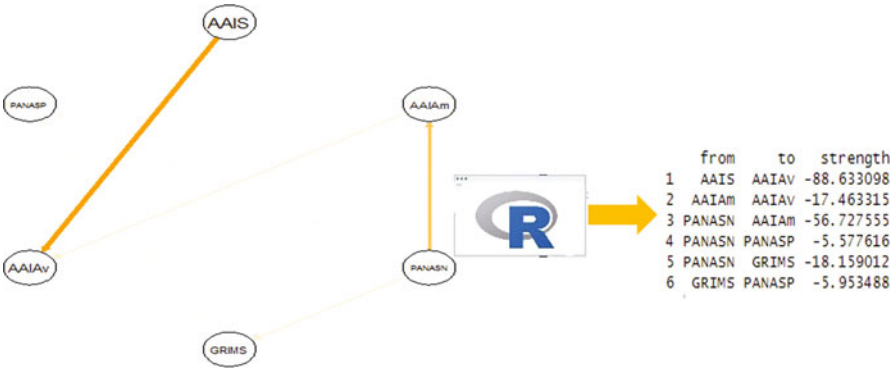


Fig. 8.17 The plot of learning the Bayesian network structure (Tabu search)

which the edge can be learned. Bnstruct can also infer the estimated probability distribution of some variables, given evidence on the values of other observed variables. In this case, a junction tree is used. The Expectation–Maximization algorithm is also implemented, which exploits a BN structure to iteratively estimate conditional probabilities from a dataset with missing values in order to impute these missing values.

As Fig. 8.16 shows, the edge between AAIS and AAIAV has the strongest negative connection and GRIMS and PANAS has the weakest negative edge based on the obtained Bayesian network using Tabu search (Fig. 8.17).

```
bootstrap replicate 200 .
learning bayesian network structure.

Bayesian network learned via Constraint-based methods

model:
 [AAIS][AAIAm][PANASP][AAIAv|AAIS:AAIAm][PANASN|AAIAm:PANASP][GRIMS|AAIS:PANASP:PANASN]
nodes:                                6
arcs:                                 7
  undirected arcs:                    0
  directed arcs:                      7
average markov blanket size:          3.67
average neighbourhood size:           2.33
average branching factor:             1.17

learning algorithm:                   PC (Stable)
conditional independence test:        Pearson's Correlation
alpha threshold:                      0.05
tests used in the learning procedure: 89
```

Fig. 8.18 Learning the Bayesian network structure using the constraint-based algorithm (PC)

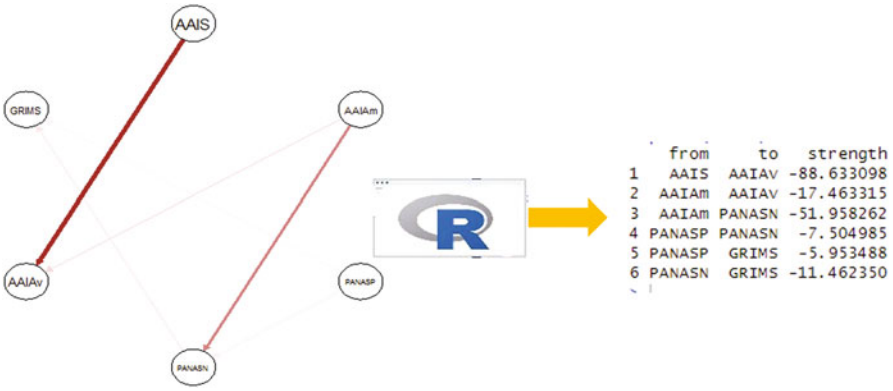


Fig. 8.19 The nodes and edges of the Bayesian network using PC

As Fig. 8.18 indicates, AAIs has the strongest negative effect on AAIAv in the network obtained using PC (Figs. 8.19 and 8.20).

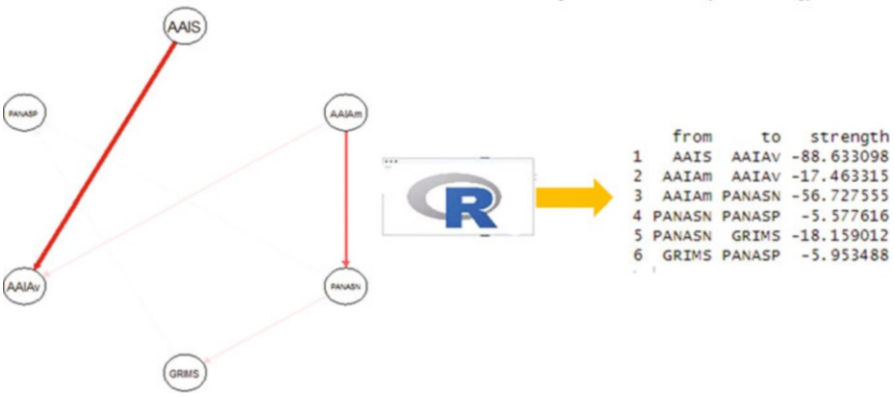
As Fig. 8.21 indicates, AAIs have the strongest negative effect on AAIAv in the network obtained using Hill-climbing.

8.8 Bayesian Model Averaging

Bayesian model averaging (BMA) works out the posterior probabilities associated with various competing linear regression models comprising different subsets of features and aggregates over them with the models containing features having the highest posterior probability being most influential in giving the weighted average of parameter estimates representing associations of predictors with life satisfaction.

```
* bootstrap replicate 200 .
* learning bayesian network structure.

Bayesian network learned via score-based methods

model:
  [partially directed graph]
nodes:                                  6
arcs:                                   8
  undirected arcs:                      3
  directed arcs:                        5
average markov blanket size:            3.67
average neighbourhood size:             2.67
average branching factor:               0.83

learning algorithm:                     Hill-Climbing
score:                                  BIC (Gauss.)
penalization coefficient:               3.054624
tests used in the learning procedure:   55
optimized:                              TRUE
```

Fig. 8.20 Learning the Bayesian network structure using the scored-based Hill-climbing



	from	to	strength
1	AAIS	AAIAv	-88.633098
2	AAIAm	AAIAv	-17.463315
3	AAIAm	PANASN	-56.727555
4	PANASN	PANASP	-5.577616
5	PANASN	GRIMS	-18.159012
6	GRIMS	PANASP	-5.953488

Fig. 8.21 The nodes and edges of the Bayesian network using the scored-based Hill-climbing)

One of the most important points in data analysis and modeling is to find a way to choose a better model. Standard statistical analyses ignore model uncertainty. In data analysis, a model is usually selected from among a class of models, which leaves this approach of uncertainty in model selection. For this reason, this approach leads to overconfident inferences. BMA provides a mechanism to consider model uncertainty when deriving parameter estimates. Therefore, in short, this method provides the uncertainty of the model that exists in the problem of variable selection by averaging the most important models according to the approximate estimation of the posterior probability of the model. For this purpose, in this method, competing models are examined through two indicators to select the best model (the most appropriate model). One of these indicators is the Bayesian Information Criterion

(BIC) presented by Schwarz in 1978, and the other is the posterior probability of each model. The most appropriate model is the model that has the lowest BIC and the highest posterior probability. This method is suitable when there are a large number of predictor variables to predict an outcome, for example 20, 30, or 40. In this method, P! = 0 is calculated for each variable in each model. See the output of the R software in the example. P! = 0 indicates how likely it is that the regression coefficient for any particular predictor variable is non-zero among the resulting models. Also, in the "EV" column, the average of the posterior distribution for each coefficient is presented, and SD indicates the standard deviation of the posterior distribution for each coefficient. To be precise, only the best five models are presented. In the output of each BIC model, the posterior probability, the number of variables in each model, P! = 0 and R^2 are presented.

BMA was used to calculate each variable's relative importance. As can be seen from the above Beck Anxiety, Emotional Processing and Emotional Expression have posterior densities which are centered away from zero and feature in the regression model with the highest posterior probability of occurrence (0.60) and lowest Bayesian Inference Criterion (BIC) with lowest being best. The four remaining models share almost the same posterior probabilities and BIC, which indicates that these competitive models are only slightly different from each other. The EV column shows the average of model coefficients. It is worth mentioning that if only one model was to be selected, this uncertainty in the model could not be justified. The average of all models' coefficients, which is calculated by the sum of posterior probabilities' weights and mentioned in the p! = 0 columns, can consider this uncertainty. Beck Anxiety, Emotional Processing, and Emotional Expression all have p! = 0 equal to 100, showing a very high likelihood of association with life satisfaction.

