ISSN : 0974 - 7435 Volume 10 Issue 19

2014

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 10(19), 2014 [11776-11780]

Application research of analysis and decision of students mark based on the big data

Yu Hua

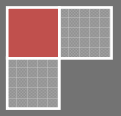
Business College of Shanxi University, Taiyuan, 030031, (CHINA)

ABSTRACT

With the increase enrollments of college and high rates of participation, the traditional approaches of analysis and decision of student's mark cannot gain a good effect which only divides marks into excellent, fine, ordinary and bad with statistic of each rank, failing to the deep analysis, even cannot find the problems in teaching. So it is bad for the revolution of education and the promotion of teaching quality. This research focuses on the application of analysis and decision of student's mark based on the big data, seeking the factors influencing students' marks and provides theoretical foundation for later teaching work.

KEYWORDS

Big data; Student's mark; Analysis and decision; Application.



INTRODUCTION

At present, many universities are increasing the enrollment scale year by year, which leads many problems to teaching. Combining the reality, this thesis and the author study the application of analysis and decision of student's mark based on the big data and seek significant information beside student's marks by the technology of decision tree and make a decision to improve the quality of teaching. The decision tree works with data prediction and classified technology. Through a group of examples without rule or order, the classified rule of decision tree's results are worked out. The study is shown in follows.

THE STUDY OF DECISION TREE

The summarize of sorting algorithms

In data mining, classification is a significant part which can be implied in the forecast and decision, in which sorting algorithms is important one mainly of the decision tree algorithm, K neighbor sorting algorithm, genetic algorithm, association rules algorithm and Bayesian classification algorithm. Analyzing the input data is the gain of classification. Through the characteristics of centralized training, the exact model or description of each kind will be found. And according to these, the description of kind will be worked out and be applied in the tests in the future. Though we don't know the label of kind in the later test, we still can predict these new data's classification based on these and have a better understand about each kind in statistics, simply speaking, we gain the knowledge of this kind.

Sorting algorithms based on decision tree

During learning symbol, inductive learning is the most extensive one which is rule contained in the example without rule or order. The decision tree algorithm is an inductive learning based on the practice, generally used to form a classifier and predication model and can classify, predict, data pre-processing, data mining the unknown data.

The description of decision tree

Being similar to flow chart, the decision tree's internal nodes are the property or the gathering of property as shown in Figure 1. The property of internal node is called test property which refers to one property of test. And the kind we need to classify is leaf node and each output of test will be represented by one branch. Root node is the top one of the tree. After a training of practical gather, a decision tree will be created. And the decision tree can classify an unknown practical gather and completed by the value of property. During classifying a practical decision tree, the test of value of property is started from the tree root and along the branches and reaches the leaf node. So this kind is one represented by the leaf node.

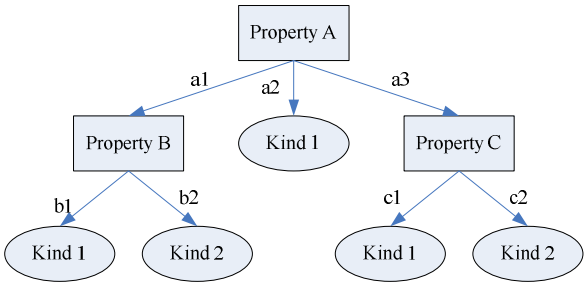


Figure 1 : The structure of decision tree

The type of decision tree

Decision tree can be divided into regression tree and classification tree. The first one is used to continuous variable and the latter is continuous variable. According to the property, decision tree can be divided into the following ones:

(1) The test property of internal node maybe univariate which means that each internal node only has one property inside. However, the internal node of decision tree may have multivariable test property which means that some nodes have exceptional property.

(2) According to the difference of test property, each node may have two or exceptional property. And the internal node has two branches which is called binary decision tree.

(3) Each property maybe the value type or enumeration type.

(4) The results may have two or more kinds. If the result is two-kind binary decision tree, we call it Boole decision tree, which is easily expressed by the disjunctive normal form, the most natural condition of decision tree.

The process of decision tree

The disjunctive normal form of decision tree is top-down and the comparison of property value is proceeding in the internal node of decision tree. And then according to the different property value, the branches will be judged from this node down and the result will be gained at the leaf node of decision tree. So between root and leaf node, the route is corresponding

to the disjunctive rule and the whole decision tree is corresponding to one group of disjunctive expression. Generating algorithm of decision tree involves two steps: the first is the generation of tree whose root node is the place where statistics exist at beginning, and then the recursive data fragment starts. The second is the pruning of the tree, which is eliminating some probably unusual and noisy statistics. The condition that the decision tree stop split is: the statistics of one node are belonging to one kind; no-property can used to split the statistics. The process of generation of tree is shown in Figure 2:

