Dean&Francis
ISSN 2959-6157

Comprehensive Analysis of Data Preprocessing and Models in Data Mining

Ning Zhou

Shenzhen internatuon fundation college,Shenzhen,518000,China

Abstract:

Nowadays, the data of the Internet shows explosive growth, in which the massive data contains great economic value. How to deal with and mine data has become an important topic nowadays. In statistics, data mining is often used to build corresponding mathematical models to analyze and process data, so as to get the results. Model to analyze and process the data, so as to get the results people want. This paper summarizes the data processing in the field of data analysis and processing technology. methods in the field of data analysis and processing technology, and first introduces the commonly used data preprocessing methods. On this basis, this paper describes in detail the commonly used data optimization models, predictive In details, this paper describes the commonly used data optimization models, prediction, decision-making models, such as: Linear programming model ,Dynamic programming model. In details, its process in data analysis and processing is detailed from the basic principles, processing steps and practical applications. Finally, we summarize the data preprocessing techniques and model building mentioned throughout the paper. This paper alsolooks ahead to the current challenges that society faces in the data domain, and speculates on future research directions with high relevance.

Keywords: DataPreprocessing;Optimization Models;Decision-Making Models;Principal Component Analysis;Hierarchical Analysis;Data Ming

---

1.Introduction

With the explosive growth of Internet data, the vast amount of data contains enormous economic value, making the effective processing and mining of data a critical issue in today's society. Faced with such large-scale data, data analysis techniques and statistical models play a key role in extracting useful information, discovering hidden patterns, and supporting decision-making[8-11]. Data mining, as an important tool, is often used to build mathematical models to analyze and process data, generating the results people need[13-14].

This paper aims to comprehensively summarize the core methods in the field of data analysis and pro-

ISSN 2959-6157

cessing technology. First, we introduces the commonly used data preprocessing techniques. As the foundational step in data analysis, data preprocessing directly influences the accuracy and efficiency of subsequent models, making a systematic review of these techniques crucial. Next, the paper provides a detailed description of several commonly used data optimization models, as well as predictive and decision-making models, focusing on methods like linear programming and dynamic programming. These models not only provide a solid mathematical framework for data analysis in theory but are also widely applied in solving complex decision-making problems in practice.

In discussing the models, the paper elaborates on their basic principles, processing steps, and specific applications in data analysis and processing. Finally, the paper summarizes the data preprocessing techniques and model-building methods, while reflecting on the current challenges faced in the data field and offering insights into future research directions. As data volume continues to grow and data types diversify, how to maintain processing efficiency while improving model accuracy and scalability will become a significant topic in future data analysis research.

2. Data preprocessing

Large-scale data in real life are often messy and incomplete. Data preprocessing refers to the necessary computer-friendly cleaning, screening and other processing of the collected data prior to transforming or enhancing most of the observed data, with the aim of transforming the raw data into a suitable form that can be used for modeling and

analysis to effectively support decision-making and forecasting[1]. The main manifestations of data clutter are incompleteness, inconsistency, redundancy, noise, and so on. This type of data cannot be directly data mined or the mining results are extremely poor. In order to improve the quality of data mining has generated a variety of data preprocessing techniques, such as data cleaning, data integration and data transformation[2].

2.1 Data cleansing and integration

Data cleansing refers to performing operations that include checking data for consistency, completeness and accuracy in order to deal with invalid and missing values, correctingdata errors, and so on[3]. Common cleansing methods include checking for and removing unnecessary duplicates to reduce data redundancy; missing data that may be due to survey, coding, and entry errors, such as the name of a supplier, regional information about a customer, etc., can be made up by estimating the sample mean or using interpolation[4]. In addition, data records containing missing values can optionally be excluded, provided the data base is large enough. This approach reduces the number of records available for analysis but leaves integrity intact[5].

Data integration is the consolidation of data from different sources, formats and nature of characteristics into a unified dataset to provide decision makers with a consistent view of the data for analyses and mining to help them make better decisions[7].

