EVALUATION METHOD OF STUDENTS BASED ON DECISION TREE MODEL

Xin Wang¹, Weihao Wang²

¹,²Science College, Beijing Information Science & Technology University, Beijing, China
Bitimath[at]163.com

Abstract—The evaluation mechanism of students in the primary school stage is one of the important means to help students to develop good learning and living habits. At this stage, the teacher summarizes the semester comprehensive quality of every student through the reasonable evaluation mode. In this paper, scientific and effective data mining way is used to build an evaluation model. We use the decision tree algorithm which is commonly used in data mining field to evaluate the comprehensive quality of students of the fifth and sixth grade in primary school. The attributes, such as the discipline, health, habits and grades are considered in our decision tree model.

Keywords—Student evaluation mechanism; Decision tree algorithm; Classification algorithm

I. INTRODUCTION

Decision tree is a commonly used method of classification in data mining and originated in concept learning system. The most influential decision tree method in the world is ID3 algorithm proposed by Quinlan[1]. The ID3 algorithm has some advantages such as clear theory, simple calculation and stronger learning ability, but it is lack of effectiveness for relatively large-scale data set and is sensitive to noise data. In addition, ID3 algorithm prefers to the attributes with more options in training dataset, which is another shortcoming. In 1992 Quinlan proposed an improved ID3 algorithm -C4.5 algorithm[2] that can not only overcome the shortcomings of the ID3 but also realize the discretization of continuous attributes and the processing of incomplete data.

The teacher in charge is an important tutor in the primary school, almost from the study to life. In order to promote the development of students' behavior habits and the progress of their academic achievements, the teacher's evaluation mechanism is particularly important.

At this stage the evaluation mechanism in China is mainly by subjective assessment of teacher and inputting into students file manually. The evaluation system modeling in scientific classification method, such as data mining algorithms, has not been used widely. In this paper, quantification evaluation method is researched and carried out. Discipline, health, habits, achievements, moral, activity, these attributes in students daily evaluation form provide natural splitting nodes for decision tree algorithm, and the classification of comprehensive performance for students can be shown from the final leaf nodes.

A. Decision Tree Algorithm

The decision tree algorithm is a kind of tree structure that describes the classification of instances in training dataset[3]. Decision tree consists of nodes and directed edges. There are two types of nodes: internal nodes and leaf nodes. An internal node represents a feature or attribute and a leaf node represents a classification. The decision tree is used to classify instance by the attributes of the instance. Starting from the root node, some instances are assigned to its child nodes according to the test result. Every child node is corresponding to an attribute value. So all instances are recursively assigned to the nodes till the leaf nodes. As shown in Figure 1, circles and squares represent the internal nodes and leaf nodes, respectively.
The decision tree can be regarded as a set of if-then rules. Each path from the root node of the decision tree to each leaf node builds a rule. The characteristics of the internal node on the path correspond to the conditions of the rules and the classification of the leaf node corresponds to the conclusion of the rules. The if-then rules set have an important property: mutually exclusion and perfect. That is to say, each instance is covered by the unique path or rule.

**B. Entropy gain**

Define the training dataset $D = \{ (x_1, y_1), ..., (x_N, y_N) \}$, where $x_i = (x_i(0), x_i(1), ..., x_i(n))$ is input instance and $y_i = (1, 2, ..., K)$ is the classification marker of $x_i$, $i = 1, 2, ..., N$. $n$ is the number of attributes and $N$ is sample size. The learning goal is to build a decision tree model for a given training dataset, so that it can correctly classify test data.

From the information theory, we know that the smaller the expected information, the greater the entropy gain, and the higher the accuracy of the model results. Therefore, the core idea of the decision tree algorithm is to select the attributes of the maximum entropy gain. Here are a few definitions to be used.

Let $D$ be the training dataset (or sub-dataset), and the entropy of $D$ can be described as:

$$H(D) = -\sum_{i=1}^{C} p_i \log_2 p_i$$

where $p_i$ is the appearance probability of the classification $i$ in the training dataset. According to maximum likelihood estimation method, $p_i$ can be estimated as the number of instances in the classification $i$ divided by the total number of instances in whole training dataset. The actual meaning of entropy is the average amount of information needed for the classification marker of instances in $D$.

