ISSN: 2688-8645

A Survey of Fuzzy Pattern Tree Classification Algorithms

Ji Zhang, Bin Ma*

Department of Computer, North China Electric Power University, Baoding, Hebei, 071000, China

*fj_mabin@163.com

Abstract

Data classification algorithm is the core content of big data mining. Its main function is to extract valuable knowledge and information,analyze the characteristics of all kinds of information, and provide data basis for further research. With the wide application of data mining technology, data classification algorithms continue to emerge and gradually improve. The classic classification algorithms include decision tree classification algorithm, naive Bayes algorithm, support vector machine classification algorithm, artificial neural network classification algorithm, fuzzy pattern tree (FPT) and so on. This paper summarizes several common algorithms in data classification algorithms, and analyzes their characteristics to understand their algorithm principles and application scenarios.

Keywords

FPT; Classification Algorithm; Data Classification.

1. Overview

Big data mining technology mainly refers to the process of collecting and dividing data information based on a specified attribute from massive information data, gradually obtaining and accumulating some effective information. Data mining technology, as a product of the development of network information technology in the era of big data, mainly involves artificial intelligence, databases, statistics, etc., and involves a lot of research content. One of the more important research branches is classification. Data classification is the basis for data analysis and obtaining correct analysis results. The data classification process generally includes two steps. The first step is to construct a model through a training set of data with known class labels. This step is often referred to as the training stage, which can be understood as training a classifier; The second step is to use the model to classify objects with unknown class labels. From this process, we can know that classification models not only need to fit known data sets, but also need to accurately predict unknown objects. Different classification algorithms are suitable for different application scenarios. The differences in classification algorithms will simulate different classifiers, which will directly affect the accuracy of classification and ultimately affect data analysis. Therefore, implementing deep classification for data with relatively complex scale systems or large amounts of data information, and selecting a reasonable classification algorithm, all have an important impact on task completion. Currently, the research on classification algorithms related to big data mining technology in the field of computer data science at home and abroad mainly focuses on the following two aspects: one is to directly apply traditional classification algorithms to actual cases, or to make simple combinations of traditional algorithms, and then apply them to actual cases. The other is to improve and upgrade traditional classification algorithms using new technologies and ideas. How to choose an appropriate classification algorithm for practical applications? Below is a description and analysis of several classic classification algorithms, summarizing the characteristics, advantages, and disadvantages of each algorithm [1]. In this paper, the

ISSN: 2688-8645

development of fuzzy pattern tree (FPT) in recent years is briefly analyzed, and an improved FPT algorithm based on multi strategy fusion is proposed to solve the classification problem.

2. Research Status and Development Trend

Currently, the classic classification algorithms in the stage of big data analysis and data mining mainly include decision tree, naive Bayesian, support vector machine (SVM), neural network classification algorithm, fuzzy pattern tree (FPT), and so on.

Decision tree classification algorithm is one of inductive learning algorithms, which mainly refers to classification rules that infer a "tree" structure from a series of irregular and unordered sample data information to predict. The decision tree classification algorithm can intuitively display the decision-making problems and key points at different stages of the entire decision-making process. The decision tree consists of a root node, an internal node, a leaf node, and the directed edges of the connecting nodes. The root node is unique and represents the set of samples to be classified; Internal nodes represent feature attributes; Leaf nodes represent classification results. The algorithm decision-making process starts from the root node, selects branches from top to bottom to reach the corresponding node according to the corresponding attribute values in the set to be classified, and repeats this step until it reaches the leaf node. The category stored by the leaf node is used as the classification result. For example, if a person applies for a loan at a bank, the bank will determine whether to approve the loan application based on the applicant's annual income, property status, marital status, and other conditions. This process can be expressed in the form of a decision tree, as shown in Figure 1.