Data transformations, on the other hand, are a method of transforming data into a form suitable for mining, which covers four common methods: data smoothing, data aggregation, attribute construction and data normalization.

2.2 Data transformation

Data smoothing, which removes data noise and discretizing continuous data. The main algorithms are binning, clustering and regression. Data aggregation, which summarises data, for example aggregation of the number of accidents per hour into the number of accidents per day[12]. Attribute construction, which involves constructing new attributes with existing given attributes and adding them to the dataset. For example, profit is constructed from the attributes sales and cost, and the process involves only simple transformations of the corresponding data.

The purpose of data normalization is to limit the range of values of an attribute to a specific range, eliminating the bias in data mining results caused by inconsistent scaling. For example, in order to normalize the range gap caused by the difference in values between the area attribute and the perimeter attribute of rice, the data is scaled to a small specific interval to eliminate the effect of unit inconsistency[15]. Data Normalization aims at limiting the range of values of an attribute to a specific range, eliminating the bias of data mining results caused by inconsistent scales, in order to eliminate the effects of unit inconsistencies.

About data normalization in data transformation, there are three methods of normalizing values for computation, which are min-max normalization, z-score normalization, and decimal scaling normalization. Among them, min-max normalization determines the interval of normalized attribute values usually as [0,1], and the normalization result is obtained by calculating the indicated metric of 1.

$$v' = \frac{v - \min_A}{\max_A - \min_A}$$

where minA and maxA are assumed to be the minimum and maximum values of attribute A, respectively, and the value v of A is mapped to the interval [0, 1] by Eq[16].

Z-score normalization on the other hand compares and analyzes the values between different data sets. In statistics and data analysis, Z-score standardization helps to normalize the data, making different data sets comparable to each other for better data analysis and mining. Z-score normalization is based on the principle of subtracting the mean from the original data and dividing it by the standard deviation, which is calculated by the following formula:

$$z = \frac{x - \mu}{\sigma}$$

In the formula, z is the normalized value, X is the original data, μ is the mean of the original data, and σ represents the standard deviation of the original data. Through Z score normalization, the raw data can be converted into a standard normal distribution centered on the mean and with standard deviation as the unit, which indicates the distance or deviation of a specific value from the mean, thus facilitating data analysis and comparison. Z-score normalization indicates the relative position of each raw data in the data set. For example, firms' weekly brown sugar production is standardized through normalization to allow for comparisons of production across firms.Decimal calibration normalization is a normalization calculation by shifting the position of the decimal point, usually mapped in the range [-1,1]. The conversion formula is:

$$X_{new} = \frac{X}{10^k}$$

where k is the number of decimal places to move, a value that depends on the maximum value of the absolute value of the attribute value. When dealing with datasets with very large or very small values, decimal normalization can effectively scale the data by simply moving the decimal point without the need for complex mathematical operations to map the values to smaller ranges.

In conclusion, the use of data preprocessing techniques greatly improves the quality of data mining and reduces the time required for actual mining. Preprocessed data is more easily understood by the model and helps to improve the accuracy and reliability of predictions.

3.Optimization Model

An optimization problem is a mathematical problem of finding an optimal solution under given constraints, usually with the goal of reducing project costs, reducing time consumption, increasing revenue, etc. The optimal solution is usually the case where a maximum or minimum value is obtained for some objective function. Optimization problems can involve a single variable or multiple variables, and constraints can take various forms, either equations or inequalities. Optimization models can be constructed based on linear programming, nonlinear programming, integer programming, and other algorithms. This chapter focuses on optimization problems in data analysis, and systematically introduces the definitions of the kinds of optimization models and elaborates on them. And then discusses the modeling steps of optimization models in real scenarios as well as the solution methods, and finally introduces the practical applications of optimization models.