Now we assume that the training dataset $D$ is divided by the attribute $A$ and $|D|$ is sample capacity. There are $i$ instance classes, $i = 1, 2, ..., I$. $|C|$ is the number of instances belongs to the classification $C_i$ and $\sum_i C_i = D$. $\{a_1, a_2, ..., a_r\}$ is the values set of attribute $A$ and $D$ is divided into $n$ subsets $D_1$, $D_2$, ..., $D_n$ by the $n$ attribute values. $D_n = D_1 \cap C_i$ and the conditional entropy of $D$ is as below:

$$H(D|A) = -\sum_{i=1}^{C} \left[ \frac{|D_i|}{|D|} \right] \log_2 \left[ \frac{|D_i|}{|D|} \right]$$

The entropy gain is the difference of them:

$$H(D) - H(D|A) = -\sum_{i=1}^{C} \left[ \frac{|D_i|}{|D|} \right] \log_2 \left[ \frac{|D_i|}{|D|} \right]$$
Decision tree algorithm is that the attribute with the largest entropy gain is selected to split. Constructing a decision tree needs a recursive process, so it is necessary to determine the stopping condition. One of the most intuitive ways is to stop program when each child node has only one type of record. However, it is easy to get too much nodes and result in over fitting problem. So another feasible method is that the classification corresponding to \( \max(P(i)) \) is taken as the current leaf node when the number of records in current node is lower than a threshold, then the split is stopped.

II. MODEL CONSTRUCTION

In this paper the training data is from the classroom performance of 136 students of Xidan Primary School in Beijing city, which ensures the rationality and representatives of the data.

As shown in Table 1, there are 6 evaluation features in training data set, namely \( A_1 \) "discipline", \( A_2 \) "hygiene", \( A_3 \) "habit", \( A_4 \) "score", \( A_5 \) "quality", \( A_6 \) "vitality". \( \{1,2,3\} \) is the class marker corresponding to different levels, such as "good", "medium", "bad"", respectively. The part correlated training data is as table 1 shown:

The empirical entropy of the training dataset can be obtained as below:

\[
H(D) = \frac{80}{136} \log_2 \frac{80}{136} + \frac{23}{136} \log_2 \frac{23}{136} + \frac{33}{136} \log_2 \frac{33}{136}
\]

Computing the entropy gains of each attribute,

\[
g(D,A_i) = H(D) - \left[ \frac{82}{136} H(D|A_i = 1) - \frac{30}{136} H(D|A_i = 2) - \frac{24}{136} H(D|A_i = 3) \right]
\]

where \( H(D|A_i = 1) \), \( H(D|A_i = 2) \), \( H(D|A_i = 3) \) are the empirical conditional entropy training of the training data set corresponding to the three attribute values of \( A_i \), respectively. Similarly, the entropy gain \( g(D,A_i), \ldots \),

<table>
<thead>
<tr>
<th>data</th>
<th>discipline</th>
<th>hygiene</th>
<th>habit</th>
<th>score</th>
<th>quality</th>
<th>vitality</th>
<th>Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_1 )</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>( x_2 )</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( x_3 )</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>( x_4 )</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>( x_5 )</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( x_6 )</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( x_7 )</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( x_8 )</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( x_9 )</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( x_{10} )</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>( x_{11} )</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>( x_{12} )</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( x_{13} )</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>
\[ g(D, A_k) \]
can be calculated. Then we select the maximum entropy gain as the root node, which divides the training data set into two subsets \( D_1 \) and \( D_2 \). Using this analogy, the entropy gain of \( D_1 \) and \( D_2 \) based on different attributes can be calculated according to the above algorithms and the attribute corresponding maximum entropy gain is chosen as splitting attributes. On the basis of what attribute is classified, we draw branch according to the best classification decision attribute, then a decision tree is constructed as shown in the figure 2.

\[ \text{Figure 2. The decision tree model} \]

A, B, C of the leaf nodes denote students with high quality, students with average quality and students with low quality, respectively. For instance, \( n = 72 \) of node 4 is the number of students with attributes \( A_1 \leq 2 \), \( A_2 \leq 1 \) and \( A_3 \leq 1 \), etc. The probability that these students with the attribute values will be classified as class label “A” is 0.97. There are most of the students for the C class in node 9, which shows that these students do not receive a good revaluation of comprehensive quality from their teacher.

When the discipline is good and the habit is not good the students received inconsistent evaluation from their teacher, and the part reason is that the amount of data is too small. It can be seen that the students’ habits are slightly different in the eyes of different teachers. It can be seen that the students' habits are slightly different from the perspective of different teachers. On the other hand, when the students is inclined to be classified as class label A by teacher when they do not have high attribute value of “activity”, but with superior attribute value of “score”.

The test data set is used to test the decision tree model, and the results are as follows:

```
testPred A B C
A 19 2 0
B 5 10 4
C 0 2 3
```

\[ \text{Figure 3. The testing result} \]
It can be seen from Figure 3, the samples of the class A and the class B has better testing result though there is little overlap. However, for class C, four of seven students are misclassified as class B. The main reason is less data, and the model precision needs to be further improved.

III. CONCLUSION

In this paper, we use one of data mining methods- the decision tree algorithm to evaluate the comprehensive quality of the primary school students. The algorithm is easy to implement and the results are obvious. When the amount of data is large, the model can further improve the prediction accuracy.

ACKNOWLEDGEMENT

This research was supported by the National Natural Science Foundation of China (71501016).

REFERENCES