

DEVELOPING AN INTELLIGENT RECOMMENDATION SYSTEM FOR A PRIVATE UNIVERSITY IN THAILAND

Kanokwan Kongsakun, Murdoch University, kokoya002@yahoo.com
Chun Che Fung, Murdoch University, L.fung@murdoch.edu.au
Tuchtawan chanakul, Hatyai University, tawan_hu@hotmail.com

ABSTRACT

In Thailand, choosing a program of study for tertiary students is significant due to the associated future job opportunities. Many students have enrolled in course majors without receiving counseling or advices from appropriate authorities or university services. This could have potential mismatch between students' aptitude, personal interest and capability, and the particular course being taken up. This may lead to low retention rate and failures. In order to improve and support the academic management processes, many universities in Thailand are developing innovative information systems and services with an aim to enhance efficiency and student relationship. Some of these initiatives are in the form of a Student Recommendation System (SRM). In Thailand, this university service is normally provided by professional counselors or advisors who have many years of experience within the organization or in the higher education sector. However, the success or appropriateness of such advice is entirely depending on expertise of the counselor and it is entirely human-driven. In addition, the process is also tedious and time consuming. This paper reports a study on an investigation of possible correlation between student historic data and their final results. Clustering techniques have been used with the aim to find structures and relationship within the data. Results from two clustering methods, k-means and TwoStep methods have been compared. This paper describes the development of the experiments, and the proposed Intelligent Recommendation System framework.

Keywords: Data Mining, Data Preprocessing, Clustering, classification, Student Relationship Management, Recommendation System, Management Information System.

INTRODUCTION AND BACKGROUND

With the increasing competition among higher education institutes in Thailand, most institutes are focusing on how to increase student retention rates and the number of completions. In addition, a

university's performance is also increasingly being used to measure its quality and reputation [1]. One form of service which is normally provided by all universities is *counseling*. Archer and Cooper [2] stated that the provision of counseling services is an important factor contributing to students' academic success. In addition, Urata and Takano [3] stated that the essence of the student counseling should include advices on career guidance, identification of learning strategies, handling of inter-personal relation, along with self understanding of the mind and body. It can be said that a key aspect of student services is to provide course guidance as this will assist the students in their selection and future university experience.

However, on the other hand, many students have chosen particular courses of study just because of perceived job opportunities. Issues may arise if a student is not interested in the course, or if the course or career is not suitably matched with the student's capability[4]. In Thailand's tertiary education sector, teaching staff may have insufficient time to counsel the students due to workload and there are inadequate tools to support them. Hence, it is desirable that some forms of intelligent recommendation tools could be developed to assist students in their enrolment. This forms the motivation of this research.

Some of the initiatives for helping the students and staff are in the form of a Student Recommendation System. Such system could be used to provide course advice and counseling for freshmen in order to achieve a better match between the student's ability and success in course completion. In the case of Thai universities, this service is normally provided by counselors or advisors who have many years of experience within the organization. However, with the increasing number of students and expanded number of choices, the workload on these human advisors is becoming too much to handle. It becomes apparent that some forms of intelligent system will be useful in assisting the advisors.

This paper is structured as follows. Section II describes Student Relationship Management (SRM)

in general. Section III and IV student Relationship Management and current solutions respectively. Section VI discusses various Data Mining (DM) techniques which could be used in the Intelligent Recommendation Systems, and next section then focuses on the proposed framework, which presents the main idea and the research methodology. Next section describes the experiments and the results. This paper concludes with discussions on the work to be undertaken and future development.

STUDENT RELATIONSHIP MANAGEMENT IN UNIVERSITIES

According to literature, the problem of low student retention in higher education could be attributed to low student satisfaction, student transfer and drop-outs [5]. This issue leads to lost of students and revenue, and increasing cost of replacement. On the other hand, it was found that the quality and convenience of support services are other factors that influence students to change educational institutes [6]. Consequently, the concept of SRM has been implemented in various universities so as to assist the improvement of the quality of learning processes and student activities.