3.1 Linear programming

Linear programming optimization algorithms are an important part of operations research and a mathematical method to assist people in conducting scientific management. As a quantitative method commonly used by enterprises for total production planning, it is the most theoretically sound and practically the most widely applied. Today it is widely used in space exploration, computer networks and problems such as optimization of food supply chains and energy supply. Based on the objective function and constraints to solve for the extreme values of the objective function within the constraints.Depending on the number of objective functions, linear programming problems can be subdivided into multi-objective linear programming and single-objective linear programming[21]. In a linear programming model, the decision variables, constraint conditions and objective functions are the three essential factors.The usual form is:

$$\max z = \sum_{j=1} c_j x_j \quad ①$$

$$\text{s.t.} \begin{cases} \sum_{j=1}^{n} a_{ij} x_j = b_i, ② \quad i = 1, 2, \ldots, m ③ \\ x_j \geq 0, ④ \quad j = 1, 2, \ldots, n ⑤ \end{cases}$$

Where Eq. 1 is the objective function, Eq. 2-5 are the constraints that must be satisfied, in the formulas,Cj, Xj are the decision variables.For example, to establish the function objective for the maximum profit, that is, to seek the maximum value of equation 1; constraints are set as equipment limitations, raw material limitations and two basic requirements (3, 5); the decision variables are the planned output of the product and the unit cost of the product.

Selection of appropriate evaluation metrics based on the problem and development of a linear programming mathematical model can help to seek the optimal solution to a practical problem. The objective of the study by Muammar Khadpi and other researchers was to optimize the production cost of sandals to maximize the profit[23]. In

ISSN 2959-6157

order to develop the production quantities of three different styles of sandals X1, X2 and X3.

Subsequently by understanding the production cost of each pair of sandals, the authors determined the objective function, which maximizing the profit from the three sandals. The constraint conditions identified include raw material limitations, operating costs and production limitations, among others. In the study, the simplex method(2-1) is used to solve the linear programming problem by modeling the objective function incorporating the constraints as a linear programming mathematical model, inputting it into the LINDO program to perform iterative computation to arrive at the optimal production plan, which resulted in maximum profit at minimum manufacturing cost[23].

3.2 Nonlinear programming

Nonlinear programming problems are similar to linear planning problems, which are an important branch of operations research. The difference is that its objective function or constraint conditions contain nonlinear functions in the programming problem. This kind of problem needs to solve the n-dimensional nonlinear function in a set of equations or inequality constraints of the extreme value of the problem to find the extreme value of a n- dimensional nonlinear function subject to a set of equations or inequality constraints, widely exist in the production management, quality control, product design and other fields[24]. Because the complex relationships in many practical problems are difficult to be accurately described by linear programming, nonlinear programming is very common in daily life.

The mathematical model of a nonlinear programming problem can be described as:

$$\text{Minimize } f(X)$$

$$\text{Subject to } h_j(X) \leq 0,\ j = 1, 2, ..., m$$

$$l_k(X) = 0,\ k = 1, 2, ..., l$$

$$x_i^{(l)} \leq x_i \leq x_i^{(u)},\ i = 1, 2, ..., n$$

Where f(x) is the objective function, as an example, in the study of Niyazi Cem GÜRSOY, this linear programming model is applied to optimize the airline's strategy of refueling at different airports with the goal of reducing the cost, then the objective function is denoted as the total fuel cost, and the optimal solution is to solve for the minimum value of f(x), $h_j(x)$ and $l_k(x)$ are the inequality and equation constraints, which involve fuel weight, aircraft performance, flight distance, etc. X_i is the decision variable, and Eq. 4 constrains the upper and lower bounds of the decision variable. The Generalized Reduced Gradient Method (GRG) is used in the thesis to solve these nonlinear planning problems. The GRG method optimizes the objective function by dividing the variables into basic and non-basic variables and using the constraint equations to directly compute the required changes in the basic variables. In practice, based on flight routes, flight distances, and fuel prices, the nonlinear planning model and the GRG method can be used to find the optimal refueling solution with savings ranging from 1% to 4% of the total fuel cost[25].

