Automatic Leveling System for E-Learning Examination Pool Using Entropy-Based Decision Tree

Shu-Chen Cheng[1], Yueh-Min Huang[2], Juei-Nan Chen[2], and Yen-Ting Lin[1]

[1] Department of Computer Science and Information Engineering, Southern Taiwan University of Technology, Tainan 710, Taiwan
kittyc@mail.stut.edu.tw, ricky014789@yahoo.com.tw
[2] Department of Engineering Science, National Cheng Kung University, Tainan 701, Taiwan
huang@mail.ncku.edu.tw, nan@www.mmn.es.ncku.edu.tw

Abstract. In this paper, we propose an automatic leveling system for e-learning examination pool using the algorithm of the decision tree. The automatic leveling system is built to automatically level each question in the examination pool according its difficulty. Thus, an e-learning system can choose questions that are suitable for each learner according to individual background. Not all attributes are relevant to the classification, in other words, the decision tree tells the importance of each attribute.

1 Introduction

Due to the rapid growth of information and communication technologies, the education environment has being enriched with those technologies and become more diversified. Obviously, the Internet-based environment has been rapidly replacing the traditional lecture-based one due to the factor of providing the multimedia-enhanced learning anytime and anywhere. Nowadays, there are more and more instructors who have been making Internet-based courseware or even curriculum of distance learning for extending education purpose. In fact, many research works have focused on employing multimedia and Internet technologies to produce more attractive or effective course contents. Besides the effective courseware, we believe that developing an effective learning assessment is another key for quality-assured e-Learning. Due to its important role, some researchers have developed some techniques for the area, such as automatic classifier for assessment items [7], question databases [8], assessment items for cooperative items [9], etc. In [7], they design an automatic classifier for Chinese assessment items in terms of particular keywords. Sumitomo et al. also employed keyword search to analyze data in their designed question-answer database. In [9], he further divided the assessments into self and peer parts for collaborative learning purpose. However, how to design an assessment with an individual background involved is still an open issue.

Many studies have been working on the development of individual course content by using artificial intelligence (AI), data mining, as well as agent technologies [1][2]. Their common purpose is to correctly assist a learner to achieve an optimal learning effectiveness according to his/her current background. Similarly, the design of an effective assessment must be able to examine an individual learning status so that an

R.W.H. Lau et al. (Eds.): ICWL 2005, LNCS 3583, pp. 273 – 278, 2005.

instructor can properly progress his/her teaching. In other words, it is desired to design an individual assessment which can explore the capability of an individual student. In 2004, Büchner and Patterson suggested platforms to keep track of learners' activities including content viewed, time spent and quiz results [3]; Kuo et al. proposed a real-time learning behavior mining algorithm [4].

Over the last years, we have witnessed an explosive growth of e-learning. Internet has been widely used in various fields. More and more learning contents have been published and shared over the Internet. Therefore, how to progress an efficient learning process becomes a critical issue. For example, there are various applications to facilitate learners, especially for distance education. However, general learning system cannot provide suitable learning materials to achieve efficient learning. More and more people try to apply the artificial intelligence techniques, such as the agent technology [1][2], to the application of distance learning. In 2003, Fei et al. designed the question classification for e-learning using artificial neural network [5]. In 2004, Merceron et al. proposed a question answering mining platform [6]. In this paper, an automatic leveling system for e-learning examination pool is proposed.

The remainder of the paper is organized as follows. In Section 2, the function of e-learning examination pool is described. In Section 3, we propose the automatic leveling system for e-learning examination pool using the algorithm of the decision tree. Finally, the conclusions are made in Section 4.

2 E-Learning Examination Pool

An example is described by the values of attributes and Boolean values for the classification of the examples. In Table 1, learners with individual backgrounds are divided

Table 1. An example of learners with individual backgrounds

	Attributes					
Learner	**Educational background**	**Department**	**Institution**	**Hobby**	**Occupation**	**Answer**
No.1	University	CSIE	TKU	Computer	Student	Correct
No.2	Vocational school	ME	LIT	Playball	Doctor	Wrong
No.3	Institute of technology	AI	NTCB	Internet	Civil servant	Wrong
No.4	Institute	EE	NCKU	Swimming	Student	Correct
No.5	Vocational school	TI	TKCVS	Shopping	Service trade	Wrong
No.6	University	ME	NCCU	Reading	Student	Wrong
No.7	University	AI	CYCU	Computer	Lawyer	Correct
No.8	University	CDIE	PU	Shopping	Service trade	Wrong
No.9	Academic	EE	NTU	Fishing	Teacher	Correct
No.10	Institute	EE	FJU	Movie	Student	Correct
⋮	⋮	⋮	⋮	⋮	⋮	⋮

into two groups. One is for learners whose answer is correct and the other is for those whose answer is wrong. For each question in the examination pool, all learners' behaviors can be recorded as shown in Table 1. After automatic leveling system is built, each question would be leveled automatically according its difficulty and an e-learning system can choose questions that are suitable for each learner.