8.8.1 Practical Example Using R

An artificial psychologist is trying to determine which of the 10 variables (Extraversion, Conscientiousness, Agreeableness, Openness, Emotional Stability, Anxiety, Emotion recognition, Emotion processing, Emotional tracing, and Emotion expression) predicts Life Satisfaction. He uses a sample of 447 male and female students, and all of them respond to these variables through the respective scales; then, BMA is fitted using the R code in Listing 8.7.

Listing 8.7
R codes for BMA implementation.

```
Call:
bicreg(x = x, y = y, strict = FALSE, OR = 20)

  7  models were selected
 Best  5  models (cumulative posterior probability =  0.9351 ):

                p!=0    EV         SD       model 1    model 2    model 3    model 4    model 5
Intercept       100.0  -21.779540  8.71250  -20.8294   -24.9205   -22.0153   -22.2798   -22.7192
Extraversion      0.0    0.000000  0.00000       .          .          .          .          .
Agreebleness      4.3    0.005295  0.03726       .          .          .        0.1220       .
Conscientiousness 19.4   0.446017  1.04828       .        2.3049       .          .          .
EmotionalStability 3.5  -0.018866  0.18936       .          .          .          .          .
Openness          6.2    0.077725  0.39247       .          .        1.2511       .          .
BeckAnxiety     100.0   -5.112318  1.20655   -5.1864    -4.8123    -5.1653    -5.0095    -5.2578
Emotionrecognition 3.0  -0.020347  0.35180       .          .          .          .          .
Emotionprocessing 100.0  10.502802  1.31917   10.6450     9.9807    10.2590    10.5579    10.3361
Emotiontracing    3.6    0.029759  0.28331       .          .          .          .        0.8369
Emotionexpression 100.0  4.164647  0.96144    4.1687     4.1655     4.1092     4.1403     4.2062

nVar                                            3          4          4          4          4
r2                                           0.293      0.300      0.296      0.295      0.294
BIC                                       -136.9818  -134.7171  -132.4447  -131.7278  -131.3287
post prob                                    0.600      0.194      0.062      0.043      0.036
```

Considering the lower BIC and higher posterior probability, it can be said that the first model is the appropriate model in which three variables have been selected, which explains 29.3% of the variance in life satisfaction (Fig. 8.22).

Figure 8.22 indicates that anxiety, emotional processing, and emotional expression are able to predict life satisfaction in model 1, which is the most appropriate model. Looking at p! = 0, the percentage of models where the coefficients for all three variables are non-zero is 100% in this example. The EV column shows the mean of the posterior distribution of the coefficient of each variable, and the SD shows the standard deviation of the mean of each EV coefficient.

Examining the graphs shows that among the density graphs, each variable with the least amount of influence has a spike at the zero point; for example, it can be seen in the graphs that Extraversion, Agreeableness, Emotional Stability, Openness, Emotion recognition, and Emotional tracing have a spike at the zero point and have the least effect. These variables have coefficients that are most likely to be zero, and the variables of anxiety, emotional processing, and emotion expression and of course the constant value of the regression equation play an important role in the resulting model (Fig. 8.23).

As we have seen, there are differences between the strength of different paths depending on what algorithm was used to learn the Bayesian network. But in all three Bayesian network algorithms, the path in the Bayesian network between secure attachment style and very strong avoidance is negative, but the path between ambivalence and negative affect is in opposite directions using the Tabu search

```
Call:
bicreg(x = x, y = y, strict = FALSE, OR = 20)

  7  models were selected
 Best  5  models (cumulative posterior probability =  0.9351 ):
```

	p!=0	EV	SD	model 1	model 2	model 3	model 4	model 5
Intercept	100.0	-21.779540	8.71250	-20.8294	-24.9205	-22.0153	-22.2798	-22.7192
Extraversion	0.0	0.000000	0.00000	.	.	.	.	.
Agreebleness	4.3	0.005295	0.03726	.	.	.	0.1220	.
Conscientiousness	19.4	0.446017	1.04828	.	2.3049	.	.	.
EmotionalStability	3.5	-0.018866	0.18936	.	.	.	.	.
Openness	6.2	0.077725	0.39247	.	.	1.2511	.	.
BeckAnxiety	100.0	-5.112318	1.20655	-5.1864	-4.8123	-5.1653	-5.0095	-5.2578
Emotionrecognition	3.0	-0.020347	0.35180	.	.	.	.	.
Emotionprocessing	100.0	10.502802	1.31917	10.6450	9.9807	10.2590	10.5579	10.3361
Emotiontracing	3.6	0.029759	0.28331	.	.	.	.	0.8369
Emotionexpression	100.0	4.164647	0.96144	4.1687	4.1655	4.1092	4.1403	4.2062
nVar				3	4	4	4	4
r2				0.293	0.300	0.296	0.295	0.294
BIC				-136.9818	-134.7171	-132.4447	-131.7278	-131.3287
post prob				0.600	0.194	0.062	0.043	0.036

Fig. 8.22 R output of BMA of the example

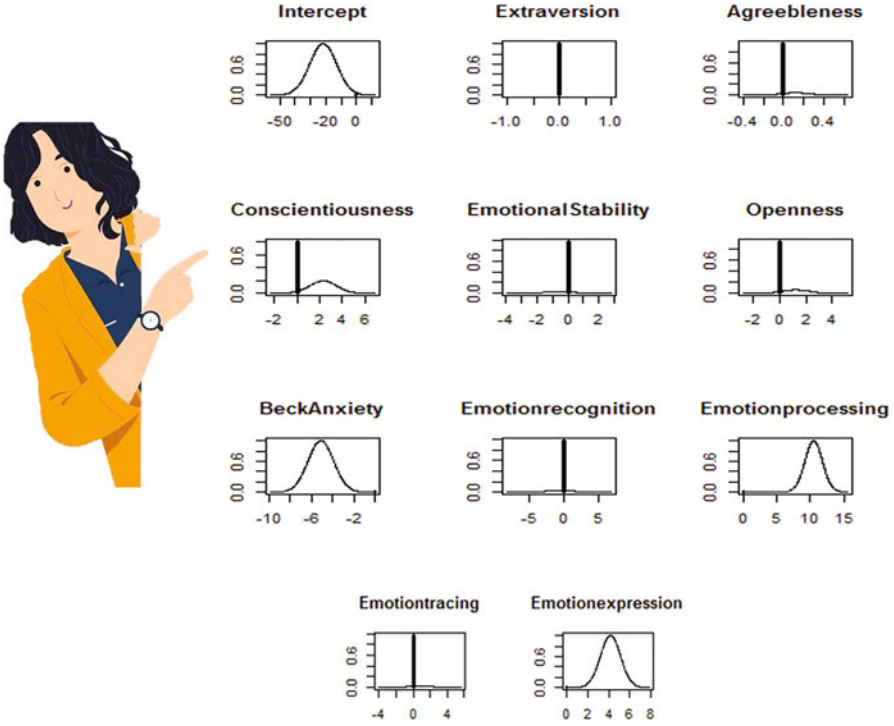


Fig. 8.23 The density of the variables

algorithm and the PC algorithm. In the former, the ambivalent attachment style is affected by negative affect (PANASN→AAIAm), while in the latter, this path is the opposite (AAIAm→PANASN). What makes these models acceptable are the theoretical foundations and the possibility of a theoretical explanation of the resulting findings, and we should not forget that statistical findings support researchers' theories.

Farahani et al. in 2022 used the BMA method to predict Social-Emotional Competence Based on Childhood Trauma, Internalized Shame, Disability/Shame Scheme, Cognitive Flexibility, Distress Tolerance, and Alexithymia.

References

Abraham, A., Milham, M. P., Di Martino, A., Craddock, R. C., Samaras, D., Thirion, B., & Varoquaux, G. (2017). Deriving reproducible biomarkers from multi-site resting-state data: An Autism-based example. *NeuroImage, 147*, 736–745.

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access, 6*, 52138–52160.

Agatonovic-Kustrin, S., & Beresford, R. (2000). Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *Journal of Pharmaceutical and Biomedical Analysis, 22*(5), 717–727.

Agrawal, P., Abutarboush, H. F., Ganesh, T., & Mohamed, A. W. (2021). Metaheuristic algorithms on feature selection: A survey of one decade of research (2009–2019). *IEEE Access, 9*, 26766–26791.

