Intelligent Placement Model Based On Decision Tree

CongYu Cai
College of Information Engineering
China Jiliang University
Hangzhou (310018) Zhejiang, China
ccyberyl@163.com

Huijuan Lu
College of Information Engineering
China Jiliang University
Hangzhou (310018) Zhejiang, China
hjlu@cjlu.edu.cn

Ke Yan
College of Information Engineering
China Jiliang University
Hangzhou (310018) Zhejiang, China
yanke@cjlu.edu.cn

Minchao Ye
College of Information Engineering
China Jiliang University
Hangzhou (310018) Zhejiang, China
yeminchao@cjlu.edu.cn

Abstract- Recently, the analysis of big data based on artificial intelligence algorithms has shown great potential in the field of education system. However, due to the complexity of the education system, the problem of students' placement in primary and secondary schools (matching problem of students and teachers) still exists. This paper proposes a placement model using ID3 and C4.5 bagging algorithms. The proposed model uses the characteristics and academic performances of students to train the decision tree and performs appropriate pruning to implement the decision trees. In the experimental results, the proposed method (ID3 algorithms and C4.5 algorithm) achieves high matching accuracy.

Keywords-Decision tree; placement model; C4.5 algorithm; ID3 algorithm; bagging;

I. INTRODUCTION

Many studies have been conducted using artificial intelligence algorithm to promote the next generation education system[2]. Due to the complexity of the education system, there are still many problems in the current education systems. Among them, the placement problem of primary and secondary school students (student and teacher matching problem) has not been well resolved.

Education should pay attention to teaching students in accordance with their aptitude. Different teachers have different teaching methods. The problem of pairing between teachers and students is difficult to solve because of the variety of teaching styles and the different learning habits of students [1].

At present, the mainstream placement scheme has the following three situations [3]:

A. Randomness

The same number of students are randomly assigned according to the computer system, that is, the pairing between teachers and students is randomly arranged. Although this method is convenient, fast and does guarantee the completion of the matching work, it does not satisfy the public.

B. Ranking

Arrange for placement exams and, with the most standardized scores, assign teachers to the class according to the student's current student performance. This approach is relatively ignoring the different personality traits of students and their future development, and may even affect students' self-confidence [5].

C. Balance

Arrange the placement examinations and assign them randomly according to the system if the average scores of the examinations for each class are the same or the number of students in the grades are the same. This approach is an improvement to the first method, so that the level of each class is as balanced as possible. However, this approach has not yet reached our optimal match.

II. RELATED WORK

In order to solve this problem, some algorithms are tried, such as clsutering. Students with similar characteristics are expected to be grouped into one class and then paired with the teachers whose teaching style are suitable to those students. The effect of this idea is not ideal.

- **General Satisfaction Matching Model:** For the general matching model, the characteristics of students and teachers need to be listed and scored with each other. Then, the data are normalized and the matrix of multiplication is used to obtain the satisfaction matrix. In theory, this method seems to be very good, however in real life, teachers and students should score each other on the premise that the teacher and the student have certain contact. This approach has no practical operability, so this is not a suitable method.
- **K-means clustering:** As to the matching problem in this paper, categorical variables (category variables are generalizations of binary variables) are needed, because each value has no numerical or ordinal meaning, such as psychological, personality, and so on. However, the effect of clustering classification is too rough, and it is not desirable. Because the character of the person is responsible and the data dimension of the habit is high, only use of the K-means clustering alone does not a reasonable measure.

2474-3828/18/$31.00 ©2018 IEEE
DOI 10.1109/ITME.2018.00188

In this paper, the classification function of the decision tree model is used to prove that the model can be well used in the placement model. In turn, the problem of placement in primary and secondary schools can be solved.

III. BASIC KNOWLEDGE

Decision tree is another algorithm that divides input space into different regions, and each region has independent parameters. Decision tree classification algorithm is a case-based inductive learning method, which can extract the tree-like classification model from a given disordered training sample. Each non-leaf node in the tree records which feature is used to determine the category, and each leaf node represents the final category.