Definitions of SRM have been adopted from CRM which focuses on the customers and are aimed to establish effective competition and new strategies in order to improve the performance of a firm [7]. In the case of SRM, the context is within the education sector. Although there have been many research focusing on CRM, few research studies have focused on SRM. In addition, the technological supports are inadequate to sustain SRM in universities. For instance, a SRM system's architecture has been proposed so as to support the SRM concepts and techniques that assist the university's Business Intelligent System [8]. This project provided a tool to aid the tertiary students in their decision-making process. The SRM strategy also provided the institution with SRM practices, including the planned activities to be developed for the students, as well other relevant participants. The study verified that the technological support to the SRM concepts and practices were insufficient at the time of writing [8].

In the context of educational institutes, the students may be considered having a role as "customers", and the objective of Student Relationship Management is to increase their satisfaction and loyalty for the benefits of the institute. SRM may be defined under a similar view as CRM and aims at developing and maintaining a close relationship between the institute

and the students by supporting the management processes and monitoring the students' academic activities and behaviors. Piedade and Santos [8] explained that SRM involves the identification of performance indicators and behavioral patterns that characterize the students and different situations under which they are supervised. In addition, the concept of SRM is "*understood as a process based on the student acquired knowledge, whose main purpose is to keep a close and effective students institution relationship through the closely monitoring of their academic activities along their academic path*". SRM can be utilized as an important means to support and enhance a student's satisfaction [9].

Understanding the needs of the students is essential for their satisfaction. It is necessary to prepare strategies in both teaching and related services to support Student Relationship Management. The significance of developing a strong relationship between university and student is a long term strategy. Hence, this paper proposed an innovative information system to assist students in universities in order to support the SRM concept.

CURRENT SOLUTIONS

An understanding of the existing information systems can assist student management, student services and market operation for the participants. It is important to develop strategy to maintain and enhance student satisfaction which is the key role in SRM. Among the researches and papers concerning SRM, there are a number of proposals on solutions for the university management and students. Some examples of such proposals are as shown below.

Piedade and Santos [8] proposed a SRM architecture which supports the SRM concepts and techniques that assist the university's Business Intelligent System. Moreover, the project provided a tool to aid the tertiary students in the decision-making process. The SRM strategy also provides the institution with SRM practices, including planned activities to be developed for the students, as well as other relevant participants. The study found that the technological support to the SRM concepts and practices were insufficient at the time of writing.

Another solution focused on the provision of counselling and careers services, which has been adopted by many universities. To enhance the university's mission, the prominent services provided by universities are counselling, careers and work-placement advice and financial assistance. In

addition, feedback from students is vital in the process.

In business, focusing on customer retention is significant [10, 11]. In terms of education systems, Ackerman and Schibrowsky [12] have applied the concept of business relationships and proposed a business relationship marketing framework. The framework provided a different view on retention strategies and an economic justification on the need for implementing retention programs. The prominent result is the improvement of graduation rates by 65% by simply retaining one additional student out of every ten. The researcher added that this framework is appropriate both on the issues of places on quality of services. Although some problems could not be solved directly, it is recognised that Information and communication technologies can be used and contributes towards maintaining a stronger relationship with students in the educational systems [8].

INTELLIGENT RECOMMENDATION SYSTEM USING DATA MINING TECHNIQUES

Data mining is a process of finding models and relationships between data, and it is used to classify and analyze large databases [13]. Digital and electronic systems have changed the traditional user behaviors. For example, e-commerce has become an essential process in today's businesses [14] and e-mail has an important role in communication. Moreover, e-learning is now a prominent element in learning. Similarly, the Intelligent Recommendation System may become a significant supporting system for universities. There are some previous studies focusing on recommendation systems, which used several data mining techniques as shown below.