Aguinis, H., & Solarin, A. (2019). Transparency and replicability in qualitative research: The case of interviews with elite informants. *Strategic Management, 40*(8), 1291–1315. https://doi.org/10.1002/smj.3015

American Institute of Aeronautics and Astronautics. (1998). Guide for the verification and validation of computational fluid dynamics simulations, AIAA-G-077-1998.

Anđelković, B. Research and development of new methods for the calculation of clamping circuits using neural networks and fuzzy logic. https://view.officeapps.live.com/op/view.aspx?src=http%3A%2F%2Fwww.hm.co.rs%2Fstatistika%2FPrincipi_fazi_i_neuro_modeliranja1.doc&wdOrigin=BROWSELINK. Last access 10.01.2022.

Andrews, M., & Baguley, T. (2013). Prior approval: The growth of Bayesian methods in psychology. *British Journal of Mathematical and Statistical Psychology, 66*(1), 1–7.

Ansari, N. (2021). A survey on feature selection techniques using evolutionary algorithms. *Iraqi Journal of Science*, 2796–2812.

Apostolopoulos, I. D., Groumpos, P. P., & Apostolopoulos, D. J. (2021). Advanced fuzzy cognitive maps: State-space and rule-based methodology for coronary artery disease detection. *Biomedical Physics & Engineering Express, 7*(4), 045007.

Arfi, B. (2010). *Linguistic fuzzy logic methods in social sciences* (Vol. 253). Springer.

Ashish, K., Dasari, A., Chattopadhyay, S., & Hui, N. B. (2018). Genetic-neuro-fuzzy system for grading depression. *Applied Computing and Informatics, 14*(1), 98–105.

Aslam, N., Khan, I. U., Mirza, S., AlOwayed, A., Anis, F. M., Aljuaid, R. M., & Baageel, R. (2022). Interpretable machine learning models for malicious domains detection using explainable artificial intelligence (XAI). *Sustainability, 14*, 7375. https://doi.org/10.3390/su14127375

H. Farahani et al., *An Introduction to Artificial Psychology*,
https://doi.org/10.1007/978-3-031-31172-7

Avci, E. (2008). Comparison of wavelet families for texture classification by using wavelet packet entropy adaptive network based fuzzy inference system. *Applied Soft Computing, 8*(1), 225–231.

Avci, E., & Akpolat, Z. H. (2006). Speech recognition using a wavelet packet adaptive network based fuzzy inference system. *Expert Systems with Applications, 31*(3), 495–503.

Avci, E., Turkoglu, I., & Poyraz, M. (2005, June). Intelligent target recognition based on wavelet adaptive network based fuzzy inference system. In *Iberian conference on pattern recognition and image analysis* (pp. 594–603). Springer.

Avci, E., Hanbay, D., & Varol, A. (2007). An expert discrete wavelet adaptive network based fuzzy inference system for digital modulation recognition. *Expert Systems with Applications, 33*(3), 582–589.

Ayre, C., & Scally, A. J. (2013). Critical values for Lawshe's content validity ratio: Revisiting the original methods of calculation. *Measurement and Evaluation in Counseling and Development 2014, 47*(1), 79–78. https://doi.org/10.1177/0748175613513808

Babyak, M. A. (2004). What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine, 66*(3), 411–421.

Baig, F., Khan, M. S., Noor, Y., Imran, M., & Baig, F. (2011). Design model of fuzzy logic medical diagnosis control system. *International Journal on Computer Science and Engineering, 3*(5), 2093–2108.

Bailey, J. D., Baker, J. C., Rzeszutek, M. J., & Lanovaz, M. J. (2021). Machine learning for supplementing behavioral assessment. *Perspectives on Behavior Science*, 1–15.

Barabási, A. L., Gulbahce, N., & Loscalzo, J. (2011). Network medicine: A network-based approach to human disease. *Nature Reviews Genetics, 12*(1), 56–68.

Basarkod, G., Sahdra, B., & Ciarrochi, J. (2018). Body image–acceptance and action questionnaire–5: An abbreviation using genetic algorithms. *Behavior Therapy, 49*(3), 388–402.

Bassett, D. S., & Sporns, O. (2017). Network neuroscience. *Nature Neuroscience, 20*(3), 353–364.

Bengio, Y. (2012). Practical recommendations for gradient-based training of deep architectures. *Neural Networks: Tricks of the Trade: Second Edition*, 437–478.

Bentler, P. M., & Satorra, A. (2010). Testing model nesting and equivalence. *Psychological Methods, 15*(2), 111.

Blake, C. (1998). UCI repository of machine learning databases. http://www.ics.uci.edu/~mlearn/MLRepository.html

Blanken, T. F., Bathelt, J., Deserno, M. K., Voge, L., Borsboom, D., & Douw, L. (2021). Connecting brain and behavior in clinical neuroscience: A network approach. *Neuroscience & Biobehavioral Reviews, 130*, 81–90.

Blazer, D. G. (1982). Social support and mortality in an elderly community population. *American Journal of Epidemiology, 115*(5), 684–694.

Bodenhofer, U., De Baets, B., & Fodor, J. (2007). A compendium of fuzzy weak orders: Representations and constructions. *Fuzzy Sets and Systems, 158*(8), 811–829.

Bollen, K. A., & Stine, R. A. (1992). Bootstrapping goodness-of-fit measures in structural equation models. *Sociological Methods & Research, 21*(2), 205–229.

Bonaccorso, G. (2017). *Machine learning algorithms*. Packt Publishing Ltd.

Borgatti, S. P. (2005). Centrality and network flow. *Social Networks, 27*(1), 55–71.

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. *Science, 323*(5916), 892–895.

Borsboom, D., van der Maas, H. L., Dalege, J., Kievit, R. A., & Haig, B. D. (2021a). Theory construction methodology: A practical framework for building theories in psychology. *Perspectives on Psychological Science, 16*, 756–766.

Borsboom, D., Deserno, M. K., Rhemtulla, M., Epskamp, S., Fried, E. I., McNally, R. J., et al. (2021b). Network analysis of multivariate data in psychological science. *Nature Reviews Methods Primers, 1*(1), 58.

Briganti, G., Decety, J., Scutari, M., McNally, R. J., & Linkowski, P. (2022a). Using Bayesian networks to investigate psychological constructs: The case of empathy. *Psychological Reports, 00332941221146711*.

Briganti, G., Scutari, M., & McNally, R. J. (2022b). A tutorial on bayesian networks for psychopathology researchers. *Psychological methods.*

Bringmann, L. F., Albers, C., Bockting, C., Borsboom, D., Ceulemans, E., Cramer, A., et al. (2022). Psychopathological networks: Theory, methods and practice. *Behaviour Research and Therapy, 149*, 104011.

Brooks, S., Gelman, A., Jones, G., & Meng, X. L. (Eds.). (2011). *Handbook of markov chain Monte Carlo*. CRC Press.

Brownlee, J. (2016). Master Machine Learning Algorithms: Discover How They Work and Implement Them From Scratch. (2016). *United States: Machine Learning Mastery*.

Buchanan, B. G. (2005). A (very) brief history of artificial intelligence. *AI Magazine, 26*(4), 53–53.

Bullmore, E., & Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience, 10*(3), 186–198.

Camuñas-Mesa, L. A., Linares-Barranco, B., & Serrano-Gotarredona, T. (2019). Neuromorphic spiking neural networks and their memristor-CMOS hardware implementations. *Materials, 12*(17), 2745.

Cao, S. G., Rees, N. W., & Feng, G. (2001). Mamdani-type fuzzy controllers are universal fuzzy controllers. *Fuzzy Sets and Systems, 123*(3), 359–367.

Chang, M., Balser, J., Roach, J., & Bliss, R. (2019). *Innovative strategies, statistical solutions and simulations for modern clinical trials*. CRC Press/Taylor & Francis Group.

Chen, Y. J., Wu, C. H., Chen, Y. M., Li, H. Y., & Chen, H. K. (2017). Enhancement of fraud detection for narratives in annual reports. *International Journal of Accounting Information Systems, 26*, 32–45.

Chopra, S., Dhiman, G., Sharma, A., Shabaz, M., Shukla, P., & Arora, M. (2021). Taxonomy of adaptive neuro-fuzzy inference system in modern engineering sciences. *Computational Intelligence and Neuroscience, 2021*.