1) Entropy [4]:

Let X be a discrete random variable with a finite n values, the probability distribution is

$$P(X = x_i) = p_i, i = 1, 2, 3, \ldots, n \tag{1}$$

By definition, entropy only depends on the distribution of X. Therefore, the entropy of X can also be written as $H(p)$, i.e.

$$H(p) = -\sum_{i=1}^{n} p_i \log p_i \tag{2}$$

With a random variable (X,Y), its joint probability distribution is

$$P\left(X = x_i, Y = y_j\right) = p_{ij}, i = 1, 2, \ldots, n; j = 1, 2, \ldots, m \tag{3}$$

Conditional Entropy represents the uncertainty of the random variable Y under the condition that the random variable X is known.

The conditional entropy of the random variable Y under the given condition of the random variable X,

$$P\left(Y \mid X\right) = \sum_{i=1}^{n} p_i H\left(Y \mid X = x_i\right) \tag{4}$$

Among them $p_i = P\left(X = x_i\right), i = 1, 2, \ldots, n$

2) Information gain

Defining: The difference of entropy before and after dividing data sets with a certain feature ;

$$g(D, A) = H\left(D\right) - H\left(D \mid A\right) \tag{5}$$

3) Information gain ratio:

Taking the information gain as the feature of dividing the training data set, there is a problem that the feature with a larger value is selected. This problem can be corrected using the information gain ratio. This is another criterion for feature selection [7].

Defining $g_R(D, A)$ -the information gain ratio of the feature A to the training data set D is defined as the ratio of its information gain entropy $H_A(D)$ of the training data set D with respect to the value of $g(D, A)$ the feature A, i.e.

$$g_R\left(D, A\right) = \frac{g\left(D, A\right)}{H_A(D)} \tag{6}$$

IV. DECISION TREE BASED ON PLACEMENT MODEL

The decision tree model is a tree structure, which is a basic classification and regression idea. It can be considered as a conditional probability distribution defined in feature space and class space. When learning, using the training data, the decision tree model is built on the principle that the loss function is minimized. When forecasting, the new data is classified using a decision tree model.

In the model of the placement problem, the scoring standard is to express the degree of likeness according to the score. For example, the degree of judgment within the personality is 0-10 points, that is, 0 is extremely introverted, and 10 is extremely extroverted [6].

A. Hypothesis

1) In this model, the factors affecting the pairing between teachers and students only consider the four factors of influence: the students' psychological age, class participation, teacher interaction, and independent thinking ability. (In actual application, users can increase or decrease factors according to actual needs. Factors do not affect the use of those models.)

2) Do not consider very few people with different personalities to appreciate each other.

B. Thinking on model

1) Decision Tree model

Conditional Entropy represents the uncertainty of the random variable Y under the condition that the random variable X is known.

The conditional entropy of the random variable Y under the given condition of the random variable X,

$$P\left(Y \mid X\right) = \sum_{i=1}^{n} p_i H\left(Y \mid X = x_i\right) \tag{7}$$

Among them $p_i = P\left(X = x_i\right), i = 1, 2, \ldots, n$

Defining (information gain or information gain rate) is

$$g\left(D, A\right) = -\sum_{K=1}^{K} \frac{\left|C_k\right|}{\left|D\right|} \log_2 \frac{\left|C_k\right|}{\left|D\right|} + \sum_{i=1}^{n} \frac{D_i}{D} \sum_{K=1}^{K} \frac{\left|D_{ik}\right|}{\left|D\right|} \log_2 \frac{\left|D_{ik}\right|}{D_i} \tag{8}$$

$$g_R\left(D, A\right) = \frac{g\left(D, A\right)}{H_A(D)} \tag{9}$$

Or

Let the training data set be D, $|D|$ for its sample size

There are K classes C_k ($k = 1, 2, \ldots, K$). $|C_k|$ for the number of samples belongs to class C_k,

$$\sum_{k=1}^{k} \left|C_k\right| = \left|D\right| \tag{10}$$

Let feature A have n different values $\{a_1, a_2, \ldots, a_n\}$, according to the value of feature A, D is divided into n subsets $D_1, D_2, \ldots, D_n$. $|D_i|$ is the number of samples D_i.

$$\sum_{i=1}^{n} |D_i| = |D| \tag{11}$$

The set of samples belonging to the class D_i in the subset C_k isC_k ($D_{ik} = D_i \cap C_k$). $|D_i|$ is the number of samples of D_i.