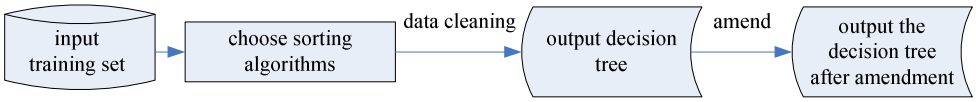


Figure 2 : The process of generation of tree

The mathematical model of construction algorithm of decision tree

Training set T can complete construction algorithm of decision tree, T={<x,cj>), while as a training practice, x=(a1, a2, …, an) has n properties, which is respectively included in the attribute table (A_1,A_2,…,A_n) and in which a1 is the value of A1. C=(C_1,C_2,…,C_n)(Cj∈C)is the classified result of X. The algorithm can be fall in to following steps:

(1) Choose A1 as the classified property from the attribute table;

(2) Property Ai has kj values, and T is classified into kj subsets T_1,T_2,…,T_{KJ}, in which, T_{ij}={<x,c>|<x,c>)∈T, and the value of X is one of k_j;

(3) Delete Ai from the attribute table;

(4) As for each T_{ij}(1≤j≤K1), T=T_{ij} ;

(5) If the attribute table is not null, go back to (1) or output.

Decision tree algorithm

The algorithms such as ID3,C4.5,CART,CHAID are the main generating algorithm in decision tree. In 1979, J. Ross Quinlan created ID3 algorithm, which is popular in the computer learning and represents for one big kind in the inductive learning. In ID3 algorithm, choosing property vigorously improved the efficiency and quality of classification so it is used widely and now is the commonest decision tree algorithm. Based on the information gain, ID3 algorithm is a typical decision tree algorithm from top to bottom featuring at one node the maximal information gain is heuristic method to decide which property and then generate the tree. Applying this algorithm can get the decision tree with simple structure and the calculated amount is small, so it applies the problem in large-scale data set.

APPLICATION OF ANALYSIS AND DECISION OF STUDENT'S MARK BASED ON THE BIG DATA

The significance of student's mark

Whether the student master the knowledge is measured by the student's mark which is also a basis of evaluating the quality of teaching. In the college, teachers would accumulate so many data during teaching, but they cannot deal with them deeply and reasonably, but only have a primary data backup, searching and simple statistic which cannot be full shown. Now teaching affair management information system only simply search input and output of marks and seldom have analysis. If the analysis of student's mark just stays in the level of excellent, fine, pass and fail, and college are not aware of reason of mark, the volume of data will lose their meaning, they will just saved in the different forms and will not be found their significant meaning, and the evaluation of teaching will not gain benefits from them. With the increasing application of decision tree in the analysis of student's mark, higher education is improved by analyzing the deep-seated reason of student's mark, seeking problems and solving them. During decision support, data mining is a method to deeply analyzing the data. It is beneficial to apply decision tree in the evaluation of teaching which can comprehensively analyze the test's result and the relations among various reasons and transform much data into classification rule. In this way, these data can be analyzed better which has remarkable meaning in the improvement of quality of education. Here is a practical example about the application of analysis and decision of student's mark based on the big data.

Determine the object

During the general evaluation of each semester, universities want to find the relationships between marks, recreational and sports activities, social activities and English marks alone. So we choose such a model: database of student condition including student number, gender, English mark, recreational and sports activities, social activities, average marks, ranking and so on.

Preparation of data

After quantization, clearing, transformation and integration etc of database above, we will get a data warehouse shown in TABLE 1 which facilitates the next step--data mining. The fields of student number are l -50; the fields of gender are male of female; the fields of English awarded marks are transformed from ones of English marks whose definitions are following: 0 refers to the failure of CET4; 0.5 refers to the pass the CET4; 1 refers to the pass of CET6. The fields of awarded marks of social

activities are transformed from the conditions of social activities whose definitions are following: 0 refers to the seldom participation of activities; 0.2 refers to moderate participation of activities; 0.4 refers to often participation of activities. The fields of awarded marks of recreational and sports activities are transformed from the conditions of recreational and sports activities whose definitions are following: 0 refers to seldom participation of activities; 0.2 refers to participation of activities and gain the better results. The comprehensive condition of mark are represented by average scores and the fields are 0-100(hundred-mark system). The fields of the ranking are 1-50 and are recorded from the top to the bottom.

