Journal of Applied Sciences

ISSN 1812-5654

Journal of Applied Sciences 13 (19): 3903-3908, 2013
ISSN 1812-5654 / DOI: 10.3923/jas.2013.3903.3908

Application and Study of Ordinal Decision Tree in the Teaching Quality Evaluation

[1]Hong-Yan Ma, [2]Jian-Kai Chen, [3]Nan Yang and [3]Li-Ling Wang
[1]Industrial and Commercial College,
[2]College of Mathematics and Computer, Hebei University, 071002, Baoding, China
[3]CNPC Beijing Richfit Information Technology Co., Ltd., 100007, Beijing, China

Abstract: Ordinal decision tree is one of the important ways of dealing with ordinal classification tasks. Ordinal decision tree based rank mutual information is representative of ordinal decision tree learning algorithms. Rank mutual information can be used to reflect the monotonous relevance between features and decision. Namely, it is useful for measuring the importance of attributes in ordinal classification. This study applies ordinal decision tree to teaching evaluation of colleges and universities for improving the level of teaching evaluation. The aim is to make teaching quality evaluation fair, reasonable and effective.

Key words: Ordinal decision tree, rank mutual information, teaching evaluation

INTRODUCTION

Nowadays, society is in an era of information explosion. There are more and more data information, which brings people' life much trouble, therefore data mining technology (Zhu *et al*., 2006) has attracted extensive attention from all fields of society. Compared with other methods of machine learning, decision tree is easy to be understood and implemented. The decision tree is one of important classifiers to deal with decision making. In any multistage decision process, the basic method is to break up a complex decision into a union of several simpler decisions, hoping the final solution obtained this method would meet the intended desired solution. Besides, decision tree (Mitchell, 1997) is one of the core technologies of data mining and the decision tree is often used to deal with the classification problems. At present, there are a variety of decision tree algorithms, such as continuous-valued attributes in decision tree generation (Wang *et al*., 2002, 2010; Wang and Shen, 2006). The ID3 (Quinlan, 1986) and C4.5 (Wu *et al*., 2008) algorithms are representative of the decision tree learning algorithm.

In addition, Ordinal classification tasks widely exist in real world life and work. In ordinal classification problems, the values of condition attribute and decision are ordinal and linearly ordinal, respectively. Specifically, if there is a constraint functions in an ordinal classification, namely, given two samples x and y, if $x \le y$, then it have $f(x) \le f(y)$, this call the ordinal classification with monotone constraints. The object of better values of condition attribute should not be assigned to a worse decision class. This call such problem is ordinal classification with monotone constraint and classification rule need to be order-preserving and monotonic. In ordinal classification problems, it is very important is to build an ordinal decision tree. There are many monotone ordinal classification (Ben-David *et al*., 1989) problems in real life and work. For example, according to the price and quality, it select commodities in a market. It evaluate consolidated results of students based on the student's performance and achievement. It evaluate teaching quality based on the information of teachers. The employers select their employees based on their education and experience. A bank wants to put loans for a client based on income, education level and criminal record of clients. In order to deal with ordinal classification tasks, Hu *et al*. (2010) put forward the concepts of rank entropy and rank mutual information in 2009. Soon afterwards, (Hu *et al*. (2012) constructed ordinal decision tree based rank mutual information, which can deal with ordinal classification tasks (the attributes and classes are ordinal). The monotonous relevance between features and decision can be reflected by rank mutual information. So, it is useful for measuring the importance of attributes in ordinal classification, multiple decisions (Dembczynski *et al*., 2008; Zopounidis and Doumpos, 2002).

With the progress of society, people have paid high attention to the education of colleges and universities. In recent years, due to the enrollment expansion of colleges and the decline of quality in new students, it is necessary to put forward higher requirements to the teachers'

Corresponding Author: Hong-Yan Ma, Industrial and Commercial College, Hebei University, Baoding, 071000, China
Tel: 13673223211

teaching quality. Teaching evaluation of colleges and universities is an important measure, which can affect the quality of teaching and the level of the teacher's teaching. Since teaching evaluation have the characteristics of complexity, multiple factors and fuzziness. How to make an objective, scientific, comprehensive evaluation is an important subject in modern teaching evaluation research. This study uses ordinal decision tree based rank mutual information to research the teaching evaluation.