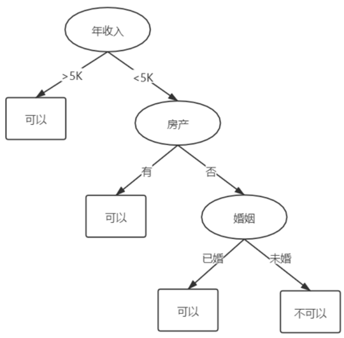

Figure 1. Decision Tree

Currently, there are many types of decision tree algorithms, including ID3 [2], C4.5 [3], and CART [4], among which C4.5 is an optimization improvement on ID3. Compared with other types of classification algorithms, the decision tree algorithm has the following advantages: First, the decision tree algorithm has clear logic, clear hierarchy, and intuitive, and its classification rules are easy for people to understand and implement, making it a relatively friendly classification algorithm. Secondly, the decision tree algorithm has high classification accuracy. In the data mining process, using the decision tree classification algorithm, each node corresponds to a classification rule, which can accurately classify each data to a leaf node. Secondly, the decision tree algorithm runs efficiently and takes less time. In addition, although the classification algorithm of decision trees has many advantages in the application stage, it can also cause problems such as over fitting. When generating a decision tree, when completely following the training set, if there are noise samples in the training set, the noise will also be fitted while fitting the training set, resulting in overly complex classification models with too many branches. Some branches are unique to the training sample itself and are not representative, resulting in over fitting during the testing phase, resulting in low accuracy of the model [5]. We need to discretize or pre order attributes based on the selected classification algorithm to achieve classification and learning as much as possible. Avoid situations where

ISSN: 2688-8645

there are more or fewer categories, resulting in over fitting, and other situations that affect prediction accuracy. Based on the analysis of the advantages and disadvantages of its decision tree, experts and scholars have improved it. In the process of forming and constructing a decision tree, pruning is usually used to reduce the impact of noise on classification. If the number of features is large, it is necessary to prune features that do not have sufficient classification capabilities based on global considerations after establishing a decision tree, reducing the complexity of the model, and making the decision tree have better generalization capabilities. There are two methods of pruning: one is called pre pruning, which stops the process of constructing a tree in advance. For example, setting the maximum depth of the decision tree in advance or setting constraints on certain feature attributes in the sample set in advance. The other method is called post pruning, which determines whether certain branches need to become nodes after the entire tree is generated and prunes them [6]. Pruning should be appropriate, avoiding both over fitting and under fitting [7].

Naive Bayesian algorithm is a commonly used algorithm in supervised learning. Its operation and principle are relatively simple, mainly based on the famous Bayesian formula:

$$P(Y|X)=(P(X|Y)*P(Y))/P(X)$$

Solving the problem of posterior probability through prior probability and conditional probability [3]. Assuming that the characteristic attributes of the sample dataset are independent of each other, when the conditional independence assumption is established, the classification probabilities P (y1), P (y2)... P (yn) in the training sample are known. By calculating the probabilities P (x1, x2... xn | y1), P (x1, x2... xn | y2)... P (x1, x2... xn | yn) of the characteristic attributes of the known classification, the classification with the attribute data to be characterized can be predicted, i.e. comparing P (y1 | x1, x2... xn) P (y2 | x1, x2... xn)... P (yn | x1, x2... xn) The object with the highest probability is classified as the object. The formula can ultimately become:

$$\hat{y} = \underset{y}{\operatorname{argmax}}\, P(y)\prod_{i=0}^{n} p(x_i \mid y)$$