3.3 Integer Programming

Integer programming is a key branch of optimization theory that has important applications in several fields, including operations research, engineering design, economics, and computer science.Unlike traditional linear or nonlinear optimization problems, integer programming requires that the variables in the solution be integers. Integer programming was formed as an independent branch after the introduction of the cut-plane method by R.E. Gomory in 1958, and has evolved to solve more problems[26]. The importance of integer programming stems from the need for real world problems. In many cases, solutions to optimization problems need to be in integer form because the solutions represent indivisible items or decisions, such as problems of resource allocation, grid design, and time scheduling. For example, an integer programming model has been developed for shopping centers with constraints on the capacity of the shopping center, upper and lower bounds on the number of shoppable, etc. In this model, the total expense of all customers staying for short and long time in a period for shopping centers has been maximized under some generalized constraints. The results of the analysis allow the estimation of the potential customer base and the number of customers for different types of shopping centers. As a result, managers and decision makers can identify the strengths and weaknesses of their business and ultimately adopt more effective management strategies and business decisions.

Integer programming can be categorized into three different types based on the characteristics and constraints of its variables: pure integer programming, mixed integer programming and 0-1 integer programming. Due to its special integer constraints, integer programming problems are usually more difficult to solve than general linear or nonlinear programming problems, and the most frequently used methods are branch and bound and cutting plane method, each with its own advantages and limitations[28]. In practical applications, the above two methods are usually used to get as close as possible to the optimal solution in an acceptable period of time.

3.4 Dynamic programming

Dynamic programming is mainly used for solving optimization problems of dynamic processes with time-divided phases. The basic idea is to decompose the problem to be solved into a number of sub-problems, and from the solution of the sub-problems to the solution of the problem to be solved, the active process is divided into a number of phases that can be interconnected[29]. Therefore, dynamic planning is often used to deal with multiple solutions and conflicting types of problems, such as resource allocation problems, Fibonacci series problems, the problem of finding the most value, etc., through the inverse deduction of the optimal strategy at any stage, to provide managers with a valuable reference basis to help make decisions. For example, logistics companies can use historical data as a source to create an effective optimization method based on dynamic planning and parallel information processing technology, construct a dynamic planning model to predict the maximum amount of goods in stock at each stage, and minimize the cost of overproduction and overstocking.Where the objective function is F(x,z).

$$x^{\min} \le x_m \le x^{\max}, \ m \in \overline{1, M}, \quad ②$$

$$F(x,z) = F1(x,z) + F2(x,z) \to \min_{x},$$

$$F1(x,z) = \sum_{m=1}^{M} w1 \cdot \max\{0, z^{\min} - (x_m + z_{m-1} - D_m)\},$$

$$F2(x,z) = \sum_{m=1}^{M} w2 \cdot \max\{0, x_m + z_{m-1} - D_m - z^{\max}\},$$

$$z_m = x_m + z_{m-1} - D_m,$$

Where F1 and F2 are the combined functions constituting the objective function, the cost of stockouts and storage cost of goods[32]. There is a restriction on the quantity of goods to be purchased from the supplier, i.e., 2. Subsequently, through the recursive equations, the principle of optimization is flexibly applied through the code in order to predict a better inventory management decision.

3.5 Discussion

Optimization models, including linear programming, nonlinear programming, integer programming, and dynamic programming models, are essential tools for decision making in various fields.

In the field of decision making, optimization models such as linear programming, nonlinear programming, integer programming, and dynamic programming are indispensable tools in decision-making in various fields Linear programming is mainly focuses on maximizing or minimizing a linear objective function and is limited by a series of linear constraints, making it particularly suitable for problems related to resourc allocation. On the other hand, nonlinear programming can adapt to objective functions or constraints of nonlinear properties, so it has greater versatility when dealing with complex situations. Integer programming is an extension of linear programming and requires integer solutions, so it is suitable for problems that require integer variables.