Let the number of leaf nodes of the tree T be $|T|$. T be the leaf node of the tree T and the leaf node has N_t sample points, wherein the sample points of the k class have $k = 1, 2, \ldots, K$. $H_t(T)$ is the leaf node t empirical entropy, $\partial \geq 0$ for the parameter, then

At this time

$$C_\partial(T) = C(T) + \partial |T| \tag{12}$$

In equation (12), $C(T)$ indicates the prediction error of the model on the training data, that is, the degree of fitting of the model to the training data.

2) Based on decision tree's placement model

(1) Assign the new eigen data to the decision tree (i=1, 2, 3... m)

(2) Decision tree produces predictive judgment

(3) Cycle n times step1 and step2, and get n prediction results

(4) Take most of the same predictions

(5) If the result with the decision tree is excellent academic performance, successful matching

(6) If the result of the decision tree is academic or generally unsatisfactory, i=i+1, repeat Step1 to Step5

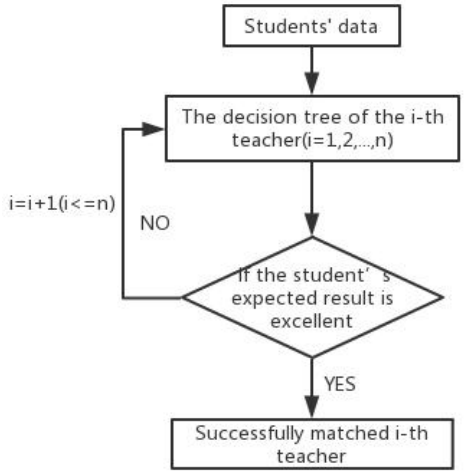

Figure 1. The flowchart of the model

C. Algorithm description

The algorithm of decision tree learning is usually a recursive selection of the optimal features and the training data is segmented according to the features, so that there is a best classification process for each sub-data set [6]. The decision tree learning algorithm includes feature selection, decision tree generation and decision tree pruning process. The commonly used algorithms for decision tree learning are ID3 and C4.5.

(1) ID3 Algorithm

The core of the ID3 algorithm is to apply the information gain criterion selection feature on each node of the decision tree to construct the decision tree recursively [8].

Input: training data set D, feature set A, threshold ε;
Output: Decision Tree T and Trimmed Subtree T_a.

(1) If all instances in D belong to the same class, then T is a single-node tree, and the class C_k is marked as a class of the node, returning T.

(2) If $A = \varnothing$, then T is a single-node tree, and the class with the largest number of instances in D. C_k is marked as the class of the node, and returns T.

(3) Otherwise, calculate the information gain of each feature pair D in A according to the algorithm, and select the feature with the largest information gain A_g.

(4) If the information gain of A_g is less than the threshold C_k, then T is a single node tree, and the class with the largest number of instances in D is used as the class C_k tag of the node, and returns T.

(5) Otherwise, a_i for each possible value of A_g, divide D into a number of non-empty subsets, and use the class with the largest number of instances as a marker to construct a child node, which consists of a node and its child nodes T. Program return T.

(6) For the i-th child node, use D_i as the training set and $A - \{A_g\}$ as the feature set, recursively call step (1) ~ step (5), get the Subtree T_i. Program return T_i.

(7) Calculate the empirical entropy of each node.

(8) Recursively retracting from the leaf nodes of the tree.

(2) C4.5 Algorithm

Input: training data set D, feature set A, threshold ε;
Output: Decision Tree T and Trimmed Subtree T_a.

