2018 4th International Conference on Systems, Computing, and Big Data (ICSCBD 2018)

Grade Evaluation Model Based on Fuzzy Decision Tree

Xinyue Zhang, Tinghuan Wang

Statistic Institute, Shanxi University of Finance and Economics, wucheng road, Taiyuan, China

Keywords: data mining;fuzzy decision tree; ID3 algorithm; performance evaluation

Abstract: Aiming at the problem that the general decision tree classification method can't deal with the data ambiguity and uncertainty well, this paper proposes a performance evaluation model based on fuzzy decision tree to realize the student's academic level prediction based on daily behavior. In this paper, the mathematical attribute and expert suggestion method are used to determine the model attribute index, the fuzzy membership function of the design is used to fuzzy the data, the fuzzy matrix is established, and the decision tree ID3 algorithm is used to make decision analysis on the achievement information related to the campus behavior of college students. The fuzzy decision tree of this model can correctly and efficiently and comprehensively analyze and predict student achievement, and provide an important basis for the information construction and teaching management decision-making work of colleges and universities.

1. Introduction

Data mining is an interdisciplinary field based on multidisciplinary integration. Data mining can mine hidden but unknown useful information and knowledge from a large number of incomplete, noisy, fuzzy and random data. Decision tree classification method is an effective data mining method. However, this approach does not handle data ambiguity and uncertainty very well. Fuzzy decision tree algorithm is a generalization of decision tree algorithm. The fuzzy decision tree combines the advantages of fuzzy theory and decision tree. It not only has strong decision analysis capabilities, but also handles ambiguity and uncertainty. In this paper, we use fuzzy ID3 algorithm to construct a performance evaluation prediction model based on students' daily behavior. When you divide the level of a performance model attribute, clearing the boundary does not correctly describe the attribute level. Therefore, this paper uses the combination of fuzzy theory and decision tree to analyze the relationship between student attendance, self-learning time, library borrowing and dormitory learning atmosphere and student achievement, in order to achieve the purpose of prediction.

2. Construction of Grade Model Based on Fuzzy Decision Tree

2.1 The Basic Principle of Fuzzy Decision Tree

The decision tree algorithm is characterized by high quality, high efficiency classification with fewer attribute values. The thinking of college students is still immature, and their behavior is sometimes accidental and sudden. The decision tree generated by the traditional decision tree algorithm is incompatible with the mutated data, resulting in a cumbersome decision tree structure and inaccurate decision results. Therefore, this paper uses the combination of fuzzy theory and ID3 to analyze the behavior data and obtain the student performance evaluation model. The core principles of the fuzzy decision tree are as follows:

1) Indicator fuzzy processing
2) Establishing fuzzy matrix
3) The establishment of fuzzy decision tree

This paper designs a decision analysis model through an improved fuzzy decision tree. The model framework is shown in Figure 1.

  DOI: 10.25236/icscbd.2018.016

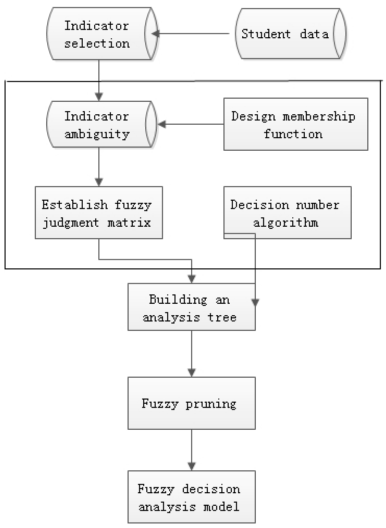

Figure 1 The model framework

2.2 Data Blurring

In the many behavioral indicators that affect students' academic performance, this paper selects the classroom attendance rate, self-learning time, library borrowing volume and dormitory learning atmosphere as the node attributes of the student achievement decision tree, and selects the student's final score as the node attribute of the decision tree. Let m be the division of the attribute level, and a be the center point that distinguishes the attribute level. The attribute A_{ij} (the jth element of attribute i) has a fuzzy membership matrix of C_i at level m_k and a matrix element of C_k^j, where j=1,2,3, k=1, 2, 3, 4, a_1, a_3 are the central points that distinguish the attribute levels, a_2 is the average of a_1 and a_3. The fuzzy set is described by the membership function. In the classical set, the feature function can only take two values of 0 and 1, while in the fuzzy set, the range of the feature function is expanded from the set of two elements to the continuous value of the [0,1] interval. In order to overcome the difference of numerical meanings, this paper designs the most common fuzzy membership function: the triangle membership function, and solves the membership level of the attribute element segmentation level degree:

$$\begin{cases} C_1^j(x) = 1, & x \le a_1 \\ C_2^j(x) = \dfrac{a_2 - x}{a_2 - a_1}, & a_1 < x < a_2 \\ C_3^j(x) = 0, & x \ge a_2 \end{cases} \qquad \begin{cases} C_1^j(x) = 0, & x \le a_1 \\ C_2^j(x) = \dfrac{x - a_1}{a_2 - a_1}, & a_1 < x \le a_2 \\ C_3^j(x) = \dfrac{a_3 - x}{a_3 - a_2}, & a_2 < x \le a_3 \\ C_4^j(x) = 0, & \mathrm{x} > a_3 \end{cases}$$

$$\begin{cases} C_1^j(x) = 0, & x \le a_2 \\ C_2^j(x) = \dfrac{x - a_2}{a_3 - a_2}, & a_2 < x < a_3 \\ C_3^j(x) = 1, & x \ge a_3 \end{cases}$$

