World Academy of Science, Engineering and Technology
International Journal of Educational and Pedagogical Sciences
Vol:8, No:1, 2014

Mining Educational Data to Support Students' Major Selection

Kunyanuth Kularbphettong, Cholticha Tongsiri

***Abstract*—**This paper aims to create the model for student in choosing an emphasized track of student majoring in computer science at Suan Sunandha Rajabhat University. The objective of this research is to develop the suggested system using data mining technique to analyze knowledge and conduct decision rules. Such relationships can be used to demonstrate the reasonableness of student choosing a track as well as to support his/her decision and the system is verified by experts in the field. The sampling is from student of computer science based on the system and the questionnaire to see the satisfaction. The system result is found to be satisfactory by both experts and student as well.

***Keywords*—**Data mining technique, the decision support system, knowledge and decision rules.

I. Introduction

CURRENTLY, selection of student's major is very important because it has a direct effect on their career path and may affect to student's life such as time consuming to study in unsuitable field or not finish study and so on. Computer Science program is one of the significant graduate programs in Thailand. However, the problem encountered by computer science students is an error in the fields of study. Most students do not know the exact needs and their skills in the fields of study.

Therefore, to select appropriated fields, the use of student information, like registration information, course information and class learning courses, was evaluated and data mining techniques was used to analyze pattern and relationship of information. Data mining is a process to create knowledge from transactional database by using statistic procedure and machine learning and training set to get the exact information for predicted decision.

This research aims to develop a decision support system which will be useful to directly affect both the student in choosing the appropriate fields on the abilities and interests of students. Moreover, it is also beneficial to improve the curriculum and resources, both personnel and equipment used in teaching to achieve maximum performance.

The remainder of this paper is organized as follows. Section II presents related works and research methodologies used in this work. Section III presents the experimental setup based on the purposed model based on data mining technique and Section IV shows the results of this experiment. Finally, in Section IV conclude the paper with future research.

Kunyanuth Kularbphettong is with Computer Science Program, Suan Sunandha Rajabhet Univerity, Bangkok, Thailand (phone: 662-150-1169; e-mail: kunyanuth.ku@ssru.ac.th).

Cholticha Tongsiri is Faculty of Information and Communication Technology, Silpakorn University, Bangkok, Thailand (phone: 662-233-4995; e-mail: c_tongsiri@hotmail.com).

II. Related Works and the Methodologies

In this section, we illustrate related works and specified methodologies used in this project.

A. Relates Works

A literature search shows that most of the related researches have deployed data mining techniques to analyze educational data by following this: According to C. Romero et al. [1], the research was shown the usefulness of the data mining techniques in course management system and the rules can help to classify students and to detect the sources of any incongruous values received from student activities. C. Marquez et al. [2] applied data mining techniques to predict school failure with 670 middle-school students and using 10 classification algorithms and 10 fold-cross validation was evaluated the project. K. Kularbphettong used data mining techniques with data log file provided by Learning Management Systems (LMSs) in relation to visits and times, resources viewed, assessments, activities and etc [3]-[5]. B. Minaei-Bidgoli et al. [6] presented an approach to classify students in order to predict their final grade based on features extracted from logged data in an education web-based system.

B. The Methodologies

Data Mining is the data analyzing technique from different perspectives also summarizing the useful information results. The data mining process uses many principles as machine learning, statistics and visualization techniques to discover and present knowledge in an easily comprehensible form. There is another definition as "the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data" [7], [8].

A decision tree is one of the most well known classification approaches that are commonly used to examine data and induce the tree in order to make predictions [9]. The purpose of the decision tree is to classify data into distinct groups or branches that generate the strongest separation in the values of the dependent variable [10].

J48 is an open source Java implementation of the C4.5 algorithm under WEKA data mining platform. C4.5 is an algorithm used to generate a decision tree developed by Ross Quinlan. C4.5 is a software extension and thus improvement of the basic ID3 algorithm designed by Quinlan [11]. J48 uses gain ratio to classify the decision tree.

Bayesian Networks algorithms are structured, graphical models of probabilistic relationships among a set of random

variables. The conditional independence assumptions in the graph are estimated by statistical and computational models [12], [13].