4. Data Reduction and Decision Models

4.1 Analytic Hierarchy Process

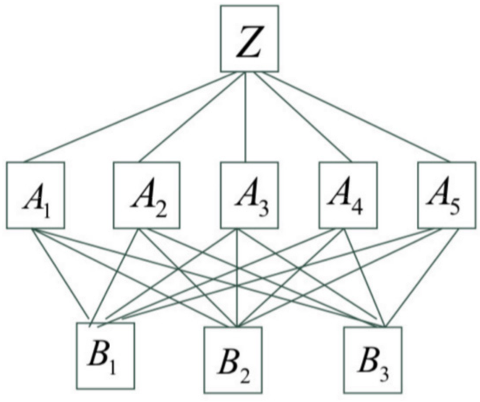
Analytic Hierarchy Process (AHP) is a hierarchical weighting decision analysis method proposed by applying network system theory and multi-objective comprehensive evaluation method[33]. AHP helps decision makers to make reasonable decisions in the face of complex problems with multiple choices and criteria by breaking down the complex problems into multiple levels, making the decision-making process mathematical by utilizing less quantitative information, and modeling the complex systems that are difficult to be fully quantified. This section will introduce the basic principles, steps and application areas of hierarchical analysis in detail.

Hierarchical analysis refers to a complex multi-objective decision-making problem as a system, according to the nature of the problem and the overall goal to be achieved, the problem will be decomposed into a number of constituent factors or criteria, and then into a number of levels of multiple indicators. Then, through a series of pairwise comparisons to determine the relative importance of the factors and their interrelated effects, a structural model of multilevel analysis is formed. The researcher calculates the priority of each level of factors over the higher level factors, and then reduces the problem to the ranking of relative advantages and disadvantages, and selects the optimal solution with the highest final weight[34-37]. Modeling using hierarchical analysis can be roughly divided

ISSN 2959-6157

into four steps: problem decomposition, pairwise comparison, weight calculation and consistency check.

First of all, decompose the problem to establish a hierarchical structure, the figure 4-1 shows the hierarchical structure of the APH, usually the highest level Z for the purpose of decision-making and the problem to be solved; the middle level, A_1-A_5, for the consideration of the factors and decision-making criteria; the lowest level B_1-B_3 for the decision-making alternatives.



4-1

As an example, the study by Chenglu Ruan et al. developed a model for evaluating orthodontic programs using AHP to investigate its feasibility and effectiveness in clinical practice. The decision-making goal in this study was the optimal orthodontic program,with intermediate tier factors selected as orthodontic cost, comfort, aesthetics, cleaning difficulties, and so on. Alternatives B1-B3 were specific orthodontic methods such as invisible braces and fixed labial braces. Then the data were compared pairwise and a judgment matrix was constructed to determine the weights between the factors at each level (Fig. 4-1), based on the scaling method proposed by Thomas L. Saaty introduced in the 1970s, and has been widely used in the field of engineering, management and economics[39-40]. These factors were rated by a 1-9 scale to indicate the differences in relative importance and superiority between the factors, as shown. The use of the scaling method reduces the difficulty of comparison due to the different nature of the factors and improves the accuracy of the model.

To construct the judgment matrix, we assign A as the target and ua and ub (a,b=1, 2...n) as factors.The u denotes the relative importance between the rated factors and these values are used to form the A-U judgment matrix P. The prioritization ordering among the factors is then calculated based on the judgment matrix, which is also the basis for weight assignment. Hierarchical analysis requires consistency test to check the consistency of the judgment matrix. Since some bias occurs when people make subjective judgments, the consistency ratio (RC) is used to assess the reasonableness of the matrix. If the range of inconsistency exceeds the specified threshold, the matrix needs to be re-evaluated.

Finally, the weights of the levels are combined to calculate the final score for each alternative, and the options are ranked according to the high or low score derived from the matrix operation, so that the optimal option can be selected.