(1) If all instances in D belong to the same class C_k, then T is a single-node tree, and C_k as the class of the node. Program return T.

(2) If $A = \varnothing$, set T to a single-node tree, and class the class with the largest number of instances in D , C_k as the class of the node. Program return T.

(3) Otherwise, the information gain ratio of each feature pair D in A is calculated according to equation (11), and the feature with the largest information gain ratio A_g is selected.

(4) If the information gain ratio of A_g is less than the threshold ε, T is a single node tree and the class with the largest number of instances in D, C_k as the class of the node. Program returns T.

(5) Otherwise, each possible value of A_g is paired with a_i divided by D into a subset of non-empty $A_g = a$, the class with the largest number of instances in D_i as a marker, constructing a child node, by the node and its child nodes Points form a tree T and return T.

(6) For node i, use D_i as the training set and $A-\{A_g\}$ as the feature set, recursively call step (1) ~ step (5). Program get Subtree T_i, return T_i.

(7) Calculate the empirical entropy of each node.

(8) Recursively retracting from the leaf nodes of the tree.

(3)Placement Algorithm

Input: data D_k set D trees T

```
For(, k<=n,k=k+1){
For(let the same group data input T_ij ， j=j+1){
    For(i=1;let the same group data input T_1n to judge the result, i=i+i){
        Call ID3 algorithm or C4.5 algorithm
    }
        Bagging
        If(expected academic performance is excellent){
            Break，end loop，pairing successfully
        }
    }
}
```

V. EXPERIMENTAL RESULT

Through data collection in all directions, different styles of teachers, A, B, and C were selected, and their typical 40 sets of data—the degree of participation, psychological age, personality (inside and outward), and independent thinking, degree of interaction with the teacher of 20 students. The values are assigned from 1-10 depending on the situation, and the data is aggregated into tables to become available data. The 40 sets of data were divided into two: one as training data and the other as test data.

A. K-means algorithm

The K-means algorithm was used to analyze the data of these 60 students. The approximate results are shown in the figure 2 .

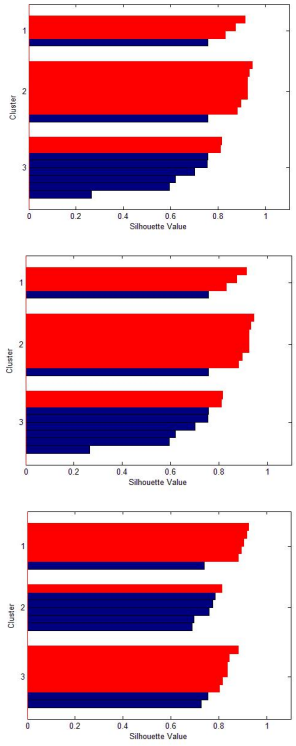

Figure 2. The outline of tested K-means algorithm

As can be seen from the contour map (figure 2), there are 18 groups of arrays with values of 30% below 0.8, which is an intolerable ratio [8].

It can also be seen from the scatter plots of the results that some data centers are relatively close in distance and cannot be well distinguished.

B. ID3 algorithm and C4.5 algorithm

After training 60 sets of data, the decision tree generated three different classification decision trees. After pruning, the correct rate of 60 sets of test data reached 100%.

Through the data, we can also analyze the teaching styles of different teachers. For example, the academic performance of B teachers is determined by the degree of independent thinking and the degree of participation in the class. From the data, we can analyze that B teacher pays more attention to students' self-learning ability and classroom. The efficiency of the lectures.

C. Data Increment

In order to prevent the lack of judgment caused by the lack of data, data is collected form the major primary and secondary schools in the Internet and obtained 3 different styles of teachers and their students' participation in the classroom, psychological age, personality (inside and outside), and like to think independently, data on the extent of the interaction with the teacher.

Plus, the concept of random forest is also considered in the experiment.