The operational characteristics of the naive Bayesian classification algorithm [8] are mainly as follows: First, the logical idea of the naive Bayesian algorithm is very simple, with strong operability and feasibility. Secondly, naive Bayesian algorithm is relatively stable and will not have a significant impact on classification results due to the different characteristics of the data itself. Thirdly, the stronger the independence between naive Bayesian data, the more accurate their classification results. However, we need to note that this classification algorithm needs to be based on the assumption of conditional independence, which is an ideal state. In practical applications, there may be connections between data attributes, which reduces classification accuracy. Therefore, this method is often difficult to achieve the theoretical maximum in terms of effectiveness. The probability distribution of a class population and the probability distribution function of various samples can be obtained by expanding the sample training set. In addition, during the classifier testing phase, if there are characteristic attributes in the test sample that are not available in the training set, the probability of all categories will be 0 regardless of how they are calculated. At this time, it is necessary to perform smoothing processing to add 1 to each sample value, and add 1 to the numerator and N * 1 to the denominator when calculating the probability. This method is called Plass smoothing processing. In actual use, lambda (1 ≥ lambda ≥ 0) can be used instead of simply adding 1. Another problem we may encounter is when calculating the product, because the probability is less than 1, the result of multiplying two numbers less than 1 will be even smaller, even if it directly becomes zero after rounding, resulting in underflow. At this time, it is necessary to take the natural logarithm of the product result to solve such problems.

K-neighbor algorithm [9] is an instance-based classification method. The specific working principle of the K-neighborhood algorithm is that it has a sample data set, and each data in the sample set has a label, that is, the corresponding relationship between each data in the sample set and the attribution category is known. After inputting new data without labels, compare each feature of the new data with the corresponding feature of the data in the sample set, and then extract a classification label for the most similar data (the closest) in the sample. Generally, only the data in the sample dataset that is most similar to k, that is, K proximity, is selected. Typically, K is an integer greater than 20. Finally, select the category with the most frequent occurrences among the K most similar data as the classification of the new data. K-neighbor method is a lazy learning method that stores samples until classification is needed. If the sample set is complex, it may incur significant computational overhead, making it difficult to apply to real-time situations.

Support Vector Machine [10] is a learning method based on statistical learning theory. It is a binary classification model. Its basic model is defined as a linear classifier with the largest interval in the feature space. Its learning strategy is to maximize the spacing and ultimately transform it into a solution to a convex quadratic programming problem. Its biggest feature is to construct the generalization ability of the learning machine by maximizing the classification interval, and construct the optimal classification hyperplane according to the structural risk minimization criterion, which better solves the problems of nonlinearity, high-dimensional, and local minima. For classification problems, the support vector machine algorithm calculates the decision surface of the region from the samples in the region, thereby determining the category of unknown samples in the region. It has high classification accuracy and good adaptability. However, processing large datasets is slower.

Neural network refers to artificial neural network [7], which simulates the structure and function of the human brain based on network topology knowledge to form an effective operational model, mainly including three parts: input layer, hidden layer, and output layer. A neural network is composed of a large number of nodes connected to each other, each node representing a specific output function, and the connection between each two nodes represents the weighted value of the signal passing through the connection, that is, the weight. Each layer of nodes weights the sum of input information and performs nonlinear transformation before outputting it. The output value is used as the input value of the next layer, and so on until the final classification node [11]. Common types of neural networks include single layer neural networks, two-layer neural networks, multi-layer neural networks, convolutional neural networks, and cyclic neural networks. In the learning stage of the neural network, the final output value is gradually approximated to the true value by adjusting the weight of each connection, and ultimately the accurate model is achieved. After training, the neural network dynamically responds to the input information and obtains classification results from the output. There are many neural network classification algorithms, such as BP neural network, RBF neural network, self-organizing feature mapping neural network, and learning vectorization neural network. Currently, BP neural network is widely used. The main characteristics of neural network classification algorithms are: First, neural networks have strong learning ability. Secondly, due to the effect of weights, neural networks have better robustness in noisy environments. Thirdly, the artificial neural network classification algorithm also has good predictive classification ability for untrained data. Fourth, because artificial neural networks are nonlinear models, they can adapt to various complex data relationships. At the same time, the shortcomings of artificial neural network classification algorithms are also prominent, mainly due to the establishment of the neural network itself. Building a relatively complete neural network requires a long learning process, and the selection and combination of activation functions, optimization functions, and loss functions can also affect the accuracy of the final model, making it difficult to work. Some scholars propose to prune the

ISSN: 2688-8645

network before extracting neural rules to delete the divine elements and chain branches that have a negligible impact on classification accuracy, thereby simplifying the neural network. Secondly, compared to decision tree classification algorithms, neural networks have poor interpretability, which may be difficult for non-technical users.