3 Automatic Leveling System

The automatic leveling system is built to automatically level each question in the examination pool according its difficulty. If most learners with similar background can answer a question correctly, the question is labeled as "easy", and vice versa. Thus, an e-learning system can choose questions that are suitable for each learner according to individual background. As shown in figure 1, the input of the automatic leveling system is the values of attributes and Boolean values for the classification of the examples. The output of the automatic leveling system is a decision tree. Figure 2 illustrates the automatic leveling system can automatically level each question so that a suitable question can be chosen for a new learner and the new data can feedback the result to the automatic leveling system. How to build the automatic leveling system using the algorithm of decision tree is illustrated in Section 3.1 and Section 3.2.

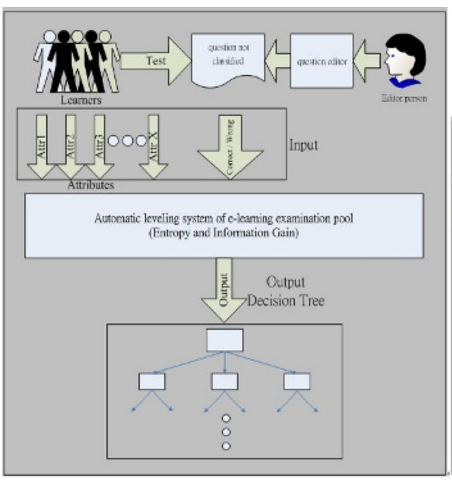

Fig. 1. System Architecture of Automatic Leveling System

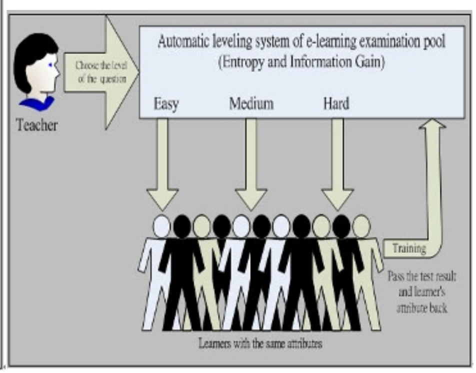

Fig. 2. Purpose of Automatic Leveling System

3.1 Entropy

Entropy measures the impurity of a data set S and is defined as:

$$\text{Entropy (S)} = -P_+ \log_2 P_+ - P_- \log_2 P_- \tag{1}$$

where P+ is the proportion of positive examples while P- is the proportion of negative ones. The entropy of S is zero when the examples are all positive or all negative. The entropy reaches its maximum value of 1 when half of examples are positive and half are negative.

3.2 Decision Tree and Information Gain

In Figure 3, we see a tree used to determine whether a learner's answer is correct or wrong. This kind of tree is called a decision tree [10][11]. A decision tree takes in the attribute values and outputs a Boolean decision. Once a decision tree is constructed, we start at the root node to check the attribute value. Based on the attribute value, the branch labeled with the corresponding value is chosen. Continue checking the next attribute value until a leaf node is reached.