PRELIMINARIES

Let $U = \{x_1, x_2,..., x_n\}$ be a set of samples and A be a set of attributes to describe the samples; D is a finite ordinal set of decisions. The value of x_i in attributes $a \in A$ or D is denoted by $v(x_i, a)$ or $v(x_i, D)$, respectively. The ordinal relations between object in terms of attribute a or D is denoted by $\leq$. There say x_j is no worse than x_i in terms of a or D if $v(x_i, a) \leq v(x_j, a)$ or $v(x_i, D) \leq v(x_j, D)$, denoted by $x_i \leq_a x_j$ and $x_i \leq_D x_j$, respectively. Correspondingly, it can also define $x_i \geq_a x_j$ and $x_i \geq_D x_j$.

Definition 1: Given an ordinal classification sample set U described with a set of attributes A, for $\forall x \in U$ $B \subseteq A$, $a \in B$, it associate x with the following sets (Hu *et al*., 2012):

$$[x]^{\leq}_a = \{y \in U: y \geq_a x\} \tag{1}$$

$$[x]^{\leq}_B = \{y \in U: y \geq_B x\} \tag{2}$$

It can easily obtain the following properties:

- If $C \subseteq B \subseteq A$ it have $[x_i]^{\leq}_C \supseteq [x_i]^{\leq}_B$
- If $x_i \leq_B x_j$, it have $x_j \in [x_i]^{\leq}_B$ and $[x_i]^{\leq}_B = \cup [x_j]^{\leq}_B$
- $[x_i]^{\leq}_B = \cup \{[x_j]^{\leq}_B | x_j \in [x_i]^{\leq}_B\}$,
- $\cup \{[x_j]^{\leq}_B | x_j \in U\} = U$

Definition 2: Given DT = <U, A, D>, $B \subseteq A$. The ascending and descending rank entropies of the system with respect to B are defined as (Hu *et al*., 2012):

$$RH^{\leq}_B(U) = -\frac{1}{n}\sum_{i=1}^{n} \log \frac{\left|[x_i]^{\leq}_B\right|}{n} \tag{3}$$

$$RH^{\geq}_B(U) = -\frac{1}{n}\sum_{i=1}^{n} \log \frac{\left|[x_i]^{\geq}_B\right|}{n} \tag{4}$$

Property 1: This have $RH^{\leq}_B(U)>0$ and $RH^{\leq}_B(U)>0$ as:

$$1 \geq \frac{\left|[x_i]^{\geq}_B\right|}{n} \geq 0$$

$RH^{\leq}_B(U)>0$ and $RH^{\leq}_B(U)>0$ if and only if $\forall x_i \in U$, $[x_i]^{\leq}_B = U$.

Definition 3: Given DT = <U, A, D>, $B \subseteq A$. The ascending and descending rank entropies of the system with respect to B are defined as (Hu *et al*., 2012):

$$RH^{\leq}_{B \cup C}(U) = -\frac{1}{n}\sum_{i=1}^{n} \log \frac{\left|[x_i]^{\leq}_B \cap [x_i]^{\leq}_C\right|}{n} \tag{5}$$

$$RH^{\geq}_{B \cup C}(U) = -\frac{1}{n}\sum_{i=1}^{n} \log \frac{\left|[x_i]^{\geq}_B \cap [x_i]^{\leq}_C\right|}{n} \tag{6}$$

Definition 4: GivenAT = <U, A, D> $B \subseteq A$, $C \subseteq A$. The Ascending Rank Mutual Information (ARMI) and Descending Rank Mutual Information (DRMI) of the set U between B and C are defined as (Hu *et al*., 2012):

$$RMI^{\leq}(B, C) = -\frac{1}{n}\sum_{i=1}^{n} \log \frac{\left|[x_i]^{\leq}_B\right| \times \left|[x_i]^{\leq}_C\right|}{n \times \left|[x_i]^{\leq}_B \cap [x_i]^{\leq}_C\right|} \tag{7}$$

$$RMI^{\geq}(B, C) = -\frac{1}{n}\sum_{i=1}^{n} \log \frac{\left|[x_i]^{\geq}_B\right| \times \left|[x_i]^{\geq}_C\right|}{n \times \left|[x_i]^{\geq}_B \cap [x_i]^{\geq}_C\right|} \tag{8}$$