Fuzzy classification is one of the important applications of fuzzy set and fuzzy logic related research. Its goal is to find a set of fuzzy rules and form a classification model. The main advantages of using fuzzy rules in classification applications are maintaining transparency and high accuracy. Machine learning based on fuzzy sets has been widely studied. Wang and Mendel [12] proposed an algorithm for generating fuzzy rules through instance learning. Inspired by Quinlan's classical decision tree induction, fuzzy decision trees (FDTs) have already done a lot of work. For example, Yuan and Shaw [14] proposed a method for fdt induction using fuzzy entropy. Janikow [15], Olaru, and Wehenkel [16] give different FDT generalizations. Su á rez and Lutsko [17] and Wang et al. [18] proposed the optimization of fdt. There are also other fuzzy based machine learning algorithms. For example, Chen et al. [19] proposed a machine learning method based on a subset. Rasmani and Shen [20] proposed a learning method based on weighted fuzzy subsets. In order to avoid exponential growth in the size of the rule base when the number of input variables increases, Raju et al. proposed a hierarchical fuzzy system. K ó czy et al. and Wong et al. [21] proposed fuzzy signatures to model the complex structure of data points in a hierarchical manner.

Huang Zhiheng et al. [22] proposed a PT induction method based on similarity measurement and fuzzy clustering, which improves classification accuracy compared to other methods; Robust against over fitting; The advantages of a compact tree structure are maintained. Robin Senge et al. [23] proposed a generalization and extension of an FPT learning algorithm. The flexibility of the average operator of the nodes within the pattern tree is increased, making the algorithm faster without affecting its prediction accuracy. Ammar Shaker et al. [24] proposed an evolutionary version of fuzzy pattern tree learning, which achieves model adaptation by predicting possible local changes in the current model and verifying these changes through statistical hypothesis testing. Su Pan et al. [25] proposed a method for constructing fuzzy pattern trees based on pre aggregation, and proved that the performance of fuzzy pattern trees using pre aggregation functions generated by nilpotent minimum t-norm is superior to other pre aggregation functions and commonly used ordered weighted average operators. Rodrigues dos Santos et al. [26] proposed a new fuzzy pattern tree induction method, which uses Cartesian genetic programming as a learning algorithm and uses NSGA II for learning in a multi-objective environment. The results show that compared to some existing optimal classifiers, FPT-CGP MO has competitive performance in classification tasks, but provides an interpretable model from which knowledge obtained during the learning process can be extracted. Various changes in multi-objective methods and standard CGP algorithms have resulted in a wasteless evaluation and smaller tree search. FPT is a viable alternative to classic rule based fuzzy models because their hierarchical structure allows for more compact representation and a compromise between accuracy and model simplicity. Zhang Xinmin et al. [27] proposed a new integrated pattern tree model to predict HMT. Integrated pattern tree is a robust nonlinear modeling method that aggregates a set of pattern tree models into a single prediction model through bagging techniques.

In addition, Muhammad Aminul Islam et al. [31] proved that the fuzzy Choquet integral (ChiI), a powerful nonlinear aggregation function, can be expressed as a multi-layer network, and proposed an improved ChiIMP (iChIMP), which implements optimization based on random gradient descent based on the exponential number constrained by the ChiI inequality. Therefore, we envisage that this method can be applied to the construction of fuzzy pattern trees, enabling them to improve their execution efficiency.