The fuzzy membership function distribution is shown in Figure 2:

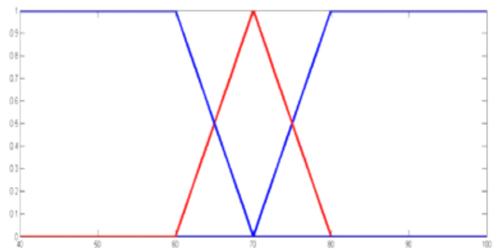

Figure2 The fuzzy membership function distribution

Thus, the fuzzy membership matrix C_i^j is a p*k order matrix, where $C_i^j \in [0,1]$. The concrete representation is as shown in equation:

$$\mathrm{C} = \begin{bmatrix} C_1^1(x) & C_2^1(x) & C_3^1(x) \\ C_1^2(x) & C_2^2(x) & C_3^2(x) \\ \vdots & \vdots & \vdots \\ C_1^p(x) & C_2^p(x) & C_3^p(x) \end{bmatrix}$$

2.3 Building a Fuzzy Decision Tree

The student achievement evaluation model established in this paper starts to test the sample node attributes gradually from the root node and walks down the corresponding branches until the sample nodes are reached. The node attributes obtained at this time are the node attribute conditions of the sample. According to the evaluation result, the membership value of the node attribute at the level m_k is the sum of the membership values of the sampled samples, namely:

$$\mathrm{F}\left(\mathrm{G}\left(C_{m_i}(\mathrm{x})\right)\right) = \sum_{j=1}^{j=p} c_{m_k}^j(x)$$

From this, the information entropy of the achievement node at the level m is as follows:

$$\mathrm{FH(G)} = -\sum_{m=1}^{m=k} \frac{\mathrm{F}\left(\mathrm{G}\left(C_{m_k}(\mathrm{x})\right)\right)}{\sum_{m=1}^{m=k} F\left(\mathrm{G}\left(C_{m_k}(\mathrm{x})\right)\right)} log_2 \frac{\mathrm{F}\left(\mathrm{G}\left(C_{m_k}(\mathrm{x})\right)\right)}{\sum_{m=1}^{m=k} F\left(\mathrm{G}\left(C_{m_k}(\mathrm{x})\right)\right)}$$

Fuzzy segmentation of attribute node G and attribute node A_i to obtain node G at node A_i fuzzy conditional entropy such as formula:

$$\mathrm{FH}\left(\frac{\mathrm{G}}{A_i}\right) = \sum_{m=1}^{m=k} \frac{F(A_i(c_m(x)) \cap F(c_m(x)))}{\sum_{m=1}^{m=k} F(A_i(c_m(x))} \mathrm{FH}(\mathrm{G} \cap A_i)$$

Finally, the corresponding information gain of the node A_i at the node G is obtained such as the formula:

$$\mathrm{FGain}(A_i, \mathrm{G}) = \mathrm{FH(G)} - \mathrm{FH}\left(\frac{\mathrm{G}}{A_i}\right)$$

Through the obtained information gain value, $\mathrm{FGain}(A_i, \mathrm{G})$ is selected as the root node of the decision tree, and then each subtree is recursively called to gradually locate the branch nodes of the tree. Finally, the results are predicted by the fuzzy decision tree.

3. Case Analysis

3.1 Example Modeling

Randomly selected 48 students from Shanxi University of Finance and Economics, through data cleaning, screening and conversion, select students' semester classroom attendance rate, self-study duration (average daily), library borrowing (semester total), dormitory learning atmosphere as decision tree node the attribute, the student's final grade is the decision tree node attribute.

The values of the center points of the model attributes are selected by the collected data results, as shown in Table 1, where a1 and a3 respectively distinguish the intermediate points of the attribute levels, and m1, m2, and m3 are attribute levels.