Constantin, M., Schuurman, N. K., & Vermunt, J. (2021). *A general Monte Carlo method for sample size analysis in the context of network models*. https://doi.org/10.31234/osf.io/j5v7u

Corbetta, M., Shulman, G. L., Miezin, F. M., & Petersen, S. E. (1995). Superior parietal cortex activation during spatial attention shifts and visual feature conjunction. *Science, 270*(5237), 802–805.

Costantini, G., & Perugini, M. (2016). The network of conscientiousness. *Journal of Research in Personality, 65*, 68–88.

Crabbe, & Schaar. (2022). Proceedings of the 39th international conference on machine learning, PMLR 162, 2022.

Crowder, J., & Friess, S. (2010). Artificial neural emotions and emotional memory. *Ic-Ai*, 373–378.

Cumming, G. S. (2011). *Spatial resilience in social-ecological systems*. Springer Science & Business Media.

Cumming, G. (2013). The new statistics: Why and how. *Psychological Science, 25*, 7–29.

Cunningham, P., Kathirgamanathan, B., & Delany, S. J. (2021). Feature selection tutorial with python examples. *arXiv preprint arXiv, 2106.06437*.

Cvetković, J., Ivanović Kovačevic, S., Cvetkovic, M., & Cvetkovic, S. (2020). Evaluation of the role of stress in patients with breast cancer and depression by paykel's life event and adaptive neuro-fuzzy approach. *Brain and Behavior, 10*(4), e01570.

Daly, R., & Shen, Q. (2007). Methods to accelerate the learning of bayesian network structures. In *Proceedings of the 2007 UK workshop on computational intelligence*. Imperial College.

Dangeti, P. (2017). *Statistics for machine learning*. Packt Publishing Ltd.

Dash, M., & Liu, H. (1997). Feature selection for classification. *Intelligent Data Analysis, 1*(1-4), 131–156.

De Schryver, T. (2015). A non-technical introduction to text mining. In *Najaarsbijeenkomst SWI, 17 november 2015*. Najaarsbijeenkomst SWI.

Diciccio, T. J., & Romano, J. P. (1988). A review of bootstrap confidence intervals. *Journal of the Royal Statistical Society: Series B (Methodological), 50*(3), 338–354.

Dikopoulou, Z., Papageorgiou, E., Jetter, A., & Bochtis, D. (2018). Open source tool in R language to estimate the inference of the Fuzzy Cognitive Map in environmental decision making, International Congress on Environmental Modelling and Software. 84. https://scholarsarchive.byu.edu/iemssconference/2018/Stream-C/84.

Dosenbach, N. U., Fair, D. A., Miezin, F. M., Cohen, A. L., Wenger, K. K., Dosenbach, R. A., et al. (2007). Distinct brain networks for adaptive and stable task control in humans. *Proceedings of the National Academy of Sciences, 104*(26), 11073–11078.

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv, 1702.08608*.

Du, H., & Zhang, N. (2008). Application of evolving Takagi–Sugeno fuzzy model to nonlinear system identification. *Applied Soft Computing, 8*(1), 676–686.

Dua, D., & Graff, C. (2017) UCI machine learning repository. https://archive.ics.uci.edu/ml/index.php

Dubois, D., & Prade, H. (1996). What are fuzzy rules and how to use them. *Fuzzy Sets and Systems, 84*(2), 169–185.

Dubois, D., & Prade, H. (2005). Fuzzy elements in a fuzzy set. International fuzzy systems association world congress, Jul 2005. pp. 55–60. ffhal-03367408f.

Edwards, S. F. (1965). The statistical mechanics of polymers with excluded volume. *Proceedings of the Physical Society, 85*(4), 613.

Efron, B. (1992). *Bootstrap methods: Another look at the jackknife* (pp. 569–593). Springer.

Eisenbarth, H., Lilienfeld, S. O., & Yarkoni, T. (2015). Using a genetic algorithm to abbreviate the Psychopathic Personality Inventory–Revised (PPI-R). *Psychological Assessment, 27*(1), 194.

Ekong, V. E., Ekong, U. O., Uwadiae, E. E., Abasiubong, F., & Onibere, E. A. (2013). A fuzzy inference system for predicting depression risk levels. *African Journal of Mathematics and Computer Science Research, 6*(10), 197–204.

Elton, A., & Gao, W. (2014). Divergent task-dependent functional connectivity of executive control and salience networks. *Cortex, 51*, 56–66.

Engelbrecht, A. P. (2007). *Computational intelligence: An introduction*. Wiley.

Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods, 23*(4), 617.

Epskamp, S., Kruis, J., & Marsman, M. (2017). Estimating psychopathological networks: Be careful what you wish for. *PLoS One, 12*(6), e0179891.

Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods, 50*, 195–212.

Erin, B., & Abiyev, R. H. (2019, January). Diagnosis of common diseases using Type-2 fuzzy system. In *Proceedings of the 3rd international conference on machine learning and soft computing* (pp. 239–243).

Eronen, M. I., & Bringmann, L. F. (2021). *Perspectives on Psychological Science, 16*(4), 779–788. https://doi.org/10.1177/1745691620970586

Fan, W., & Yan, Z. (2010). Factors affecting response rates of the web survey: A systematic review. *Computers in Human Behavior, 26*(2), 132–139.

Fang, X., Van Kleef, G. A., & Sauter, D. A. (2018a). Person perception from changing emotional expressions: Primacy, recency, or averaging effect? *Cognition and Emotion, 32*(8), 1597–1610.

Fang, X., Sauter, D. A., & Van Kleef, G. A. (2018b). Seeing mixed emotions: The specificity of emotion perception from static and dynamic facial expressions across cultures. *Journal of Cross-Cultural Psychology, 49*(1), 130–148.

Farahani, H. (2020). Comparing linear regression to shrinkage regression algorithms (RR, Lasso, El Net) Using PTSD Patients' data. *Journal of Applied Psychological Research, 11*(3), 193–206.

Farahani, H. A., Kazemi, Z., Aghamohamadi, S., Bakhtiarvand, F., & Ansari, M. (2011). Examining mental health indices in students using Facebook in Iran. *Procedia-Social and Behavioral Sciences, 28*, 811–814.

Farahani, H., LeighPiotr, L., & Piotr, O. (2018). *Fuzzy matrix model as a new method to find optimal diagnostic accuracy points of psychological tests*. European Mathematical Psychology Group.

Farahani, F. V., Karwowski, W., & Lighthall, N. R. (2019). Application of graph theory for identifying connectivity patterns in human brain networks: A systematic review. *Frontiers in Neuroscience, 13*, 585.

Farahani, H., & Azadfallah, P. (2020). P-Value is Hacking Psychological Results: A New Method for Meta-Analysis on "Null Findings" Using Bayesian Statistics,International Conference on Mathematical Psychology and Psychometics, https://doi.org/10.34171/ICMPP.33.151.

Farahani, H., Azadfallah, P., Chesli, R. R., Pourmohamadreza-Tajrishi, M., Esrafilian, F., Lavasani, F. F., & Chiniforoushan, F. (2021a). Methodology of inquiring "therapy failure" in psychotherapy research: Practical guide for clinical practitioners and researchers. *Psychotherapy, 7*, 01.

Farahani, H., Azadfallah, P., Watson, P., & Blagojević, M. (2021b). Bayesian hypothesis testing in linear models: A case study predicting mental health. https://doi.org/10.13140/RG.2.2.32071.37283

Farahani, H., Nápoles, G., & Azadfallah, P. (2021c). Fuzzy cognitive maps for impact assessment in psychological research: Case study of psychological well-being. In *3th international conference on modern approach in humanities and social sciences (ICMHS)*.

Farkas, I., Ábel, D., Palla, G., & Vicsek, T. (2007). Weighted network modules. *New Journal of Physics, 9*, 1–18. https://doi.org/10.1088/1367-2630/9/6/180

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex (New York, NY: 1991), 1*(1), 1–47.

Fielder, E., Von Zglinicki, T., & Jurk, D. (2017). The DNA damage response in neurons: Die by apoptosis or survive in a senescence-like state? *Journal of Alzheimer's Disease, 60*(s1), S107–S131.

Fonseca-Pedrero, E. (2018). Network analysis in psychology. *Papeles del Psicólogo, 39*(1), 1–12.