3. Conclusion

From the above research status, many scholars have made a relatively good summary of existing classification methods, including decision tree learning, naive Bayesian method, K-nearest neighbor algorithm, artificial neural network algorithm, fuzzy decision tree learning, fuzzy pattern tree learning, and so on. We can find that the models obtained by classification algorithms based on neural network algorithms have poor interpretability, and classification algorithms based on decision tree learning are prone to over fitting, ignoring the correlation between attributes; FPT has good interpretability and can generate compact models that are very attractive in interpretation based on attribute features. However, FPT has a fatal disadvantage, which is high computational complexity. In the case of large amounts of data, the running time may be unacceptable. Therefore, an improved FPT algorithm based on multiple fusion methods (including operator selection optimization, introduction of neural network models, etc.) is proposed to achieve an improved running time without losing accuracy as much as possible.

References

[1] Yin Tingjun, Li Linghui, Zhou Rui. Overview of data classification algorithms in big data mining [J]. Digital Technology and Application, 2021,39 (01): 102-104.

[2] Yu Jianjun, Zhang Qiongzhi. Rough Set Based Decision Tree ID3 Algorithm [J]. Computer System Application, 2020,29 (04): 156-162.

[3] Guo Xingchen, Wang Qingqing, Wang Ya. Research on the application of decision tree algorithm in medical data classification [J]. Journal of Anqing Normal University (Natural Science Edition), 2021,27 (02): 49-53.

[4] Chen Leishi, Gao Xiaxia, Liao Yufang, Deng Jianbo, Zhou Bi. A Classification Method for Wetlands in Dongting Lake Area Based on the CART Decision Tree [J]. Surveying and Mapping Bulletin, 2021, {4} (06): 12-15.

[5] Li Yingchun Research on Decision Tree Classification Algorithms in Data Mining [D]. Hunan Normal University, 2015.

[6] Lin Yifeng Application of Decision Tree Algorithm in Student Achievement Prediction [D]. Ningbo University, 2019.

[7] Cheng Yifang. Overview of Data Classification Algorithms in Data Mining [J]. Digital Communications World, 2021, {4} (02): 136-137+140.

[8] Chen Hong et al. Improved naive Bayes classification algorithm for traffic risk management[J]. EURASIP Journal on Advances in Signal Processing, 2021, 2021(1).

[9] Salau Jennifer et al. Determination of Body Parts in Holstein Friesian Cows Comparing Neural Networks and k Nearest Neighbour Classification. [J]. Animals: an open access journal from MDPI, 2020, 11(1).

[10] Wang Shaowei, Xu Feng, Chao Haitao, Liu Yu, Zuo Dunwen. Research on fault classification of textile hot rolling mills based on support vector machines [J]. Modern Manufacturing Engineering, 2021, {4} (06): 116-121.

[11] Yang Fan, He Zhengwei, He Fan. Research on channel water depth prediction method based on depth network and ship traffic flow [J]. Journal of Wuhan University of Technology (Traffic Science and Engineering Edition), 2019,43 (01): 130-135.

[12] L. X. Wang and J. M. Mendel, “Generating fuzzy rules by learning from examples,” IEEE Trans. Syst., Man, Cybern., vol. 22, no. 6, pp. 1414–1427, Nov./Dec. 1992.

[13] J. R. Quinlan, “Decision trees and decision making,” IEEE Trans. Syst.,Man, Cybern., vol. 20, no. 2, pp. 339–346, Mar./Apr. 1990.

[14] Y. Yuan and M. J. Shaw, “Induction of fuzzy decision trees,” Fuzzy Sets Syst., vol. 69, no. 2, pp. 125–139, 1995.

[15] C. Z. Janikow, “Fuzzy decision trees: Issues and methods,” IEEE Trans. Syst., Man, Cybern., vol. 28, no. 1, pp. 1–14, Feb. 1998.

[16] C. Olaru and L. Wehenkel, “A complete fuzzy decision tree technique,”Fuzzy Sets Syst., vol. 138, no. 2, pp. 221–254, 2003.