4.2 Principal component analysis

Principal Component Analysis (PCA) is a fundamental technique used in data science and statistics for dimensionality reduction and exploratory data analysis. When solving practical problems, the number of informative variables proposed to reflect the correlation is often quite large in order to be analyzed and considered comprehensively. PCA aims to transform a set of correlated variables into a smaller set of uncorrelated variables called principal components by linear transformation[42]. These components, such as the first and second principal components, are mutually orthogonal and capture most of the variance present in the original data. This makes them useful for summarizing and visualizing complex datasets while preserving essential information. PCA works by identifying the directions (principal components) along which the data varies the most.

The first principal component captures the largest possible variance in the data, with each succeeding component capturing the highest variance orthogonal (uncorrelated) to the previous ones. This orthogonality ensures that each principal component is a linear combination of the original variables and is uncorrelated with the others.

For example, principal component analysis was used in the study of shell and tube heat exchangers (STHEs)[44]. Twelve variables predictive of hydrothermal and hydraulic performance and cost, such as baffle spacing and shell diameter, were used as components of the analysis. The 12 indicators considered as 12 random variables, denoted as X1, X2, ..., X12. Principal Component Analysis (PCA) aims to transform the issue of these p indicators into a discussion of linear combinations thereof. These new indicators F1, F2, ..., F12, based on the principle of retaining the major information content, fully reflect the information of the original indicators and are mutually independent. The newly transformed features are called principal components (PCs). The original data is projected onto the new PCs to maximize the variance, and variables with high variance have higher weights.

The process of reducing multiple variable factors into a smaller set of principal components is known as dimensionality reduction, which involves different linear combinations of the original variables. Dimensionality reduction algorithms, such as Principal Component Analysis (PCA) and Kernel PCA (KPCA), utilize various computational methods and serve distinct purposes. The choice of dimensionality reduction method can significantly impact the resulting principal components. This technique is valuable for reducing the dimensionality of datasets while retaining essential information, leading to simpler models, improved visualization, and more efficient computations. Consequently, PCA is widely applied across diverse fields, including image and signal processing, genetics, finance, and marketing research. By effectively reducing dimensions, PCA aims to preserve as much relevant data as possible while eliminating redundant information, enhancing data handling capabilities.

Overall,PCA serves as a powerful tool for exploring and summarizing high-dimensional data, offering insights into the fundamental structures and relationships between variables. This understanding aids in better decision-making based on complex datasets.

Hierarchical analysis and principal component analysis (PCA) are powerful statistical techniques used for data analysis and dimensionality reduction.Hierarchical analysis organizes data into a tree structure, allowing researchers to understand relationships between variables by creating clusters based on similarities.In contrast, PCA converts correlated variables into a set of uncorrelated variables or principal components that capture the maximum variance of the data.While hierarchical analysis is particularly useful for identifying patterns and groupings in a dataset, PCA simplifies complex datasets and facilitates visualization and interpretation.Together, these methods provide complementary approaches for analyzing and understanding high-dimensional data in fields as diverse as psychology, biology, and the social sciences.

5.Conclusion

Data preprocessing, optimization models, hierarchical analysis and principal component analysis are essential components of modern data mining. Data preprocessing involves cleaning and transforming raw data to improve its quality and ensure that subsequent analysis produces accurate results. Optimization modeling focuses on identifying the best solution among alternatives, usually under constraints, to maximize or minimize a specific objective. Hierarchical analysis organizes data into clusters based on similarities,revealing patterns and relationships, while PCA reduces dimensionality. Current challenges in these areas include handling large datasets, ensuring data privacy, and addressing model complexity. Future research is likely to explore advanced algorithms for more efficient data preprocessing, combining machine learning techniques with optimization models, and improved methods for interpreting results from hierarchical and principal component analysis. In addition, the development of more robust frameworks for handling high-dimensional data and enhancing the interpretability of complex models will be critical to advancing these analytical approaches. As data continues to grow in volume and complexity, the integration of these techniques with emerging technologies like artificial intelligence and machine learning will become increasingly important. Future research will likely focus on enhancing the scalability and efficiency of these methods, as well as developing more adaptive models that can handle real-time data streams. The continuous evolution of data analysis techniques will open up new possibilities for more accurate predictions, better decision-making, and deeper insights into ever-expanding datasets, making these methods indispensable in tackling the challenges of the data-driven era.

