Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 34 (2008) 596–602
Expert Systems with Applications
www.elsevier.com/locate/eswa

Course planning of extension education to meet market demand by using data mining techniques – an example of Chinkuo technology university in Taiwan

Tai-Chang Hsia [a,*], An-Jin Shie [b], Li-Chen Chen [c]

[a] Department of Industrial Engineering and Management, Chienkuo Technology University, Changhua, Taiwan, ROC
[b] Department of Industrial Engineering and Management Yuan Ze University, Taoyuan, Taiwan, ROC
[c] Extension Education Center Chienkuo Technology University Changhua, Taiwan, ROC

Abstract

This study used data mining techniques to analyze the course preferences and course completion rates of enrollees in extension education courses at a university in Taiwan. First, extension courses were classified into five broad groups. Records of enrollees in extension courses from 2000-5 were then analyzed by three data mining algorithms: Decision Tree, Link Analysis, and Decision Forest. Decision tree was used to find enrollee course preferences, Link Analysis found the correlation between course category and enrollee profession, and Decision Forest found the probability of enrollees completing preferred courses. Results will be used as a reference for curriculum development in the extension program.

Keywords: Data mining; Extension education; Decision tree algorithm; Link analysis algorithms; Decision forest algorithm

1. Introduction

In the last two decades, a revolution in education has followed on the heels of economic growth and political reform. The number of institutions of higher education has increased steadily, and vocational junior colleges offering two year degrees are being phased out, upgrading to universities/institutes of technology offering 4 year degree programs. To meet the demand for higher education, the number of institutions of higher education in Taiwan is expected to increase to 155 over the next twenty years, up from the current 30 (Department of Higher Education, 2005). Education has gone from being the province of elites to a commodity accessible to anyone.

As the number of colleges has increased, colleges and universities have found themselves facing stiff competition for students. As a result, many colleges and universities have established Continuing/Extension Education Centers to provide accredited and non-accredited courses as well as short- and medium-term vocational training for adults, both in and out of the employment pool. These institutions take as a foundational principle that successful long-term management of extension education must be focused on satisfying the needs of the community, local industry, and local professional development.

The purpose of this research was to use data mining techniques to uncover the preferences and future choices of continuing education students of an Extension Education Center at a university in Taiwan. This data would then be used to better target curriculum on student needs. Decision Tree Algorithms, Link Analysis Algorithms, and Decision Forest Algorithms developed from the theory of Data mining were used in this research. Data consisted of the student records of enrollees in courses in the 5 academic years from 2000 to 2005 planned by the Extension Education Center of Chienkuo Technology University (hereinafter CTU).

* Corresponding author. Tel.: +886 919063119; fax: +886 47111141.
E-mail address: tchsia@ctu.edu.tw (T.-C. Hsia).

0957-4174/$ - see front matter
doi:10.1016/j.eswa.2006.09.025

Results of the research included (1) course preferences of enrollees; (2) the relationship between course categories offered and enrollee profession; (3) the relationship between course preferences, enrollee profession and the probability of course completion. These results will be used as a reference in future curriculum development at the Extension Education Center of CTU.

2. Literature review

Scholars use a variety of different definitions for data mining. Frawley, Paitetsky-Shapiro, and Matheus (1991) declared that data mining is actually a process of discovering of nonobvious, unprecedented, and potentially useful information. Curt (1995) defined data mining as a database transformation process, in which the information is transformed from unorganized vocabulary and number to organized data, and later turned into knowledge from which a decision can be made. Fayyad, Piatesky-Shapiro, and Smyth (1996) stated that data mining is an uncomplicated process of discovering the valid, brand new, potentially useful, and comprehensive patterns from data. Hui and Jha (2000) defined data mining as an analysis of automation and semi-automation for the discovery of meaningful relationships and rules from a large amount of data in a database. Peacock (1998) declared that data mining can be categorized as Narrow and Broad. The narrow definition is limited by the methodology of mechanical learning which emphasizes the discovery process and uses artificial intelligence such as Neural Networks, Correlation Rule, Decision Tree Algorithms and Genetic Algorithms. By contrast, the broad definition emphasizes the knowledge discovery in database (KDD), the process of obtaining, transforming, clarifying, analyzing, confirming and enduring the meaning of the data within existing customers or outside of the cooperation, and then results in a backup system of decision making which is continuously being modified and maintained. Hand, Blunt, Kelly, and Adams (2000) stated that data mining is a process that discovers interesting and valuable information from a database. Berson, Smith, and Thearling (2001) argued that the appeal of data mining lies in its forecasting competence instead of merely in its ability to trace back. Hui and Jha (2000) indicated that the data mining process should include the following steps:

1. Establishing the mining goals: using domain knowledge to select data relevant to the research goal.
2. Selection of data: identifying the characteristics of variables on which mining can be performed.
3. Data pre-processing: removing noisy, erroneous, and incomplete data.
4. Data transformation: transforming the data into a new format in order to mine additional information.
5. Data warehousing: the process of envisioning, planning, building, using, managing, maintaining, and enhancing databases.
6. Data mining: discovering correlations among variables after performing data mining and finding interesting, meaningful, and valuable knowledge based on the research topic.
7. Evaluating the mining results: elaborating and evaluating the results after knowledge is obtained.

To summarize the foregoing definitions, data mining is a process of obtaining knowledge. The key to the process is comprehension of the research application, and then constructing a data set by collecting data relevant to the research field, purifying the data in the targeted database to eliminate erroneous data, supplementing missing data, simplifying and transforming the data, and at last discovering the patterns and among the data and presenting them as useful knowledge.

Artificial intelligence has made remarkable progress in recent decades. A diversity of data mining algorithms have been developed for application to different types of data. Brief descriptions of these algorithms follow.

1. Market Basket Analysis (MBA): MBA clusters similar data. It is similar to cluster analysis, but MBA can analyze the probability for the types of goods purchased not only at the same time, but also at different times. Thus, when customers repeatedly purchase certain goods within a specific and short period of time, MBA can cluster the goods into a combination of goods in terms of rules and probability. MBA is commonly used on large amounts of trading data, when the topic to be discovered is not known. MBA is superior to cluster analysis in discovering combinations of goods.
2. Memory Based Reasoning (MBR): this algorithm combines the algorithm of Artificial Intelligence Neural Networks and Genetic Algorithms. MBR uses the theory of *k*-Nearest Neighbor to carry out k data clustering of data in the database, and outputs the type of clustering dots by using the distance function of the algorithm data type. It obtains the best clustering result through a search by the genetic algorithm. MBR is mainly used for clustering and attribute forecasting of continuity and discreteness. This means that MBR largely analyzes previous data and finds neighboring data for newly added data, clusters them, and then forecasts the continuity and discreteness. MBR excels at text and image processing.
3. Automatic Cluster Detection: this algorithm is actually derived from a map of self-organizing neural networks. The data are clustered into subsets based on degree of similarity. Data grouped into the same subset are more similar to each other than they are to data in other subsets. The algorithm automatically detects differences between the groups in the database, selects groups with similar characteristics, and clusters them with respect to segmentation and differentiation among different groups. When the data and variables are massive and large, this algorithm can be used for clustering before