[17] A. Suarez and J. F. Lutsko, “Globally optimal fuzzy decision trees for ´classification and regression,” IEEE Trans. Pattern Anal. Mach. Intell.,vol. 21, no. 12, pp. 1297–1311, Dec. 1999.

[18] X. Wang, B. Chen, G. Olan, and F. Ye, “On the optimization of fuzzy decision trees,” Fuzzy Sets Syst., vol. 112, no. 1, pp. 117–125, 2000.

[19] S. M. Chen, S. H. Lee, and C. H. Lee, “A new method for generating fuzzy rules from numerical data for handling classification problems,” Appl. Artif. Intell., vol. 15, no. 7, pp. 645–664, 2001.

[20] K. A. Rasmani and Q. Shen, “Weighted linguistic modelling based on fuzzy subsethood values,” in Proc. IEEE Int. Conf. Fuzzy Syst., 2003, vol. 1, pp. 714–719.

[21] L. A. Zadeh, “Fuzzy sets,” Inf. Control, vol. 8, no. 3, pp. 338–353, 1965.

[22] Z. Huang, T. D. Gedeon and M. Nikravesh, "Pattern Trees Induction: A New Machine Learning Method," in IEEE Transactions on Fuzzy Systems, vol. 16, no. 4, pp. 958-970, Aug. 2008, doi: 10.1109/TFUZZ.2008.924348.

[23] R. Senge and E. Hüllermeier, "Top-Down Induction of Fuzzy Pattern Trees," in IEEE Transactions on Fuzzy Systems, vol. 19, no. 2, pp. 241-252, April 2011, doi: 10.1109/TFUZZ.2010.2093532.

[24] Ammar Shaker and Robin Senge and Eyke Hüllermeier. Evolving fuzzy pattern trees for binary classification on data streams[J]. Information Sciences, 2013, 220 : 34-45.

[25] P. Su, T. Chen, H. Mao, J. Xie, Y. Zhao and J. Liu, "On the Application of Preaggregation Functions to Fuzzy Pattern Tree," 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2019, pp. 1-6, doi: 10.1109/FUZZ-IEEE.2019.8858922.

[26] A. Rodrigues dos Santos, J. L. Machado do Amaral, C. A. Ribeiro Soares and A. Valladão de Barros, "Multi-objective Fuzzy Pattern Trees," 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2018, pp. 1-6, doi: 10.1109/FUZZ-IEEE.2018.8491689.

[27] Xinmin Zhang and Manabu Kano and Shinroku Matsuzaki. Ensemble pattern trees for predicting hot metal temperature in blast furnace[J]. Computers and Chemical Engineering, 2019, 121 : 442-449.

[28] Robin Senge et al. Fuzzy Pattern Trees: Ein alternativer Ansatz zur Fuzzy-Modellierung[J]. Methoden und Anwendungen der Steuerungs-, Regelungs- und Informationstechnik, 2012, 60(10) : 622-629.

[29] Dovom E M , Azmoodeh A , Dehghantanha A , et al. Fuzzy Pattern Tree for Edge Malware Detection and Categorization in IoT[J]. Journal of Systems Architecture, 2019, 97:1-7.

[30] R. Senge and E. Hüllermeier, "Fast Fuzzy Pattern Tree Learning for Classification," in IEEE Transactions on Fuzzy Systems, vol. 23, no. 6, pp. 2024-2033, Dec. 2015, doi: 10.1109/ TFUZZ. 2015. 2396078.

[31] M. Islam, D. T. Anderson, A. J. Pinar, T. C. Havens, G. Scott and J. M. Keller, "Enabling Explainable Fusion in Deep Learning With Fuzzy Integral Neural Networks," in IEEE Transactions on Fuzzy Systems, vol. 28, no. 7, pp. 1291-1300, July 2020, doi: 10.1109/TFUZZ.2019.2917124.