Clustering Techniques

Clustering Techniques such as K-Means and the Two-Step Algorithm are commonly used to group the data [15]. Clustering focuses on identifying groups of similar records and labeling the records according to the group to which they belong. The process of clustering has been used to describe target customer's behavior in marketing and CRM [16]. In this research, the two clustering methods have been used to classify the target students' performance into subgroups based on historic data. In this paper, the objective of clustering is to identify students' performance based on history records from the university. The set of data items from the clustering

process will be applied in this process. Firstly, K-means algorithm aims to discover k number of clusters. K-means algorithm finds a partition such that the squared error between the empirical mean of a cluster and the points in the cluster is minimized. Finally, the TwoStep cluster method is a scalable cluster analysis algorithm, which can handle both continuous and categorical variables or attributes.

Association Rule

Association Rule is used to seek an existence of the most commonly sought patterns [17]. It is used to establish the relationships or associations between values of categorical variables in large data sets [18]. For instance, Liao, Zou and Chang proposed an association rules and sequential rules based recommendation system in [16]. In this study, association rule will be used to examine the relationships between the clustered subgroups [16, 19]. This process will be used to identify the student features that appear most often. The coverage of Association Rule is the number of instances that it predicts correctly [15]. The most common output will be analyzed and compared with the clustering output to ensure more accurate predictions.

Fuzzy Logic and Rough Set

The concept of rough set is a mathematical approach in order to represent and process vagueness and uncertainty in data analysis and it has been applied to various applications. In addition, it has produced significant results in machine learning, knowledge discovery, decision analysis, expert systems and other applications [20]. Another application was focused on an intelligent recommendation in virtual mall for products selection. The study was based on machine learning using rough set and fuzzy set. [14]

PROPOSED ARCHITECTURE OF RECOMMENDATION SYSTEM

To serve the students, an Intelligent Recommendation System plays the important role to support students in various ways. There have been several solutions supporting SRM for university students, however, no system has focused on the recommendation system for students using historical data. A recommendation system could apply statistical, artificial intelligence and data mining techniques to the problem of making appropriate recommendation for the students.

Figure 1 illustrates the proposed recommendation system architecture. The proposal aims to analyse student background such as school at which the

student studied previously, high school results and GPA in the university from the enterprise database. The result will then be used to compare the profile of the new students. The recommendation system will aim to provide suggestions for the most appropriate courses and subjects for students based on past historical records.

both continuous and categorical variables or attributes, and it requires only one data pass.

Prediction Model

The prediction model is an integrated process which evaluates the outputs from the previous stage and applies a weighted formula to determine the final recommendation. The weights will be determined or learnt from previous historic data and subject to human intervention based on the experience of the counselor or advisor.

Application program

The new intelligent model will be tested through an online system of a private university in Thailand. The students can use the application while choosing the course on the enrolment process. The online recommendation system will suggest appropriate information decision such as courses, schedules and activities to support student’s decision. Finally, students are able to make the final decision as regard to the course of study.