5875 data have been quantified with the same standard. The correctness is obtained for 2000, 4000, and 5875, as shown in the table I.

TABLE I. ACCURACY RATE TABLE

Data size	Three types of algorithm			
	K-means algorithm	*ID3 algorithm*	*C4.5 algorithm*	*Bagging of ID3 and C4.5*
2000	***88.65%***	***96.98%***	***97.33%***	***99.79%***
4000	***90.02%***	***98.00%***	***98.02%***	***100.00%***
5875	***86.67%***	***98.67%***	***99.23%***	***100.00%***

era. Years of 2000 to 2017

VI. DISCUSSION AND CONCLUSION

In this paper, decision trees-based placement model is proposed to solve the problem of placement in primary and secondary schools. Firstly, through data analysis, the characteristics of each teacher's students can be obtained, such as the psychological age, the degree of participation in the class, and the academic performance of the corresponding students. After data is quantized, the data is trained to appropriate pruning for decision tree. Secondly, for the new placement, only the new students' characteristics are acquired. The new data will be cycled once in the teachers' decision tree. If the predicted academic achievement is excellent, the assignment will be completed.

In conclusion, the random forest generated by ID3 algorithm and C4.5 algorithm bagging has excellent performance and is suitable to be used in placement of the primary and secondary school. Plus, this model can also be used in n to n matching problem.

ACKNOWLEDGMENT

This work is supported by National Natural Science Foundation of China (Nos. 61272315, 61402417, 61602431, 61701468 and 61850410531), International Cooperation Project of Zhejiang Provincial Science and Technology Department (No. 2017C34003).

REFERENCES

[1] Wu Y. H. , B. W. Liu, and X. L. Ma, "Constructing the Ecosystem of "AI + education," Proc. JOURNAL OF DISTANCE EDUCATION (G40-057), Press, May . 2017, pp. 68-73, doi: 10.15881/j.cnki.cn33-1304/g4.2017.05.003.

[2] Z. X. Cai, "40 years of Artificial Intelligence in China," Proc. SCIENCE AND TECHNOLOGY REVIEW (G40-34), Press, May . 2016, pp. 12-32, doi: 10.3981/j.issn.1000-7857.2016.15.001.

[3] D. M. Wang, C. H. Lu, W. W. Jiang, M. X. Xiao and B. R. Li, "Research on Decision Tree SVM Multi-classification Method based on Particle Swarm Optimization," Proc. Journal of Electronic Measurement and Instrumentation (TP181), Press, Feb .2015, pp. 611-615, doi: 10.13382/j.jemi.2015.04.018.

[4] L. Sun, Y. X. Cheng, "The Research and Realization of Learning Achievement Prediction in Network Education in the Age of Big Data: T. Zhang, J. Cao, "Decision Tree Algorithm For Big Data Analysis" Proc. Computer Science (TP181), Press, Jan . 2016, pp. 173-179.

[5] J. P. Zhang, "Thinking about Artificial Intelligence Education" Proc. E-education Research (G420), Press, Jan . 2003, pp. 27-39, doi: 10.13811/j.cnki.eer.2003.01.005.

[6] Q. Zhong, S. X. Zhong, "Integrated Application of Artificial Intelligence in Education" Proc. Journal of Gannan Normal University (G434), Press, Jun . 2006, pp. 66-69, doi: 10.13698/j.cnki.cn36-1037/c.2011.06.025.

[7] R. Y. Li, L. Cheng, "Construction of Decision Tree based on SVM optimal decision surface" Proc. Journal of Electronic Measurement and Instrumentation (TP181), Press, Mar . 2016, pp. 342-351, doi: 10.13382/j.jemi.2016.03.003

[8] J. Shuai, L. P. Li, Y. Q. Chen, "The role of Decision Tree Model and Logistic Regression Model in Influencing Factors of Injury Occurrence" Proc. Chinese Journal of Disease Control & Prevention (R195), Press, Feb . 2015, pp. 185-189, doi: 10.16462/j.cnki.zhjbkz.2015.02.021.