References

[1]Jansen B J, Aldous K K, Salminen J, et al. Understanding Audiences, Customers, and Users Via Analytics: An Introduction to the Employment of Web, Social, and Other Types of Digital People Data[M]. Springer Nature, 2023.

[2]Albahra S, Gorbett T, Robertson S, et al. Artificial intelligence and machine learning overview in pathology & laboratory medicine: A general review of data preprocessing and basic supervised concepts[C]//Seminars in Diagnostic Pathology. WB Saunders, 2023, 40(2): 71-87.

[3]Côté P O, Nikanjam A, Ahmed N, et al. Data cleaning and machine learning: a systematic literature review[J]. Automated Software Engineering, 2024, 31(2): 54.

[4]Gomathi S. Data Migration and Cleaning[J].

[5]Chicco D, Oneto L, Tavazzi E. Eleven quick tips for data cleaning and feature engineering[J]. PLOS Computational Biology, 2022, 18(12): e1010718.

[6]Gracias A, Shad R, Broklyn P. DATA INTEGRATION AND ANALYSIS IN PRECISION MEDICINE[J].

[7]Nargesian F, Asudeh A, Jagadish H V. Responsible data integration: Next-generation challenges[C]//Proceedings of the 2022 International Conference on Management of Data. 2022: 2458-2464.

[8]Nascimento J, Silva J, Cupertino Bernardes R, et al. Statistical data transformation in agrarian sciences for variance analysis: a systematic review[J]. F1000Research, 2024, 13: 459.

[9]Drosu A, Cofaru C, Popescu M V, et al. The influence of data aggregation levels on accident severity researches[C]//IOP Conference Series: Materials Science and Engineering. IOP Publishing, 2024, 1311(1): 012041.