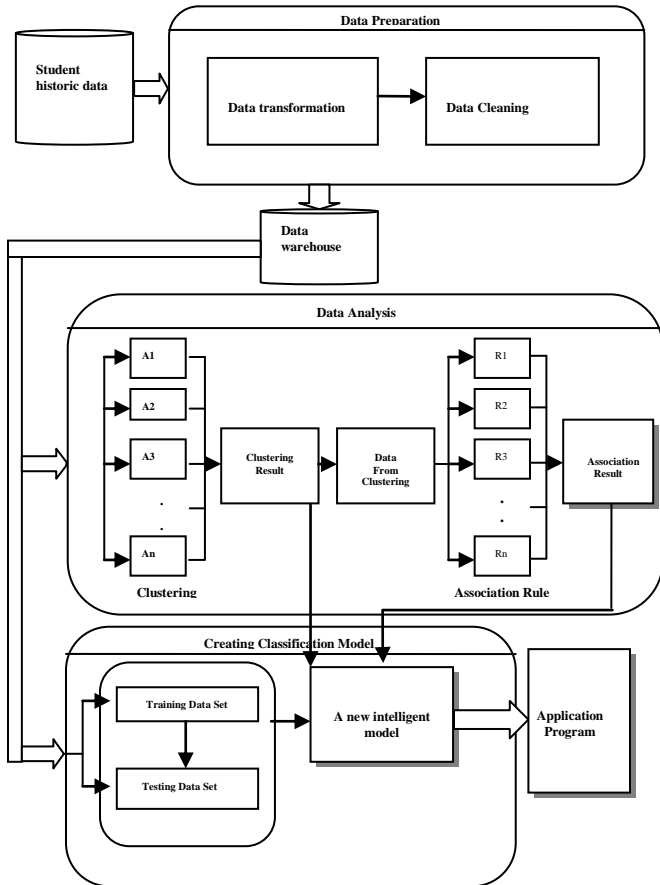


Figure 1. Proposed Recommendation System Framework to Support Student Relationship Management (SRM)

Data Preparation

After data collection, the data is re-formatted in the stage of data transformation. In the data cleaning process, the parameters used in the data analysis is identified and the missing data is handled [21] that permit null values.

Data Analysis

The k-means method, a clustering method, is used to group records based on similarity of values for a set of data, whereas The TwoStep cluster method, is designed to handle very large data sets. The TwoStep clustering algorithm involves two steps, pre-clustering and clustering. It can handle

EXPERIMENT

The data preparation and selection process involves a dataset of 3,500 student records from five academic years. All the student data have included records from the first year to graduation. Due to privacy issue, the data in this study do not involve any personal information and no student is identified in the research. The university has randomized the data and all private information have been removed. After data collection, the data has been re-formatted in the stage of data transformation. In the cleaning process, the parameters used in the data analysis has been selected and the missing or incomplete data records have been removed [22]. The student records have been separated into five cases, and each case has been separated into training and testing data sets as shown in the table below.

Table 1. Examples of training sample dataset

Randomized Student ID	GPA from previous study	Types of previous school	Awards from pervious school	Talent and Interest	Major(in university)	GPA In University
4800	2.35	High School	0.2	1	Accounting	3.75
4801	3.55	Technical & Technology college	0.3	4	Accounting	3.05
5001	2.55	Other	0.9	3	Information Technology	2.09
5002	2.75	University(re-study)	0.4	5	Accounting	2.58
5003	3.00	Community college	0.2	7	Politic	2.77
5101	2.00	Other	0.1	2	Business Computer	2.11

The table above shows the randomized student ID, GPA from previous study, types of previous school, awards, talent and interest majors in university and result in the university.

RESULTS

In the experiment, the results have shown that Two Step clustering technique has a maximum of 15 clusters and a minimum of 2 clusters. Table 2 shows that seven clusters have been used in order to compare with the results from the K-means method.

Table 2. Number of cluster and iteration used by the K-means and TwoStep Clustering Techniques

	K-means clustering	TwoStep Clustering
Number of clusters	7	7
Number of Iterations	13	-

Results in Table 3 show the percentage of records which have been allocated to each cluster. The differences between the two methods are also shown in the same table as follows.

Table 3. Results of K-means and TwoStep Clustering

	K-means clustering (%)	TwoStep Clustering (%)
Cluster 1	23.000	6.457
Cluster 2	7.486	4.914
Cluster 3	5.514	10.286

Cluster 4	5.314	6.914
Cluster 5	8.600	8.114
Cluster 6	11.229	8.571
Cluster 7	2.143	9.457

The values give a measurement of the quality of the results from these two methods as they indicate the average cluster distance within each cluster.

CONCLUSIONS

The article describes a proposal on a recommendation system in support of SRM and to address issues related to the problem of course advice or counseling for university students. The recent work is focusing on the development and implementation of each process in the framework. The first experiment has been tested using clustering techniques. The prediction model will be refined, and the recommendation system will provide a useful service for students supporting Student Relationship Management strategy in Thai private universities.