Fornito, A., Zalesky, A., & Bullmore, E. (2016). *Fundamentals of brain network analysis*. Academic press.

Foygel, R., & Drton, M. (2010). Extended Bayesian information criteria for Gaussian graphical models. *Advances in Neural Information Processing Systems, 23*.

Francis, G. (2014). The frequency of excess success for articles in psychological science. *Psychonomic Bulletin & Review, 21*(5), 1180–1187.

Fraser, A. S. (1957). Simulation of genetic systems by automatic digital computers I. Introduction. *Australian Journal of Biological Sciences, 10*(4), 484–491.

Fried, E. I., & Cramer, A. O. (2017). Moving forward: Challenges and directions for psychopathological network theory and methodology. *Perspectives on Psychological Science, 12*(6), 999–1020.

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics, 9*(3), 432–441.

Gandhi, S. P., Heeger, D. J., & Boynton, G. M. (1999). Spatial attention affects brain activity in human primary visual cortex. *Proceedings of the National Academy of Sciences, 96*(6), 3314–3319.

Gao, W., Gilmore, J. H., Giovanello, K. S., Smith, J. K., Shen, D., Zhu, H., & Lin, W. (2011). Temporal and spatial evolution of brain network topology during the first two years of life. *PLoS One, 6*(9), e25278.

Giabbanelli, P. J., & Crutzen, R. (2014). Creating groups with similar expected behavioural response in randomized controlled trials: A fuzzy cognitive map approach. *BMC Medical Research Methodology, 14*(1), 1–19.

Gianfagna, L., & Di Cecco, A. (2021a). Model-agnostic methods for XAI. In *Explainable AI with python* (pp. 81–113). Springer.

Gianfagna, L., & Di Cecco, A. (2021b). *Explainable AI with python* (pp. 1–202). Springer.

Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., & Kruger, L. (1990). *The empire of chance: How probability changed science and everyday life* (Vol. 12). Cambridge University Press.

Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018, October). Explaining explanations: An overview of interpretability of machine learning. In *2018 IEEE 5th international conference on data science and advanced analytics (DSAA)* (pp. 80–89). IEEE.

Golberg, D. E. (1989). Genetic algorithms in search, optimization, and machine learning. *Addion Wesley, 1989*(102), 36.

Goldberg, D. E., & Richardson, J. (1987, July). Genetic algorithms with sharing for multimodal function optimization. In *Genetic algorithms and their applications: Proceedings of the second international conference on genetic algorithms* (Vol. 4149). Lawrence Erlbaum.

González, A., & Perez, R. (1999). SLAVE: A genetic learning' system based on an iterative approach. *IEEE Transactions on Fuzzy Systems, 7*(2), 176–191.

González, A., Pérez, R., & Verdegay, J. L. (1994). Learning the structure of a fuzzy rule: A genetic approach. *Fuzzy Systems and Artificial Intelligence, 3*(1), 57–70.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Grazioli, S., Rosi, E., Mauri, M., Crippa, A., Tizzoni, F., Tarabelloni, A., et al. (2021). Patterns of response to methylphenidate Administration in Children with ADHD: A personalized medicine approach through clustering analysis. *Children, 8*(11), 1008.

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: A guide to misinterpretations. *European Journal of Epidemiology, 31*, 337–350.

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022). *A primer on partial least squares structural equation modeling (PLS-SEM)* (3rd ed.). Sage.

Hájek, P. (2013). *Metamathematics of fuzzy logic* (Vol. 4). Springer Science & Business Media.

Harper, G. Viability of Artificial Neural Networks in Mobile Healthcare. https://www.theseus.fi/bitstream/handle/10024/28868/Harper_Gavin.pdf?sequence=1. Last access 10.01.2022.

Harrison, A. J., McErlain-Naylor, S. A., Bradshaw, E. J., Dai, B., Nunome, H., Hughes, G. T. G., Kong, P. W., Benedicte Vanwanseele, J., VilasBoas, P., & Fong, D. T. P. (2020). Recommendations for statistical analysis involving null hypothesis significance testing. *Sports Biomechanics, 19*(5), 561–568. https://doi.org/10.1080/14763141.2020.1782555

Hasan, M. F., & Sobhan, M. A. (2020). Describing fuzzy membership function and detecting the outlier by using five number summary of data. *American Journal of Computational Mathematics, 10*, 410–424. https://doi.org/10.4236/ajcm.2020.103022

Haslbeck, J. M., & Waldorp, L. J. (2018). How well do network models predict observations? On the importance of predictability in network models. *Behavior Research Methods, 50*, 853–861.

Haslbeck, J., Ryan, O., Robinaugh, D. J., Waldorp, L. J., & Borsboom, D. (2021). *Modeling psychopathology: From data models to formal theories*. Psychological Methods.

Hayes, A. M., & Andrews, L. A. (2020). A complex systems approach to the study of change in psychotherapy. *BMC Medicine, 18*, 1–13.

Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015a). v. The extent and consequences of p-hacking in science. *PLoS Biology, 13*(3), e1002106.

Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015b). The extent and consequences of p-hacking in science. *PLoS Biology, 13*(3), e1002106.

Heaton, J. (2015). *Artificial intelligence for humans, volume 3: Deep learning and neural networks*. Heaton Research, Inc.

Hevey, D. (2018). Network analysis: A brief overview and tutorial. *Health Psychology and Behavioral Medicine, 6*(1), 301–328. https://doi.org/10.1080/21642850.2018.1521283. PMID: 34040834; PMCID: PMC8114409.

Higashi, N., & Iba, H. (2003, April). Particle swarm optimization with Gaussian mutation. In *Proceedings of the 2003 IEEE swarm intelligence symposium. SIS'03 (Cat. No. 03EX706)* (pp. 72–79). IEEE.

Hilger, K., Ekman, M., Fiebach, C. J., & Basten, U. (2017). Efficient hubs in the intelligent brain: Nodal efficiency of hub regions in the salience network is associated with general intelligence. *Intelligence, 60*, 10–25.

Horowitz, L., & Malle, B. (1993). Fuzzy concepts in psychotherapy research. *Psychotherapy Research, 3*(2), 131–148.

Hu, D., & Zeng, L. L. (2019). *Pattern analysis of the human connectome*. Springer Singapore.

Huette, S., & Spivey, M. (2012). Fuzzy consciousness. In *Being in time: Dynamical theories of phenomenal experience* (pp. 149–164). John Benjamins Pub. Co.

Huette, S., Winter, B., Matlock, T., & Spivey, M. (2012). Processing motion implied in language: Eye-movement differences during aspect comprehension. *Cognitive Processing, 13*(1), 193–197.

Indrayan, A. (2019). The conundrum of P-values: Statistical significance is unavoidable but need medical significance too. *Journal of Biostatistics and Epidemiology, 5*(4), 259–267.

Ishibuchi, H., Nakashima, T., & Murata, T. (1999). Performance evaluation of fuzzy classifier systems for multidimensional pattern classification problems. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 29*(5), 601–618.

Isvoranu, A. M., van Borkulo, C. D., Boyette, L. L., Wigman, J. T., Vinkers, C. H., Borsboom, D., & Group Investigators. (2017). A network approach to psychosis: Pathways between childhood trauma and psychotic symptoms. *Schizophrenia Bulletin, 43*(1), 187–196.

Ito, M., & Gilbert, C. D. (1999). Attention modulates contextual influences in the primary visual cortex of alert monkeys. *Neuron, 22*(3), 593–604.

Izquierdo, S., & Izquierdo, L. R. (2017). Mamdani fuzzy systems for modelling and simulation: A critical assessment. *Available at SSRN, 2900827*.

Jaganathan, K., Rehman, M. U., Tayara, H., & Chong, K. T. (2022). XML-CIMT: Explainable machine learning (XML) model for predicting chemical-induced mitochondrial toxicity. *International Journal of Molecular Sciences, 23*, 15655. https://doi.org/10.3390/ijms232415655

Jang, J. S. (1993). ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics, 23*(3), 665–685.

Jeffreys, H. (1998). *The theory of probability*. OuP Oxford.

Jiao, L., Pan, Q., Denoeux, T., Liang, Y., & Feng, X. (2015). Belief rule-based classification system: Extension of FRBCS in belief functions framework. *Information Sciences, 309*, 26–49.

Jones, P. J., Mair, P., & McNally, R. J. (2018). Visualizing psychological networks: A tutorial in R. *Frontiers in Psychology, 9*, 1742.