ISSN 2959-6157

[10]Pang K, Zou L, Kang N, et al. The construction of fuzzy linguistic attribute partial ordered structure diagram[J]. Computational and Applied Mathematics, 2023, 42(6): 240.
[11]Dudic B, Nguyen N T, Ašonja A. Data Normalization for Root Assessment Methodology[J]. International Journal of Industrial Engineering and Management, 2024: in press-in press.
[12]Celik A. Evaluating the impact of data normalization on rice classification using machine learning algorithms[J]. Acadlore Trans. Mach. Learn, 2024, 3(3): 162-171.
[13]Setiawan R, Imelda I. Application of min-max normalization in backpropagation to detect early malnutrition[J]. Indonesian Journal of Multidisciplinary Science, 2024, 3(10).
[14]Rahmad Ramadhan L, Anne Mudya Y. A Comparative Study of Z-Score and Min-Max Normalization for Rainfall Classification in Pekanbaru[J]. Journal of Data Science, 2024, 2024(04): [15]Pambudi E A, Fahrezi A G, Badharudin A Y. Forecasting Brown Sugar Production Using k-NN Minkowski Distance and Z-Score Normalization[J]. Journal of Information Systems and Informatics, 2023, 5(2): 580-589.
[16]Rahmad Ramadhan L, Anne Mudya Y. A Comparative Study of Z-Score and Min-Max Normalization for Rainfall Classification in Pekanbaru[J]. Journal of Data Science, 2024, 2024(04): [17]BALASUBRAMANIAM S. Practical Application of Linear programming technique in textile spinning industry[J].
[18]Alzalg B. Combinatorial and algorithmic mathematics: From foundation to optimization[M]. John Wiley & Sons, Inc., 2024.
[19]Chakraborty A, Chandru V, Rao M R. A linear programming primer: from Fourier to Karmarkar[J]. Annals of Operations Research, 2020, 287: 593-616.
[20]Khadapi M, Ramadani S. Optimization Of Sandal Production Using Linear Programming[J]. Journal of Artificial Intelligence and Engineering Applications (JAIEA), 2024, 3(3): 631-638.
[21]Nakayama K. Eigenvalue Problem and Nonlinear Programming Problem: For Economic Studies[M]. Springer Nature, 2023.
[22]Gürsoy N C, Alptekin N. Modelling Strategy of Airline Tankering with Nonlinear Programming[J]. Verimlilik Dergisi, 2023, 57(2): 365-392.
[23]Ponce-Ortega J M, Hortua A C, El-Halwagi M, et al. A property-based optimization of direct recycle networks and wastewater treatment processes[J]. AIChE Journal, 2009, 55(9): 2329-2344.
[24]Eisenbrand F, Hunkenschröder C, Klein K M, et al. Sparse Integer Programming Is Fixed-Parameter Tractable[J]. Mathematics of Operations Research, 2024.
[25]Ribeiro C C, Urrutia S, de Werra D. Combinatorial Models for Scheduling Sports Tournaments[M]. Berlin: Springer, 2023.
[26]Laaksonen A. Guide to competitive programming[M]. Cham: Springer, 2020.
[27]Hinderer K, Rieder U, Stieglitz M. Dynamic optimization[M]. Springer, 2016.
[28]Bellman R. Dynamic programming[J]. science, 1966, 153(3731): 34-37.
[29]ФЕДОРОВ Є, НЕЧИПОРЕНКО О, НЕСКОРОДЄВА Т, et al. DYNAMIC PROGRAMMING FOR SOLVE THE INVENTORY MANAGEMENT PROBLEM IN LOGISTICS[J]. Computer systems and information technologies, 2024 (1): 118-126.
[30]Buckley J J. Fuzzy hierarchical analysis[J]. Fuzzy sets and systems, 1985, 17(3): 233-247.
[31]Pauer F, Schmidt K, Babac A, et al. Comparison of different approaches applied in analytic hierarchy process–An example of information needs of patients with rare diseases[J]. BMC medical informatics and decision making, 2016, 16: 1-11.
[32]Stanton N A. Hierarchical task analysis: Developments, applications, and extensions[J]. Applied ergonomics, 2006, 37(1): 55-79.
[33]Tramarico C L, Salomon V A P, Marins F A S. Analytic hierarchy process and supply chain management: A bibliometric study[J]. Procedia Computer Science, 2015, 55: 441-450.
[34]Bridges Jr C C. Hierarchical cluster analysis[J]. Psychological reports, 1966, 18(3): 851-854.
[35]Raudenbush S W. Hierarchical linear models: Applications and data analysis methods[J]. Advanced Quantitative Techniques in the Social Sciences Series/SAGE, 2002.
[36]Hillerman T, Souza JCF, Reis ACB, Carvalho RN. Applying clustering and AHP433-459.
[37]methods for evaluating suspect healthcare claims. J Comput Sci. 2017;19.97-111. https://doi.org/10.1016/jjocs.2017.02.007.
[38]Marza B, Bratu R, Serbu R, Stan S, Oprean-Stan C. Applying ahp and fuzzy Ahp Management methods to assess the level of financial and digital inclusion. Economic Comput Economic Cybernetics Stud Res. 2021;55(4):165-82. https://doi.org/10.24818/18423264/55.4.21.11.
[39]Johnson R A, Wichern D W. Applied multivariate statistical analysis[J]. 2002.
[40]Abdi H, Williams L J. Principal component analysis[J]. Wiley interdisciplinary reviews: computational statistics, 2010, 2(4): 433-459.
[41]Makkulau M, Ampa A T, Yahya I, et al. Variance The Estimation Eigen Value of Principal Component Analysis and Nonlinear Principal Component Analysis[C]//ITM Web of Conferences. EDP Sciences, 2024, 58: 04001.
[42]Nargundkar A, Gulia V, Sur A. Multi-Variate Analysis of Shell & Tube Heat Exchanger using Principal Component Analysis[C]//E3S Web of Conferences. EDP Sciences, 2024, 552: 01025.