REFERENCES

1. Jusoff, K., S.A.A. Samah, and P.M. Isa. *Promoting university community's creative citizenry.* in *Proceedings of world academy of science.* 2008: Engineering and technology.
2. Archer, J.J. and S. Cooper, *Counselling and Mental Health Services on Campus,* in *A handbook of Contemporary Practices and Challenges* Jossy-Bass, Editor. 1998: San Francisco, CA.
3. Uruta, U. and A. Takano, *Between psychology and college of education.* *Journal of Educational Psychology,* 2003. **51**: p. 205-217.
4. Gamage, D.T., et al., *The impact of quality assurance measures on student services at the Japanese and Thai private universities.* *Quality Assurance in Education,* 2008. **16**(2): p. 181-198.
5. Caison, A.L. *Determinates of Systemic Retention: Implications for improving retention practice in higher education.* *Journal of College Student Retention* 2004-2005 [cited 6 4]; 425-441].

6. Helland, P., H.J. Stallings, and J.M. Braxton, *The fulfillment of expectations for college and student departure decisions*. Journal of College Student Retention, 2001-2002. **3**(4): p. 381-396.
7. Gao, Y. and C. Zhang. *Research on Customer Relationship Management Application System of Manufacturing Enterprises*. in *Wireless Communications, Networking and Mobile computing, 2008 Wicom'08.4th International conference*. 2008. Dalian.
8. Piedade, M.B. and M.Y. Santos, *Student Relationship Management: Concept, Practice and Technological Support*. IEEE Xplore, 2008: p. 2-5.
9. Harej, K. and R.V. Horvat, *Customer Relationship Management Momentum for Business Improvement*. Information Technology Interfaces(ITI), 2004: p. 107-111.
10. Verhoef, P.C., *Understanding the effect of Customer Relationship Management Efforts on Customer Retention and Customer Share Development*. Journal of Marketing, 2003. **67**: p. 30-45.
11. Bolton, R.N., P.K. Kannan, and M.D. Bramlett, *Implications of Loyalty Program Membership and Service Experiences for Customer Retention and Value*. Journal of the Academy of Marketing Science, 2000. **17**: p. 45-65.
12. Ackerman, R. and J. Schibrowsky, *A Business Marketing Strategy Applied to Student Retention: A higher Education Initiative*. Journal of College Student Retention, 2007-2008. **9**(3): p. 330-336.
13. Fayyad, U., G. Piatetsky-Shapiri, and P. Smith, *From Data Mining to Knowledge Discovery in Databases*, in *AAAI*. 1997. p. 37-54.
14. Yang, H.W., et al. *Machine Learning based Intelligent Recommendation in Virtual Mall*. in *Proceeding of the third international conference on Machine Learning and Cybernetics*. 2004. Shanghai.
15. Witten, I.H., *Data mining: Practical Machine Learning Tools and techniques*, ed. 2nd. 2005, United state of America: Elsevier.
16. Liao, S., T. Zou, and H. Chang, *An Association Rules and Sequential Rules Based Recommendation System*, in *Wireless Communications, Networking and Mobile Computing, 2008*. 2008. p. 1-4.
17. R.Agrawal, T. Imielinski, and A.Swami. *Mining association rules between sets of items in large databases*. in *ACM SIGMOD International Conference on Management of Data*. 1993. Washington, D.C.
18. Cios, K.J., et al., *Data Mining: A knowledge Discovery Approach*. 2007: Springer Science and Business Media. 600.
19. Anand, A. and D. Khots. *A data mining framework for indentifying craim overpayments for the health insurance industry*. in *The 3rd INFORMS Workshop on Data Mining and Health Informatics 2008*. 2008.
20. Pawlak, Z., ed. *Rough Sets and Data Mining analysis of imprecise data*. Rough Sets, ed. T.Y. Lin and N. Cercone. 1997, Kluwer Academic group: Dordrecht, The Natherlands. 436.
21. K.W.Wong, C.C.Fung, and T.D.Gedeon. *Data Mining Using Neural Fuzzy for Student Relationship Management*. in *International Conference of Soft Computing and Intelligent Systems*. 2002. Tsukuba, Japan.
22. Wong, K.W., et al., *Intelligent Data Mining and Personalisation for Customer Relationship Management*, in *International Conference on Control, Automation, Robotics and Vision Kunming 8th*. 2004: China. p. 1796-1801.