Table1 Attribute center point and horizontal value selection

Attributes	a1	a3	m1	m2	m3
Class attendance rate	80%	60%	High	Medium	Low
Self-study time	3	2	Long	Medium	Short
Library borrowing	8	4	More	Medium	Less
Final grade	80	60	Excellent	Average	Poor

The original data instance sample is fuzzified according to the triangle membership function designed in section 2.2, and the fuzzy membership matrix of the student achievement and each evaluation attribute is obtained, as shown in Table 2:

Table2 Fuzzy membership matrix

Class attendance rate			Self-study time			Library borrowing			Final grade		
High	Medium	Low	Long	Medium	Short	More	Medium	Less	Excellent	Average	Poor
1	0	0	1	0	0	1	0	0	1	0	0
1	0	0	0.4	0.6	0	0	0	1	0	0.1	0.9
1	0	0	1	0	0	0	0	1	1	0	0
….	…	…	…	…	…	…	…	…	…	…	…
0	0	1	0	0	1	0	0.5	0.5	0	0	1

The fuzzy decision tree model is calculated by calculation as shown in the Figure3.

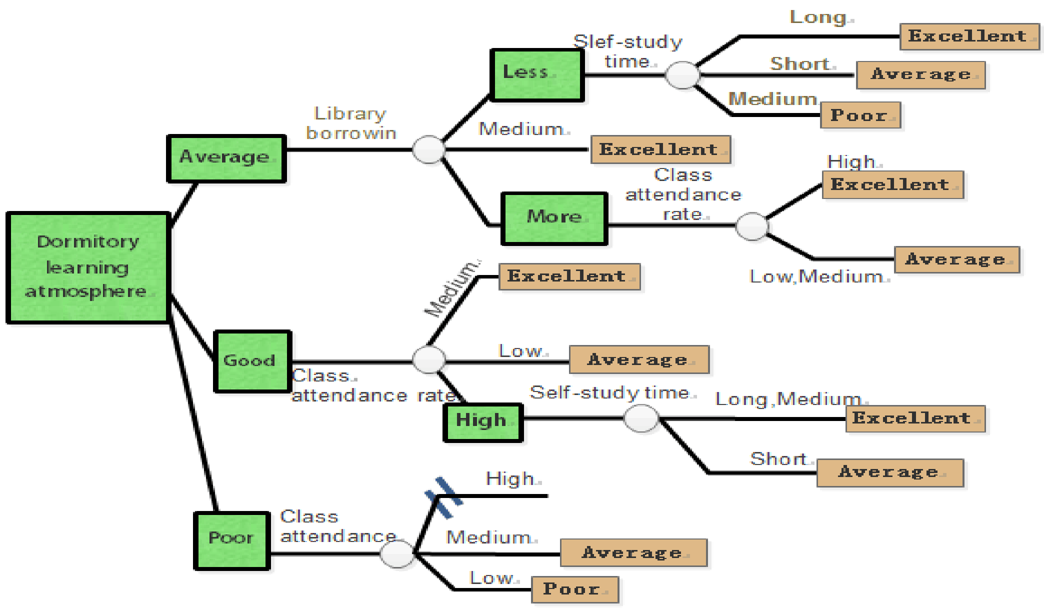

Figure3 The fuzzy decision tree

4. Comparison and analysis of results

1) Based on the daily learning behavior of college students, this paper uses fuzzy theory to design membership function, and combines ID3 decision tree algorithm to deeply explore the relationship between students' daily behavior and final grades, and establishes a fuzzy decision tree. The experiment proves that the fuzzy decision tree can analyze and predict students' performance correctly, efficiently and comprehensively, and provides an important basis for the information construction and teaching management decision-making work of colleges and universities.

2) Experimental results show that the fuzzy decision tree is better than the decision tree ID3 algorithm in terms of test accuracy. Therefore, the fuzzy decision tree algorithm has stronger classification ability. In the natural and social phenomena, the difference of objective things often goes through a form of intermediary transition. The rules generated by the decision tree ID3 algorithm are clear, ignoring the uncertainty of classification, and the fuzzy decision tree fully considers the uncertainty of classification, so the fuzzy decision tree has stronger robustness. The rules generated by the fuzzy decision tree are marked with a certain degree of confidence, which is consistent with the facts.

References

[1] Baoling Qi. Application Research of Decision Tree Technology in Teaching Quality Evaluation

[J]. Artificial Intelligence & Identification Technology, 2007(7):191-192.

[2] Chu He, Jian Song, Tong Zhuo. Course relevance classification model based on frequent pattern spectral clustering and student achievement prediction algorithm [J]. Application Research of Computers, 2015(32):2930-2933.

[3] Shili Xuan. Application of Decision Trees in English Proficiency Test Results Analysis[J]. Computer &Digital Engineering, 2016,42(5): 844-846.

[4] Wei Zheng, Nan Ma. An Improved Post-Pruning Algorithm for Decision Tree[J]. Computer &Digital Engineering, 2016(26):114-118.

[5] Mehmed Kantardzic. Data Mining - Concepts, Models, Methods and Algorithms [M]. Beijing: Tsinghua University Press, 2003.

[6] Liya Fu. The Application of Data Mining's C4. 5 Algorithm in the Students' Result Analysis[J]. Journal of Chongqing University of Technology (Natural Science), 2013(27):78-82.

[7] Min Liu, Gencai Chen. Research on Data Mining Technology in Decision Support System[J]. Application Research of Computers, 1999, 16(11):83.