This can also get the upwards and downwards ranking mutual information of the set U regarding B and D: $RMI^{\leq}(B, D)$. In essence, the rank mutual information RMI(B, D) can measure the monotonic degree of the ordinal consistency between attributes B and decisions D. That is it can be used to measure the importance of attributes in ordinal classification tasks, multiple decisions and ranking analysis.

ALGORITHM AND MONOTONIC CONSISTENT RULES

Ordinal decision tree algorithm: Ordinal decision tree is one of the important ways of dealing with ordinal classification problems. It is a crucial problem to select expanded attributes in building ordinal decision trees. This first needs to compute the rank mutual information of each cut for each of the continuous-valued attributes during the selection of expanded attributes for learning of decision trees and compare these values of rank mutual information to get the maximum which corresponds to the expanded attribute. The expanded attribute builds a new node and divides samples to two different subsets according to the values of the expanded attribute. Recursively it produces new splits until the class of samples is same in each subset.

The ordinal decision tree algorithm based on rank mutual information is formulated as follows:

Input: attributes of samples.
Decision: decision of samples.
Stopping criterion: If the maximal rank mutual information is less than ε, the branch stops growing, or if the number of sample is 1 or all the samples come from the same class, the branch stops growing
Output: ordinal decision tree T.
Begin: generate the root node.
Step 1: For (each attribute $a_i \in$ Criteria)
Step 2: For (each $c_j \in a_i$)
Step 3: Divide samples into two subsets according to c_j.
If $a_i(x) \le c_j$, then $a_i(x) = 1$, else $a_i(x) = 2$.
Compute $RMI(a_i, c_j) = RMI(\text{Criteria, Decision})$
End j
Step 4: Select c^*_j, $c_j^* = \arg\max_j RMI(a_i, c_j)$
End i
Step 5: Select the best feature and the corresponding splitting point: $(a, c^*) = \max_i RMI(a_i, c_j^*)$
Step 6: If $RMI(a, D) < \varepsilon$, stop.
Step 7: Build a new node and split samples with a, c*
Step 8: Recursively produce new splits according to step1-step5 until stopping criterion is satisfied.
End

Monotonic consistent rules: Now the following study the properties of ordinal decision trees based rank mutual information generated with the above procedure.

Definition 5: Hu *et al*. (2012) given ordinal decision tree T, the rule from the root node to a leaf l is denoted by R_l. If two rules R_l and R_m are generated from the same attributes, it say R_l and R_m are comparable; otherwise they are not comparable. In addition, this denote $R_l < R_m$ if the feature value of R_l is less than R_m. (R_l and R_m are also called left node and right node, respectively).

As to a set of rules R, it call it is monotonically consistent if for any comparable pair of rules R_l and R_m, if $R_l < R_m$, then $D(R_l) < D(R_m)$, where $D(R_l)$ and $D(R_m)$ are the decisions. Otherwise, it says R is not monotonically consistent.

This gives a toy example to show the difference between ordinal decision tree and decision tree. Now there introduces a monotonic function to produce monotone datasets:

$$f(x_1, x_2) = 1 + x_1 + \frac{1}{2}\left(x_2^2 - x_1^2\right) \quad (9)$$

where, x_1, x_2 are two random variables independently drawn from the uniform distribution over the unit interval. In order to generate ordered class labels, the resulting numeric values are discretized into k interval [0, 1/K], (1/k, 2/k],..., (k-1/k]. Firstly generating artificial data of 30 samples, the class number is 4. Now there employs CART and the above ordinal decision tree algorithm to train decision trees, respectively. Figure 1 and 2 give the trained models.