Kant, I. (1908). Critique of pure reason. 1781. In *Modern classical philosophers* (pp. 370–456). Houghton Mifflin.

Kello, C. T., & Van Orden, G. C. (2009). Soft-assembly of sensorimotor function. *Nonlinear Dynamics, Psychology, and Life Sciences, 13*(1), 57.

Kello, C. T., Anderson, G. G., Holden, J. G., & Van Orden, G. C. (2008). The pervasiveness of 1/f scaling in speech reflects the metastable basis of cognition. *Cognitive Science, 32*(7), 1217–1231.

Keltner, D., Sauter, D., Tracy, J., & Cowen, A. (2019). Emotional expression: Advances in basic emotion theory. *Journal of Nonverbal Behavior, 43*(2), 133–160.

Khefacha, I., & Belkacem, L. (2015). Modeling entrepreneurial decision-making process using concepts from fuzzy set theory. *Journal of Global Entrepreneurship Research, 5*, 1–21.

Khosla, M., Jamison, K., Ngo, G. H., Kuceyeski, A., & Sabuncu, M. R. (2019). Machine learning in resting-state fMRI analysis. *Magnetic Resonance Imaging, 64*, 101–121.

Klawonna, F., & Novák, V. (1996). The relation between inference and interpolation in the framework of fuzzy systems. *Fuzzy Sets and Systems, 81*(3), 331–354.

Klir, G. J., & Yuan, B. (1995a). *Fuzzy sets and fuzzy logic: Theory and application*. Prentice-Hall.

Klir, G., & Yuan, B. (1995b). *Fuzzy sets and fuzzy logic* (Vol. 4, pp. 1–12). Prentice hall.

Koutsoukas, A., Monaghan, K. J., Li, X., & Huan, J. (2017). Deep-learning: Investigating deep neural networks hyper-parameters and comparison of performance to shallow methods for modeling bioactivity data. *Journal of Cheminformatics, 9*(1), 1–13.

Kruschke, J. (2014). *Doing Bayesian data analysis*. A tutorial with R, JAGS, and Stan.

Kruschke, J. K., Aguinis, H., & Joo, H. (2012). The time has come: Bayesian methods for data analysis in the organizational sciences. *Organizational Research Methods, 15*(4), 722–752.

Kursa, M. B., & Rudnicki, W. R. (2010). Feature selection with the Boruta package. *Journal of Statistical Software, 36*(11), 1–13. https://doi.org/10.18637/jss.v036.i11

Lakens, D. (2015). On the challenges of drawing conclusions from p-values just below 0.05. *PeerJ, 3*, e1142.

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences, 23*(11), 571–579.

Langer, N., Pedroni, A., Gianotti, L. R., Hänggi, J., Knoch, D., & Jäncke, L. (2012). Functional brain network efficiency predicts intelligence. *Human Brain Mapping, 33*(6), 1393–1406.

Larsen, J. T., & McGraw, A. P. (2011). Further evidence for mixed emotions. *Journal of Personality and Social Psychology, 100*(6), 1095.

Lavin, E. A., Giabbanelli, P. J., Stefanik, A. T., & Steven, A. (2018). Should we simulate mental models to assess whether they agree? In Gray & R. Arlinghaus (Eds.), *Proceedings of the annual simulation symposium* (pp. 1–12).

Lawrence, J. (1993). *Introduction to neural networks*. California Scientific Software.

Leggett, D. J. (Ed.). (2013). *Computational methods for the determination of formation constants*. Springer Science & Business Media.

Li, J., Gao, X., & Tian, C. (2006). FCM-based clustering algorithm ensemble for large data sets. In *Fuzzy systems and knowledge discovery: Third international conference, FSKD 2006, Xi'an, China, September 24–28, 2006. Proceedings 3* (pp. 559–567). Springer.

Lipkova, J., Chen, R. J., Chen, B., Lu, M. Y., Barbieri, M., Shao, D., et al. (2022). Artificial intelligence for multimodal data integration in oncology. *Cancer Cell, 40*(10), 1095–1110.

Liu, H., Lafferty, J., & Wasserman, L. (2009). The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *Journal of Machine Learning Research, 10*(10).

Lyu, Z., Peng, K., & Hu, C.-P. (2018). P-value, confidence intervals, and statistical inference: A new dataset of misinterpretation. *Frontiers in Psychology, 9*, 868.

Lyu, X.-K., Xu, Y., Zhao, X.-F., Zuo, X.-N., & Hu, C.-P. (2020). Beyond psychology: Prevalence of p value and confidence interval misinterpretation across different fields. *Journal of Pacific Rim Psychology, 14*. https://doi.org/10.1017/prp.2019.28

Ma, S., & Ji, C. (1999). Performance and efficiency: Recent advances in supervised learning. *Proceedings of the IEEE, 87*(9), 1519–1535.

Mamdani, E. H. (1974). Applications of fuzzy algorithms for control of simple dynamic plant. *Proceedings of the IEEE, 121*, 1585–1588.

Mamdani, E. H., & Assilian, S. (1975). An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies, 7*(1), 1–13.

Mao, Y., Wong, C. S., Tao, X., & Jiang, C. (2018). The impact of affect on organizational justice perceptions: A test of the affect infusion model. *Journal of Management & Organization, 24*(6), 893–916.

Markus, A. F., Kors, J. A., & Rijnbeek, P. R. (2021). The role of explain ability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies. *Journal of Biomedical Informatics, 113*, 103655.

Marques, D. B., Barradas Filho, A. O., Romariz, A. R., Viegas, I. M., Luz, D. A., Barros Filho, A. K., et al. (2014). Recent developments on statistical and neural network tools focusing on biodiesel quality. *International Journal of Computer Science and Application, 3*(3), 97–110.

Marsman, M., & Wagenmakers, E. J. (2017). Bayesian benefits with JASP. *European Journal of Developmental Psychology, 14*(5), 545–555.

Martínez-Jiménez, M. A., Ramirez-GarciaLuna, J. L., Kolosovas-Machuca, E. S., Drager, J., & González, F. J. (2018). Development and validation of an algorithm to predict the treatment modality of burn wounds using thermographic scans: Prospective cohort study. *PLoS One, 13*(11), e0206477.

Massaro, D. W. (1989). Testing between the TRACE model and the fuzzy logical model of speech perception. *Cognitive Psychology, 21*(3), 398–421.

Massaro, D. W., & Cohen, M. M. (2000). Tests of auditory–visual integration efficiency within the framework of the fuzzy logical model of perception. *The Journal of the Acoustical Society of America, 108*(2), 784–789.
Maxwell, J. A. (2004). Causal explanation, qualitative research, and scientific inquiry in education. *Educational Researcher, 33*(2), 3–11.
McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics, 5*, 115–133.
McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC.
Medaglia, J. D., Lynall, M. E., & Bassett, D. S. (2015). Cognitive network neuroscience. *Journal of Cognitive Neuroscience, 27*(8), 1471–1491.
Mehran, K. (2008). Takagi-sugeno fuzzy modeling for process control. *Industrial Automation, Robotics and Artificial Intelligence (EEE8005), 262*, 1–31.
Milosavljević, M. (2019). *Artificial Intelligence*.
Minsky, M., & Papert, S. (1969). *Perceptrons: An essay in computational geometry*. MIT Press.
Mischel, W. (2008). The toothbrush problem. *APS Observer, 21*.
Mkhitaryan, S., Giabbanelli, P., Wozniak, M. K., Nápoles, G., De Vries, N., & Crutzen, R. (2022). FCMpy: A python module for constructing and analyzing fuzzy cognitive maps. *PeerJ Computer Science, 8*, e1078. https://doi.org/10.7717/peerj-cs.1078
Molenaar, P. C. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement, 2*(4), 201–218.
Morey, R. D., Hoekstra, R., Rouder, J. N., & Wagenmakers, E.-J. (2016). Continued misinterpretation of confidence intervals: Response to Miller and Ulrich. *Psychonomic Bulletin & Review, 23*, 131–140.
Mosoiu, C., Sumedrea, A. G., Burtea, V., & Ifteni, P. (2010, June). Fuzzy system approach to symptoms in schizophrenia. In *Proceedings of the 11th WSEAS international conference on neural networks and 11th WSEAS international conference on evolutionary computing and 11th WSEAS international conference on fuzzy systems* (pp. 268–272).
Motter, B. C. (1993). Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *Journal of Neurophysiology, 70*(3), 909–919.
Mushtaq, F., Bland, A. R., & Schaefer, A. (2011). Uncertainty and cognitive control. *Frontiers in Psychology, 2*, 249.
Nguyen, T. H., & Walker, E. (1977). *A first course in fuzzy logic*. CRC Press.
Noetel, M., Ciarrochi, J., Sahdra, B., & Lonsdale, C. (2019). Using genetic algorithms to abbreviate the mindfulness inventory for sport: A substantive-methodological synthesis. *Psychology of Sport and Exercise, 45*, 101545.
Novák, V. (1994). Fuzzy control from the logical point of view. *Fuzzy Sets and Systems, 66*(2), 159–173.
Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological Review, 85*(3), 172.
Oliveira, C. R, & Werlan, T. (2007). *Ergodic hypothesis in classical statistical mechanics*.
Pal, S. K., & Mandal, D. P. (1991). Fuzzy logic and approximate reasoning: An overview. *IETE Journal of Research, 37*(5-6), 548–560.
Pamučar, D. (2010). *Application of fuzzy logic and artificial neural networks in the decision-making process of the traffic support authority*.
Papageorgiou, E. (2011). Learning algorithms for fuzzy cognitive maps—A review study. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 42*(2), 150–163.
Papageorgiou, E. I. (2013). *Fuzzy Cognitive Maps for Applied Sciences and Engineering: From Fundamentals to Extensions and Learning Algorithms*, 54, Springer Science & Business Media.
Papageorgiou, E. I., Stylios, C. D., & Groumpos, P. P. (2004). Active hebbian learning algorithm to train fuzzy cognitive maps. *International Journal of Approximate Reasoning, 37*(3), 219–249.