III. Experimental Setup

The data of this experiment was collected from the computer science program, Suan Sunandha Rajabhat University, during the period of 2006-2012. The student data set was composed of 312 personal records, registered course records and students 'quizzes in computer skills. In the gathering data phrase, Computer Science students were subjected to take quizzes in computer skills. There were four quizzes in Database System, Software Engineering, Multimedia, and Network and Communication fields. As presented in Fig. 1, it showed the process of this research to generate the student's academic performance model by using data mining techniques and used PHP to develop a decision support system based on the model.

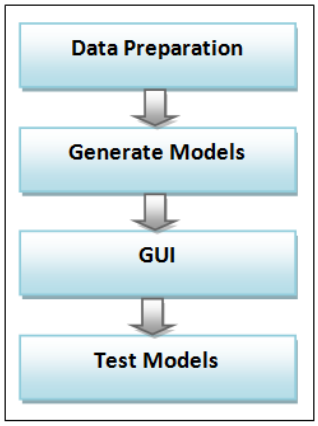

Fig. 1 The process of the student's academic performance model

Also, the results of students' quizzes are collected in the last. The data is preprocessed, and transformed to be appropriated format so as to apply data mining techniques to generate model. The equal width method was used to partition the value of continuous attributes into five nominal values: VERY POOR, POOR, FAIR, GOOD and VERY GOOD.

After preparation phrase as shown in Fig. 2, Data was analyzed by WEKA. WEKA, the Waikato Environment for Knowledge Analysis, is a collection of machine learning algorithm to analyze data set for data mining tasks [14].

J48 and Bayesian Networks algorithms were used in this project to estimate and evaluate for creating the model. To take measure the result, the K-fold cross validation method was provided to validate the result. And to create the effectiveness model, the results of each algorithm were evaluated by the percentage correct, precision, recall and F-measure.

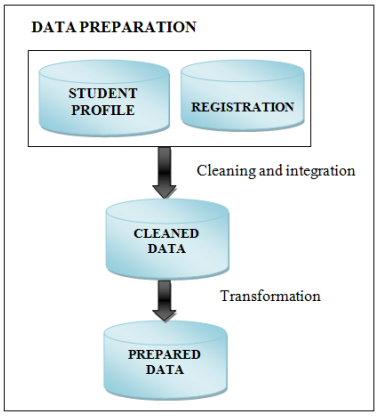

Fig. 2 The data preparation process

IV. Results

In this research, the data was analyzed by using J48 and Bayesian Networks algorithms and using the appropriated result was created the model to develop the decision support system to guide student major selection. The results of this experiment were shown in Table I, compared the effectiveness of each algorithm.

TABLE I
The Experiment Results of Each Algorithm

	percentage of collect	Precision	Recall	F-measure
J48	89.23	0.84	0.92	0.88
Bayesian Networks	92.13	0.93	0.9	0.91

The results of the experiment show that the Bayesian Networks algorithm is more efficient than J48 algorithm. The percentage of prediction is accurate 92.13% and precision and F-measure are 0.93 and 0.91 respectively.

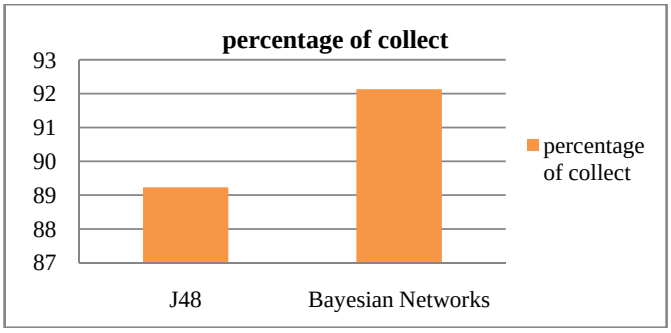

Fig. 3 The results of the percentage of collect

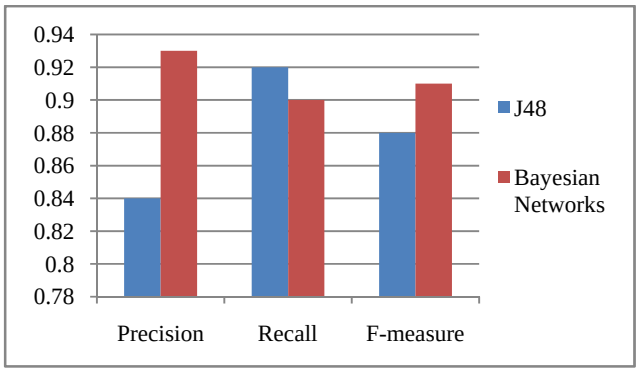

Fig. 4 The results of J48 and Bayesian Networks

For this reason, the Bayesian Networks was used to create model of this project. Fig. 5 was presented the main web page of the prototype of decision support system to select student major.