Figure 1 and 2 show the decision tree generated with CART and ordinal decision tree algorithm based rank mutual information, respectively. Most of all, the tree trained with CART is not monotonically consistent for there are two pairs of conflicted rules (Fig. 3). From left to right, it is easy to see the fourth rule is not monotonically consistent with the fifth rule. The sixth rule is not monotonically consistent with the seventh rule, where the samples with the better features get the worse decision Fig. 4. However, the ordinal decision tree algorithm based rank mutual information gets a consistent decision tree.

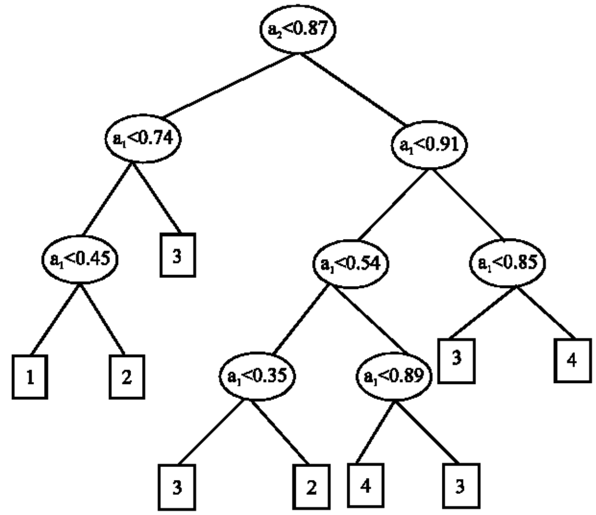

Fig. 1: Decision tree generated with CART

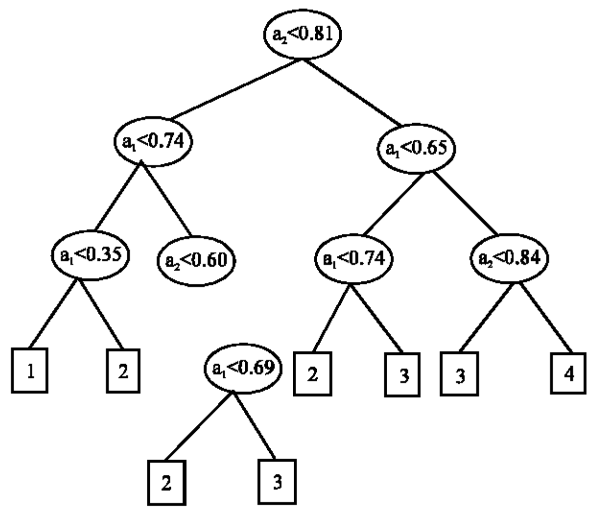

Fig. 2: Ordinal decision tree

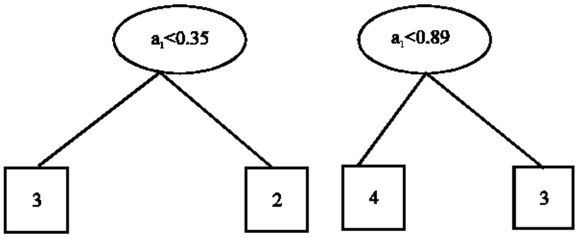

Fig. 3: Non-monotonic consistent rules

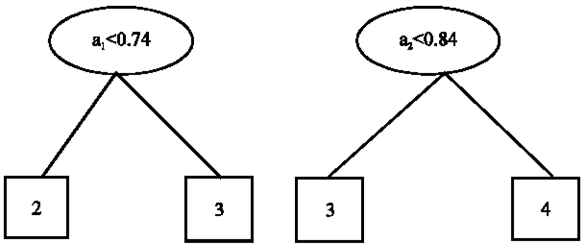

Fig. 4: Monotonic consistent rules

APPLICATION

Through the above theoretical knowledge, this section applies ordinal decision tree based rank mutual information to the teaching quality evaluation. It includes four parts: problems analysis, data collection, data classification mining and extraction of classification rules.

Problems analysis: With the development of the society, education is more and more become the hot topics, which people talk about. There is no doubt that the school teaching is an important part of education, while teaching quality is the key link of the teaching process, is also an important part of teaching evaluation. The school teaching quality not only affects the students' learning situation, further influence the progress of the society.