Papageorgiou, K., Carvalho, G., Papageorgiou, E. I., Bochtis, D., & Stamoulis, G. (2020). Decision-making process for photovoltaic solar energy sector development using fuzzy cognitive map technique. *Energies, 13*(6), 1427.

Papakostas, G. A., Polydoros, A. S., Koulouriotis, D. E., & Tourassis, V. D. (2011). Training fuzzy cognitive maps by using hebbian learning algorithms: A comparative study. In *2011 IEEE international conference on fuzzy systems (FUZZ-IEEE 2011)* (pp. 851–858). IEEE.

Pearl, J. (2000). Causal inference without counterfactuals: Comment. *Journal of the American Statistical Association, 95*(450), 428–431.

Perfilieva, I., & Vajgl, M. (2015, June). Autoassociative fuzzy implicative memory on the platform of fuzzy preorder. In *2015 conference of the international fuzzy systems association and the European Society for Fuzzy Logic and Technology (IFSA-EUSFLAT-15)* (pp. 1598–1603). Atlantis Press.

Pervaiz, U., Vidaurre, D., Woolrich, M. W., & Smith, S. M. (2020). Optimising network modelling methods for fMRI. *NeuroImage, 211*, 116604.

Petruo, V. A., Mückschel, M., & Beste, C. (2018). On the role of the prefrontal cortex in fatigue effects on cognitive flexibility-a system neurophysiological approach. *Scientific Reports, 8*(1), 1–13.

Poggio, G. F., Gonzalez, F., & Krause, F. (1988). Stereoscopic mechanisms in monkey visual cortex: Binocular correlation and disparity selectivity. *Journal of Neuroscience, 8*(12), 4531–4550.

Pourabdollah, A., Mendel, J., & M & John, R.I. (2020). Alpha-cut representation used for defuzzification in rule-based systems. *Fuzzy Sets and System, 399*(15), 110–132. https://doi.org/10.1016/j.fss.2020.05.008

Pratt, M. G., Lepisto, D., & Dane, E. (2019). The hidden side of trust: Supporting and sustaining leaps of faith among firefighters. *Administrative Science Quarterly, 64*, 398–434. ISI.

Ragin, C. C. (2000). *Fuzzy-set social science*. University of Chicago Press.

Rayan, F., Nanjayan, S. K., Quah, C., Ramoutar, D., Konan, S., & Haddad, F. S. (2015). Review of evolution of tunnel position in anterior cruciate ligament reconstruction. *World Journal of Orthopedics, 6*(2), 252.

Reinertsen, K. V., Engebraaten, O., Loge, J. H., Cvancarova, M., Naume, B., Wist, E., et al. (2017). Fatigue during and after breast cancer therapy—A prospective study. *Journal of Pain and Symptom Management, 53*(3), 551–560.

Riza, L. S., Bergmeir, C., Herrera, F., & Benítez, J. M. (2015). frbs: Fuzzy rule-based systems for classification and regression in R. *Journal of Statistical Software, 65*, 1–30.

Robinaugh, D. J., Hoekstra, R. H., Toner, E. R., & Borsboom, D. (2020). The network approach to psychopathology: A review of the literature 2008–2018 and an agenda for future research. *Psychological Medicine, 50*(3), 353–366.

Rolstad, S., Adler, J., & Rydén, A. (2011). Response burden and questionnaire length: Is shorter better? A review and meta-analysis. *Value in Health, 14*(8), 1101–1108.

Rudin, S. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*, 206–215.

Russell, S. J. (2010). *Artificial intelligence a modern approach*. Pearson Education, Inc.

Ryan, O., Bringmann, L. F., & Schuurman, N. K. (2022). The challenge of generating causal hypotheses using network models. *Structural Equation Modeling: A Multidisciplinary Journal, 29*(6), 953–970.

Sadeghian, F., Hasani, H., & Jafari, M. (2021). Feature selection based on genetic algorithm in the diagnosis of autism disorder by fMRI. *Caspian Journal of Neurological Sciences, 7*(2), 74–83.

Sahdra, B. K., Ciarrochi, J., Parker, P., & Scrucca, L. (2016). Using genetic algorithms in a large nationally representative American sample to abbreviate the multidimensional experiential avoidance questionnaire. *Frontiers in Psychology, 7*, 189.

Sambataro, F., Murty, V. P., Callicott, J. H., Tan, H. Y., Das, S., Weinberger, D. R., & Mattay, V. S. (2010). Age-related alterations in default mode network: Impact on working memory performance. *Neurobiology of Aging, 31*(5), 839–852.

Sandy, C. J., Gosling, S. D., & Koelkebeck, T. (2014). Psychometric comparison of automated versus rational methods of scale abbreviation. *Journal of Individual Differences, 35*, 221.

Schmidt, F. L., & Oh, I. S. (2016). The crisis of confidence in research findings in psychology: Is lack of replication the real problem? Or is it something else? *Archives of Scientific Psychology, 4*(1), 32.

Schopenhauer, A., Frauenstädt, J., & Hübscher, A. (1859). *Die welt als wille und vorstellung* (Vol. 2). Brockhaus.

Scrucca, L. (2013). GA: A Package for Genetic Algorithms in R. *Journal of Statistical Software, 53* (4), 1–37. https://doi.org/10.18637/jss.v053.i04

Seising, R. (2008). On the absence of strict boundaries—Vagueness, haziness, and fuzziness in philosophy, science, and medicine. *Applied Soft Computing, 8*(3), 1232–1242.

Sellke, T., Bayarri, M. J., & Berger, J. O. (2001). Calibration of ρ values for testing precise null hypotheses. *The American Statistician, 55*(1), 62–71.

Sharpe, D. (2013). Why the resistance to statistical innovations? Bridging the communication gap. *Psychological Methods, 18*(4), 572.

Shen, B., Wang, Z., & Liu, X. (2011). A Stochastic sampled-data approach to distributed $ H_ {\infty} $ filtering in sensor networks. *IEEE Transactions on Circuits and Systems I: Regular Papers, 58*(9), 2237–2246.

Shepherd, G. M., & Koch, C. (1990). Introduction to synaptic circuits. In G. M. Shepherd (Ed.), *The synaptic organization of the brain*. Oxfrod University Press.

Siew, C. S., Wulff, D. U., Beckage, N. M., & Kenett, Y. N. (2019). Cognitive network science: A review of research on cognition through the lens of network representations, processes, and dynamics. *Complexity, 2019*.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2016). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*, 1359.

Smithson, M. (1982). Applications of fuzzy set concepts to behavioral sciences. *Mathematical Social Sciences, 2*(3), 257–274.

Smithson, M., & Oden, G. C. (1999). Fuzzy set theory and applications in psychology. In *Practical Applications of fuzzy technologies* (pp. 557–585). Kluwer Academic.