Fig. 5 The Decision Support System

To evaluate the performance of the system, the 5 expert respondents and 30 students were used to test this system. The results of the accuracy and satisfaction from the respondents were affected to satisfaction levels.

V. Conclusion and Future Work

In this paper, we presents the preliminary result showing a promising progress in this prototypes model for the ongoing improvement project and also this model can be beneficial to help student for selecting his/her major. However, in term of the future experiments, we are looking forward to research about other data mining techniques to enhance this project and also apply the tool to support student's decision.

Acknowledgment

The authors gratefully acknowledge the financial subsidy provided by Suan Sunandha Rajabhat University.

References

[1] C. Romero, S. Ventura, E. García, "*Data mining in course management systems: Moodle case study and tutorial*" Computers & Education, Volume 51, Issue 1, August 2008, pp. 368–384.

[2] R. Agrawal, T. Imielinski, and A.N. Swami, A. N., "*Mining association rules between sets of items in large databases*", In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, pp. 207-216,1993.

[3] K. Kularbphettong, and C.Tongsiri."*Student Motivation Behavior on e-Learning based on Data Mining Techniques*".Proceeding of International Conference on Data Analysis and Decision Making. Prague, Czech Republic July 08-09, 2013.

[4] K. Kularbphettong, and C. Tongsiri. "*Mining Educational Data to Analyze the Student Motivation Behavior*" .Proceeding of International Conference on Information Technology and Computer Science. Paris, France Augest 22-23, 2012.

[5] K. Kularbphettong, C. Tongsiri and P. Waraporn. "Analysis *of Student Motivation Behavior on e-Learning Based on Association rule mining*".Proceeding of International Conference on Education and Information Technology. Paris, France July 27-28, 2012.

[6] B. Minaei-Bidgoli, D. A. Kashy, G. Kortemeyer and, W. F. Punch."*Predicting student performance: an application of data mining methods with the educational web-based system LON-CAPA*" In Proceedings of ASEE/IEEE Frontiers in Education Conference, Boulder, CO: IEEE, 2003.

[7] U. M. Fayyad, G. Pitatesky-Shapiro, P. Smyth, and R. Uthurasamy, *"Advances in Knowledge Discovery and Data Mining"*, AAAI/MIT Press, 1996.

[8] W. Frawley, G. Piatetsky-Shapiro, and C. Matheus, "*Knowledge Discovery in Databases: An Overview*". AI Magazine, Fall 1992, pp. 213-228.

[9] H. Edelstein, "Introduction to Data Mining and Knowledge Discovery", Third Edition. Two Crows Corporation, Potomac, MD, USA, 1999.

[10] Parr Rud, O. "Data Mining Cookbook. Modeling Data for Marketing, Risk, and Customer Relationship Management". John Wiley & Sons, Inc.; 2001.

[11] A. Kumar Sharma and S. Suruchi, "A Comparative Study of Classification Algorithms for Spam Email Data Analysis", International Journal on Computer Science and Engineering (IJCSE), Vol. 3 No. 5, pp. 1891-1895. May 2011.

[12] Kevin B.Korb and Ann E. Nicholson. "Introduction to Bayesian Networks", Bayesian Artificial Intelligence. 2010, pp 29-54. CRC Press.

[13] M. Singh and M. Valtorta. Construction of Bayesian Network Structures from Data: a Brief Survey and an efficient Algorithm. International Journal of Approximate Reasoning, 1995, volume 12, pages 111-131, Elsevier Science Inc.

[14] WEKA Source: http://www.cs.waikato.ac.nz/ml/weka/

Kunyanuth Kularbphettong received the B.S. degree in Computer Business, M.S. degree in Computer Science, and Ph.D degree in Information Technology. Her current research interests are in Multi-agent System, Web Services, Semantic Web Services, Ontology and Data mining techniques.

Cholticha Tongsiri is Faculty of Information and Communication Technology, Silpakorn University, Thailand (phone: 662-233-4995; e-mail: c_tongsiri@ hotmail.com)