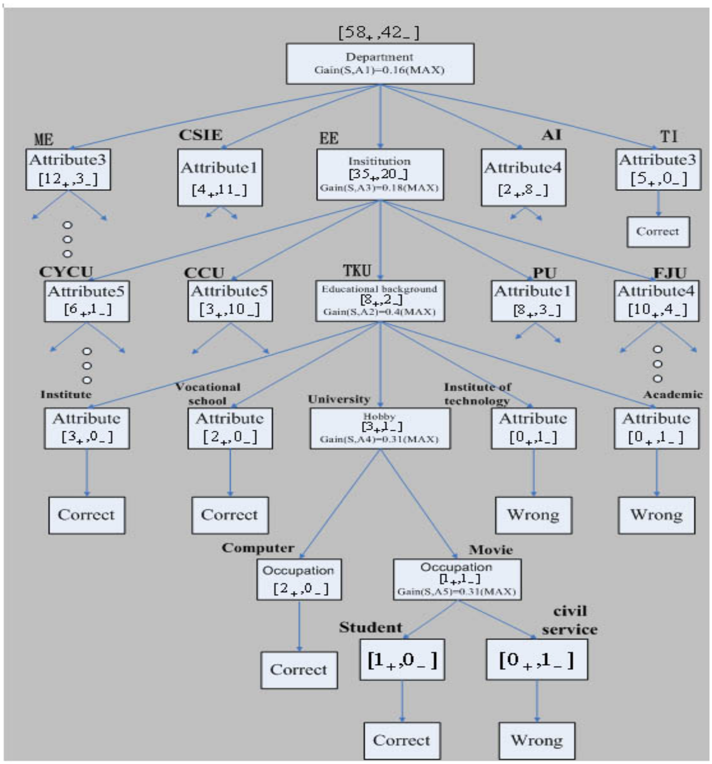

Fig. 3. A decision tree for classifying learners' answers

Decision tree induction involves a set of training data to generate a decision tree that can classify the training data correctly. If the training data represent the entire space of possible data adequately, the decision tree will then correctly classify new input data as well. The best-known decision tree induction algorithm is ID3, which was proposed by Quinlan in 1980s. The algorithm builds a decision tree from top down and finds the shortest possible decision tree to classify the training data correctly. The method used by the algorithm to determine which attribute to be chosen for each node is to select the attribute that provides the greatest information gain at each stage [10][11]. A perfect attribute divides the examples into sets that are all positive or all negative. The information gain of an attribute tells us how close to perfect the attribute is. Information gain is defined as the reduction in entropy and calculated by:

$$Gain(S,A) = Entropy(S) - \sum_{V \in values(A)} \frac{|S_V|}{|S|} Entropy(S_V) \tag{2}$$

Where Entropy(S) denotes the entropy of a data set S and can be calculated by equation 1. $Gain(S,A)$ is the expected reduction in entropy due to sorting S on attribute A. Looking at the attribute "Department" as shown in figure 3, we can calculate the entropy for the data set and its subsets. The total information gain, $Gain(S,A)$, is defined as the entropy of the set minus the weighted sum of the entropies of subsets. Hence, at this stage, the information gain for the attribute "Department" is 0.16. Considering another attribute "Institution", the information gain for this attribute is 0.02. After considering all of the attributes, we found the attribute "Department" provides the greatest gain; therefore, the attribute "Department" is placed as the root at this stage. Apply the method recursively until a decision tree is constructed.

4 Conclusions

In this paper, we propose an automatic leveling system for e-learning examination pool using the algorithm of the decision tree. After the decision tree is constructed, we can see that not all attributes are relevant to the classification; therefore, the classification might be made by only 2 or 3 attributes. In other words, the decision tree tells the importance of each attribute. Unfortunately, it is quite possible that even when vital information is missing, the decision tree learning algorithm will find a decision tree that is consistent with all the examples. This is because the algorithm uses the irrelevant attributes. In the future, some technique to eliminate the dangers of over-fitting must be developed in order to select a tree with good prediction performance.

References

1. Y.M. Huang, J.N. Chen, K.T. Wang, C.H. Fu, "Agent-Based Web Learning System," IEEE Learning Technology Newsletter, Vol. 6, Issue 2, pp. 38 – 41, April 2004
2. Y.M. Huang, J.N. Chen, S.C. Cheng, W.C. Chu, "Agent-Based Web Learning System Applying Dynamic Fuzzy Petri Net," Proceeding of International Conference on Web-Based Learning, pp. 338 – 345, 2004
3. Alex G. Büchner, David Patterson, "Personalised E-Learning Opportunities", Proceedings of 15th International Workshop on Database and Expert Systems Applications, pp.410 – 414, 2004

4. Y.H. Kuo, J.N. Chen, Y.M. Huang, “Real-time Learning Behavior Mining Algorithm,” IEEE Learning Technology Newsletter, Vol. 6, Issue 4, pp. 89 – 92, October 2004
5. Ting Fei, Wei jhy Heng, Kim Chuan Toh, Tian Qi, “Question Classification for E-learning by Artificial Neural Network”, ICICS-PCM, 2003
6. Duval, P., Merceron, A., Rinderknecht, C., Scholl, M., “LeVinQam: a question answering mining platform”, ITHET 2004. Proceedings of the Fifth International Conference on, 2004
7. Chein B.C. And Liau S.T., “An Automatic Classifier for Chinese Items Categorization”, Proceedings of 2004 National Conference on Artificial Intelligence and Its Application, Taiwan, 2004
8. Chisa Sumitomo, Akira Okada and Yahiko Kambayashi, “Design and Evaluation of lecture Support Functions for Question Databases”, Proceedings of 1st International Conference on Web-based Learning, ICWL2002, Hong Kong, China, 2002
9. Park C., “Development of Self and Peer Assessment Items in Web-based Cooperative Learning”, Proceedings of 3rd International Conference on Web-based Learning, ICWL2004, China, 2004
10. Stuart Russell, Peter Norvig, “Artificial Intelligence: A Modern Approach”, Prentice Hall, pp.531-544, 1995
11. Ben Coppin, “Artificial Intelligence Illuminated”, Jones and Bartlett, pp.276-286, 2004