TABLE 1 : Database of student condition

Student number	Gender	Awarded marks of english	Awarded marks of social activities	Awarded marks of recreational and sports activities	Average results	Ranking
17	Male	0.5	0.4	0	85.27	1
25	Female	0.5	0.2	0.2	82.12	2
22	Female	0.5	0	0	81.11	3
20	Female	0.5	0	0	79.82	4
…	…	…	…	…	…	…
44	Male	0	0	0	67.22	47
18	Male	0	0	0	66.39	48
39	Female	0	0	0	66.33	49
11	Male	0	0	0	61.84	50

Data mining

In this period, an ID3 decision tree should be established and p positive examples and n negative examples will be determined. Now the first 15^{th} students are positive examples; and the lasting 35^{th} students are negative examples that is p=15, n=35

$I(p,n)=-15/50\log_2 15/50—35/50\log_2 35/50=0.88129$

E(awarded marks of English)=$24/50I(p_1,n_1)+7/50I(p_3,n_3)$

=$24/50I(11,13)+7/50I(4,3)$

=0.4775952+0.1379322=0.6155274

Gain(awarded marks of English)=I(p,n)-E(awarded marks of English)=0.88129-0.6155274=0.265726

Similar counts are:

Gain（awarded marks of social activities

)=I(p,n)-E(awarded marks of social activities

)=0.88129-0.79244=0.088847

Gain(awarded marks of recreational and sports activities)=I(p,n)-E(awarded marks of recreational and sports activities)=0.88129-0.79244=0.0001214

So we can see the awarded marks of English are the biggest one which is considered as the root node and from this node start the expansion, and so on shown in the Figure 3.

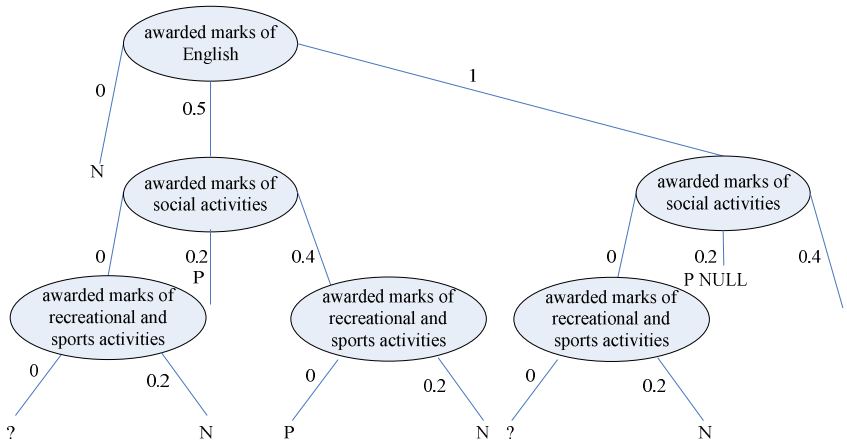


Figure 3 : Decision tree of analysis of marks

Analysis of results

Observe and analyze the decision tree on picture 3, and the conclusion are:

(1) Students who have not passed the CET4 or CET6, perform poorly in school;

(2) Students who have passed the CET6 pay more attention to study and do not take part in too many activities, some of whom take part in moderate activities can gain fine marks.

(3) The condition of those who pass the CET4 is complex. Marks of students who reasonably plan study, social activities, recreational and sports activities are not bad. While too many activities affects other student's study.

(4) Through analyzing the decision tree based on the database of students' conditions, the research primarily confirm one point that there are relationships between student's study and social activities, recreational and sports activities. If student can reasonably plan his or her study and social activities, recreational and sports activities, his or her study will be improved. But if students attach more importance to training their ability of social activities and neglect their study, their study will be affected. At the same time, it is not difficult to see those who reasonably plan their study and social activities comparatively have responsibility and desire to advance and meanwhile pay attention to their study. While students who do not gain fine marks also cannot plan their study and activities very well or they may not care about other things except study.

CONCLUSION

This research focuses on the application and decision of student's mark based on the big data. Firstly, it introduce the study of decision tree algorithm including summarization of sorting algorithms, sorting algorithms based on decision tree, the generative process of decision tree, construction algorithm model of decision tree and decision tree algorithm. Secondly, the application and decision of student's mark based on the big data is unfolded by analyzing the function of big data of factors influencing student's mark with the help of a practical example. It mainly includes steps of determining the object, preparing data, data mining, and analyzing results. Through the practical example, it is easy to see that with the help of decision tree, the relationships between student's marks and other factors are well analyzed, and teacher can base on which to carry out the measures to improve quality of teaching. If this idea is expanded, many results with significance will be gained; the evaluation of teaching will be improved; finally the quality of education will be enhanced by taking corresponding measures.

REFERENCES

[1] Wu Fa-Ti, Mou Zhi-Jia; Construction and implementation method of model analyzing student's individuation based on bid data in digitization schoolbag, China Educational Technology, **3**, **(2014)**.

[2] Zhu JIan-Hua; Education of news and communication in era of big data: Major setup, Student's kill and source of teachers, Journalist University, **4**, **(2013)**.

[3] Li Ke; Data filtering of students' perfunctory teaching evaluation in college P.E.Journal of Wuhan Institute of Physical Education, **9**, **(2013)**.