Everyone know that the quality of teaching evaluation index system can be divided into several items: teaching attitude, teaching methods, teaching contents, communication ability, teaching effect and the evaluation results. Assuming that the values of evaluation indexes and evaluation results attributes include four grades (excellent, good, pass and fail), respectively. It is obvious that these values have an order relation (fail<pass<good<excellent). Namely, the evaluation indexes and evaluation results are ordered. In addition, the evaluation index and the evaluation results exists certain monotone ordered relation. For example, if one teacher' teaching methods are scientific and reasonable, the evaluation results should not be bad. There is a certain order relation between them but how to measure the relation of the four evaluation indexes and evaluation results or which evaluation index is most important. It has a great help to improve the quality of teaching, if these problems will de solved reasonably.

Table 1: Data information

	Teaching evaluation index attributes				
ID	α_1	α_1	α_3	α_4	Evaluation results (D)
1	Good	Good	Good	Excellent	Good
2	Good	Good	Good	Good	Good
3	Pass	Pass	Pass	Pass	Pass
4	Good	Pass	Pass	Pass	Pass
5	Excellent	Excellent	Excellent	Excellent	Excellent
6	Excellent	Excellent	Excellent	Good	Excellent
7	Pass	Pass	Pass	Good	Pass
8	Good	Good	Good	Pass	Good

Through the above analysis, if the evaluation indexes and the evaluation results are seen as the condition attributes and decision attributes, respectively, this problem can be converted into ordinal classification task. Therefore that can deal with the problem with ordinal decision tree, which is given in the following subsection.

Data collection: Teaching evaluation analysis aims to find the function relation between the teacher's basic information and evaluation results. It needs the teacher's basic information and evaluation results from all aspects. In the past, it takes a lot of time to evaluate the teaching information by making questionnaire and error rate is quite high. To solve the problem of large data statistics, there need to build a system to collect data and mining the rules of data, such as web teaching evaluation system. 500 samples are randomly extracted from the system, which is from questionnaire. Evaluation indexes have four features: teaching attitude, teaching methods, teaching contents, teaching effect. Let a_1 = {teaching attitude} a_2 = {teaching contents}, a_3 = {teaching methods} a_4 = {teaching effect}. Assuming that the values of evaluation indexes and evaluation results attributes include four grades (excellent (90-100), good (80-89), pass (60-79) and fail (less than 60)) Table 1.

Data classification mining: Ordinal decision tree is also one of the important methods of data mining. It is a process of classifying the training data by a series of monotonic rules, which comes form the generated ordinal decision tree. The process is shown in Fig. 5.

Before data mining, data should be made a preprocessing. But this should follow a principle that it will not affect the results of data mining. Since ordinal decision tree only handle for the data of the numerical attributes in data mining, it make a preprocessing for the data. The evaluation results will be divided into four classes (0-fail, 1-pass, 2-good, 3-excellent) and the values

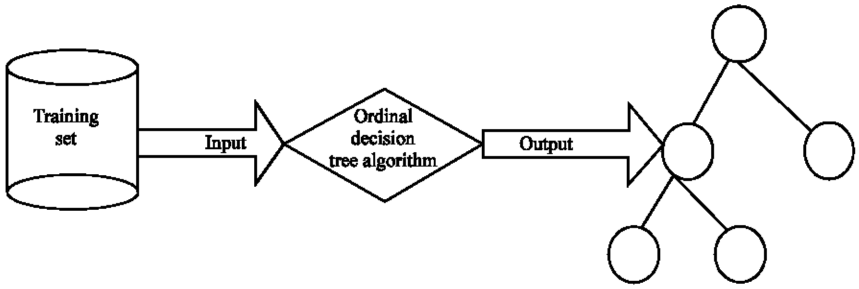

Fig. 5: Mining process of ordinal decision tree

Table 2: After data preprocessing

Teacher ID	Teaching evaluation index attributes α_1	α_1	α_3	α_4	Evaluation results
1	0.80	0.82	0.88	0.91	2
2	0.81	0.85	0.86	0.88	2
3	0.75	0.70	0.71	0.65	1
4	0.82	0.63	0.75	0.77	1
5	0.90	0.91	0.93	0.98	3
6	0.95	0.97	0.95	0.88	3
7	0.65	0.76	0.68	0.78	1
8	0.75	0.82	0.85	0.75	2

of the four evaluation indexes are discretized into the interval [0, 1]. After data preprocessing, the dataset is shown in Table 2.