Smithson, M., & Verkuilen, J. (2006). *Fuzzy set theory: Applications in the social sciences* (Vol. 147).

Spence, C., Driver, J., & Driver, J. C. (Eds.). (2004). *Crossmodal space and crossmodal attention*. Oxford University Press.

Spirtes, P., Glymour, C. N., Scheines, R., & Heckerman, D. (2000). *Causation, prediction, and search*. MIT Press.

Spivey, M. J., & Spirn, M. J. (2000). Selective visual attention modulates the direct tilt aftereffect. *Perception & Psychophysics, 62*(8), 1525–1533.

Sporns, O. (2014). Contributions and challenges for network models in cognitive neuroscience. *Nature Neuroscience, 17*(5), 652–660.

Sporns, O., Tononi, G., & Kötter, R. (2005). The human connectome: A structural description of the human brain. *PLoS Computational Biology, 1*(4), e42.

Stanley, M. L., Simpson, S. L., Dagenbach, D., Lyday, R. G., Burdette, J. H., & Laurienti, P. J. (2015). Changes in brain network efficiency and working memory performance in aging. *PLoS One, 10*(4), e0123950.

Stephen, D. G., & Mirman, D. (2010). Interactions dominate the dynamics of visual cognition. *Cognition, 115*(1), 154–165.

Stoklasa, J., Talašová, J., & Holeček, P. (2011). Academic staff performance evaluation–variants of models. *Acta Polytechnica Hungarica, 8*(3), 91–111.

Stoklasa, J., Talášek, T., & Musilová, J. (2014). Fuzzy approach-a new chapter in the methodology of psychology? *Human Affairs, 24*(2), 189–203.

Sugeno, M., & Kang, G. (1988). Structure identification of fuzzy model. *Fuzzy Sets and Systems, 28*(Oct. 1988), 15–33.

Sugeno, M., & Yasukawa, T. (1993). A fuzzy-logic-based approach to qualitative modeling. *IEEE Transactions on Fuzzy Systems, 1*(1), 7–31.

Sullivan, E. D. (2022). *Neurodevelopmental factors contributing to substance use and mental illness* (Doctoral dissertation, Dartmouth College).

Tang, J., Alelyani, S., & Liu, H. (2014). Feature selection for classification: A review. *Data Classification: Algorithms and Applications, 37*.

Taylor, S. (2022). The psychology of pandemics. *Annual Review of Clinical Psychology, 18*, 581–609.

ten Donkelaar, H. J., ten Donkelaar, H. J., Broman, J., & van Domburg, P. (2020). The somatosensory system. In *Clinical neuroanatomy: Brain circuitry and its disorders* (pp. 171–255).

Terziyska, M., Doukovska, L., & Petrov, M. (2015). Implicit GPC based on semi fuzzy neural network model. In *Intelligent Systems' 2014* (pp. 695–706). Springer.

Throckmorton, C. S., Mayew, W. J., Venkatachalam, M., & Collins, L. M. (2015). Financial fraud detection using vocal, linguistic and financial cues. *Decision Support Systems, 74*, 78–87.

Tibshirani, R. (1996). A comparison of some error estimates for neural network models. *Neural Computation, 8*(1), 152–163.

Torres, A., & Nieto, J. J. (2006). Fuzzy logic in medicine and bioinformatics. *Journal of Biomedicine and Biotechnology, 2006*.

Tsai, C. F., & Chen, M. Y. (2010). Variable selection by association rules for customer churn prediction of multimedia on demand. *Expert Systems with Applications, 37*(3), 2006–2015.

Turkoglu, I., & Avci, E. (2008). Comparison of wavelet-SVM and wavelet-adaptive network based fuzzy inference system for texture classification. *Digital Signal Processing, 18*(1), 15–24.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage, 15*(1), 273–289.

Van Borkulo, C. D., Borsboom, D., Epskamp, S., Blanken, T. F., Boschloo, L., Schoevers, R. A., & Waldorp, L. J. (2014). A new method for constructing networks from binary data. *Scientific Reports, 4*(1), 1–10.

Van Den Heuvel, M. P., Stam, C. J., Kahn, R. S., & Pol, H. E. H. (2009). Efficiency of functional brain networks and intellectual performance. *Journal of Neuroscience, 29*(23), 7619–7624.

Van Der Heide, A., Sánchez, D., & Trivino, G. (2011, August). Computational models of affect and fuzzy logic. In *Proceedings of the 7th conference of the European Society for Fuzzy Logic and Technology* (pp. 620–627). Atlantis Press.

Van Leekwijck, W., & Kerre, E. E. (1999). Defuzzification: Criteria and classification. *Fuzzy Sets and Systems, 108*(2), 159–178.

Velez, F., & Kim, B. (2017). *Towards A Rigorous Science of Interpretable Machine Learning*. arXiv: Machine Learning

Vélez, J. I. (2021a). Machine learning psychology: Advocating for a data-driven approach. *International Journal of Psychological Research, 14*(1), 6–11. https://doi.org/10.21500/20112084.5365

Vélez, J. I. (2021b). Machine learning based psychology: Advocating for A data-driven approach. *International Journal of Psychological Research, 14*(1), 6–11.

Vennix, J. A. M. (1996). *Group model building*. Facilitating Team Learning Using System Dynamics.

Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., & Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review, 25*(1), 35–57. https://doi.org/10.3758/s13423-017-1343-3. PMID: 28779455; PMCID: PMC5862936.

Wang, X. (2017). *Optimization for training deep models*.

Wang, L., & Jiang, T. (1994). On the complexity of multiple sequence alignment. *Journal of Computational Biology, 1*(4), 337–348.

Wang, L., Dong, J. Y., & Li, S. L. (2015). Fuzzy inference algorithm based on quantitative association rules. *Procedia Computer Science, 61*, 388–394.

Wanga, H., Yanga, B., & Li, W. (2022). Some properties of fuzzy t-norm and vague t-norm. https://doi.org/10.48550/arXiv.2205.09231

Wierman, M. J. (2010). An introduction to the mathematics of uncertainty. *Creighton University*, 149–150.

Williams, D. R., & Rast, P. (2020). Back to the basics: Rethinking partial correlation network methodology. *British Journal of Mathematical and Statistical Psychology, 73*(2), 187–212.

Wu, K., Taki, Y., Sato, K., Hashizume, H., Sassa, Y., Takeuchi, H., et al. (2013). Topological organization of functional brain networks in healthy children: Differences in relation to age, sex, and intelligence. *PLoS One, 8*(2), e55347.

Xin, Y., Kong, L., Liu, Z., Chen, Y., Li, Y., Zhu, H., Gao, M., Hou, H., & Wang, C. (2018). Machine learning and deep learning methods for cybersecurity. *IEEE Access, 6*, 35365–35381.

Yang, L., & Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing, 415*, 295–316.

Yarkoni, T. (2010). The abbreviation of personality, or how to measure 200 personality scales with 200 items. *Journal of Research in Personality, 44*(2), 180–198.

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science, 12*(6), 1100–1122.

Zeki, S. (2001). Localization and globalization in conscious vision. *Annual Review of Neuroscience, 24*(1), 57–86.

Zeki, S. (2003). The disunity of consciousness. *Trends in Cognitive Sciences, 7*(5), 214–218.

Zétényi, T. (Ed.). (1988). *Fuzzy sets in psychology*. Elsevier.

Zhou, J., & Chen, F. (2018). *Human and machine learning: Visible, explainable, trustworthy and transparent; human–computer interaction series*. Springer; ISBN 978-3-319-90402-3.

Zhou, B., Sun, Y., Bau, D., & Torralba, A. (2018). Interpretable basis decomposition for visual explanation. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 119–134).

Zhou, J., Gandomi, A. H., Chen, F., & Holzinger. (2021). A. Evaluating the quality of machine learning explanations: A survey on methods and metrics. *Electronics, 10*, 593. https://doi.org/10.3390/electronics10050593

Ziliak, S., & McCloskey, D. N. (2008). *The cult of statistical significance: How the standard error costs us jobs, justice, and lives*. University of Michigan Press.

Zimmermann, H. J. (2011). *Fuzzy set theory—And its applications*. Springer Science & Business Media.

Index

H. Farahani et al., *An Introduction to Artificial Psychology*,
https://doi.org/10.1007/978-3-031-31172-7