Through the problem analysis, teaching evaluation index attributes and evaluation results are ordinal, which conform to the ordinal classification tasks. In order to make the generated rules easy to understand, this study chooses to use ordinal decision tree method. Ordinal decision tree algorithm based on rank mutual information select rank mutual information $RMI^{\geq}(B, D)$ as splitting rule. It selects expanded attributes, in which the value of rank mutual information gets the maximum. Recursively it produces new splits until the class of samples is same in each subset.

According to ordinal decision tree algorithm based on rank entropy, an ordinal decision tree will be constructed with the datasets Table 2. The process is as follows.

In order to simplify the process and be easy to understand Here randomly draw 50 training samples from the set and guarantee that there is at least one representative sample from each class (excellent, good, pass). By the ordinal decision tree algorithm based on rank mutual information (step 3), There first compute the rank mutual information $RMI(a_i, c_j, D)$, which measures the importance of attributes in ordinal classification tasks:

$$\text{Max}\,RMI(a_1, c_j, D) = RMI^{\geq}(a_1, 0.85, D) = -\frac{1}{50}\sum_{i=1}^{50}\log\frac{\left|[x_i]^{\geq}_{a_1}\right|\times\left|[x_i]^{\geq}_{D}\right|}{50\times\left|[x_i]^{\geq}_{a_1}\cap[x_i]^{\geq}_{D}\right|}$$

$$= -\frac{1}{50}\log\frac{50\times50}{50\times50} - \frac{1}{50}\log\frac{50\times48}{50\times48}\ldots\ldots - \frac{1}{50}\log\frac{25\times3}{50\times3} - \frac{1}{50}\log\frac{25\times2}{50\times2} - \frac{1}{50}\log\frac{25\times2}{50\times2} = 0.5683$$

By the algorithm step 1-4, analogically:

$$RMI^{\geq}(a_2, 0.91, D) = 0.6128$$

$$RMI^{\geq}(a_3, 0.81, D) = 0.5128$$

$$RMI^{\geq}(a_4, 0.91, D) = 0.5049$$

The result is:

$$RMI^{\geq}(a_2, 0.91, D) > RMI^{\geq}(a_1, 0.85, D) = 0.6128 > RMI^{\geq}(a_3, 0.81, D) > RMI^{\geq}(a_4, 0.79, D)$$

Rank mutual information $RMI^{\geq}(a_2, D)$ gets the maximum value on c = 0.91, therefore a_2 is selected as root node. Recursively producing new splits according to step 1-5 until stopping criterion is satisfied, it produces an ordinal decision tree. As shown in Fig. 6.

Application of classification rules: Ordinal decision tree can also be converted into classification rules. Namely, it can be indicated the rules IF-THEN, which is easy to understand. For example, this can extract the classification rules from the Fig. 1. If (a_2>0.91 and a_1>0.85), then (evaluation result is excellent); If (a_2>0.91, a_1>0.76, a_2>0.78 and a_3>0.87), then (evaluation result is good); If (a_2>0.91, a_1>0.76 and a_2>0.78), then (evaluation result is pass) and so on. In addition, the rules from the ordinal decision tree (Fig. 6) are monotonically consistent, which pays an important role to make evaluation efficiently. From the above analysis, it shows that the teachers' teaching contents is the key in the teaching process, which dominates teaching evaluation stats. Therefore teaching contents is firstly considered, teaching attitude take second place. Then other factors should be considered in

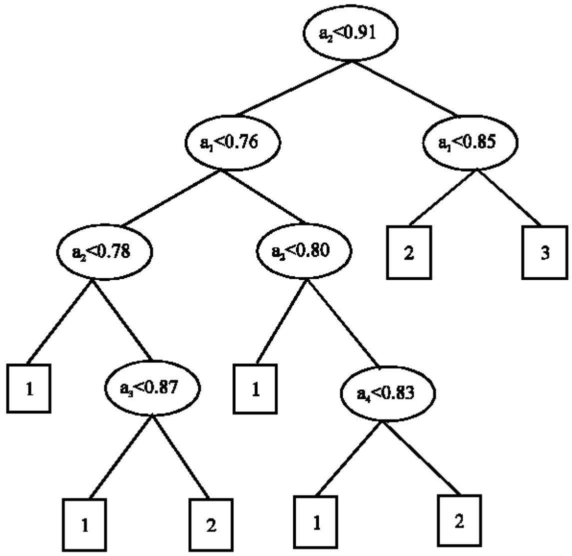

Fig. 6: Ordinal decision tre

the process of teaching evaluation. Through the practice test, the results are consistent with the reality. This method can make an effective judgment due to the generated monotonic rules. In addition, the generated monotonous rules can make a prediction for one teacher. The teachers can find shortcomings by the system and take measures to improve themselves, which can improve the level of teaching.

CONCLUSION

Ordinal classification is a kind of important tasks in management decision, evaluation and assessment, where attributes and classes of objects are ordinal. Ordinal decision tree is one of the main ways of dealing with ordinal classification problems. This study applies it to the research of the teaching quality evaluation. In theory, it can make a fair, reasonable and effective teaching evaluation and can solve some problems of teaching. In the future, it will be more perfect services in the field of education management, because data mining applied to the field of education will have a vast potential for future development under the trend of educational informationization.

ACKNOWLEDGMENTS

This work was supported by the Natural Science Foundation of China (Grant No.11101115) ,the National Natural Science Foundation of China (Grant No. 61170254), the Natural Science Foundation of Hebei Province (Grant No. F2011201039), and the Science and Technology Research and Development Guidance Plan Project of Baoding City (Grant No. 13ZG012).

REFERENCES

Ben-David, A., L. Sterling and Y.H. Pao, 1989. Learning and classification of monotonic ordinal concepts. Comput. Intell., 5: 45-49.

Dembczynski, K., W. Kotlowski and R. Slowinski, 2008. Ensemble of decision rules for ordinal classification with monotonicity constraints. Proceedings of the 3rd International Conference on Rough Sets and Knowledge T echnology, May 17-19, 2008, Chengdu, China, pp: 260-267.

Hu, Q.H., M.Z. Guo, D.R. Yu and J.F. Liu, 2010. Information entropy for ordinal classification. Sci. China Inform. Sci., 53: 1188-1200.

Hu, Q.H., X.J. Che, L. Zhang, D. Zhang, M. Guo and D. Yu, 2012. Rank entropy-based decision trees for monotonic classification. IEEE Trans. Knowl. Data Eng., 24: 2052-2064.

Mitchell, T.M., 1997. Machine Learning. WCB/McGraw-Hill, Boston, MA., USA.

Quinlan, J.R., 1986. Induction of decision trees. Mach. Learn., 1: 81-106.

Wang, X.Z. and J. Shen, 2006. Using special structured fuzzy measure to represent interaction among IF-THEN rules. Proceedings of the 4th International Conference on Advances in Machine Learning and Cybernetics, August 18-21, 2005, Guangzhou, China, pp: 459-466.

Wang, X.Z., H.W. Yang, M.H. Zhao and J. Sun, 2002. A decision tree based on hierarchical decomposition. Proceedings of the International Conference on Machine Learning and Cybernetics, Volume 4, November 4-5, 2002, Beijing, China, pp: 1824-1828.

Wang, X.Z., X.H. Gao and Q. He, 2010. Side effect of cut in decision tree generation for continuous attributes. Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, October 10-13, 2010, Istanbul, pp: 1364-1369.

Wu, X., V. Kumar, J.R. Quinlan, J. Ghosh and Q. Yang *et al*., 2008. Top 10 algorithms in data mining. Knowl. Inform. Syst., 14: 1-37.

Zhu, Y.Q., H.B. Yang and L. Sun, 2006. Data Mining Technology. Southeast University Press, Nanjing, China.

Zopounidis, C. and M. Doumpos, 2002. Multicriteria classification and sorting methods: A literature review. Eur. J. Oper. Res., 138: 229-246.