Improving Student Academic Performance by An Application of Data Mining Techniques

Sajadin Sembiring[1*], Abdullah Embong[2], Mohd. Azwan Mohamad[3], Muhammad Furqan[4.]

[1]Departemen Teknik Informatika Sekolah Tinggi Teknik Harapan Medan, Indonesia
[2,3,4,] Faculty of Computer System & Software Engineering
University Malaysia Pahang, Kuantan, Pahang Darul Makmur, Malaysia
[1]sajadinbiring@gmail.com, [2,3]{ae,azwan}@ump.edu.my, [4] mhdfurqan@gmail.com

Abstract

Data mining is a technology used in different disciplines to search for significant relationships among variables in large data sets. Data mining is mainly used in commercial applications.

Educational data mining concerns with developing methods for discovering knowledge from data, that comes from educational environment. In this study we concentrate on application of data mining techniques to analyze the relationships between student's behavior and their success. This is done by using Smooth Support Vector Machine (SSVM) classification and kernel k-means clustering algorithm techniques.

The results can be used both to help teachers understand their student's learning pattern and to enhance student performance.

Keywords: DM, SSVM, kernel-k-means, Student Performance.

1. Introduction

The amount of the data in educational environment maintained in electronic format has seen a dramatic increase in recent time. The data can be collected from historical and operational data reside in the databases of educational institutes. The search to determine the relationships among variables in the data has become a slow and subjective process. As a possible solution to this problem, the concept of *Knowledge Discovery in Databases-KDD* has emerged. The process of the formation of significant models and assessment within KDD is referred to as Data Mining [2].

There are increasing research interests in education field using data mining. Application of Data mining techniques concerns to develop the methods that discover knowledge from data and used to uncover hidden or unknown information that is not apparent, but potentially useful [1]. The discovered knowledge can be used to better understand students' behavior, to assist instructors, to improve teaching, to evaluate and improve e-learning system, to improve student academic performance; to improve curriculums and many others benefits [5].

This study investigates the educational domain of data mining using a case study from data bases class. It showed what kind of data could be collected, how could we preprocess the data, how to apply data mining on the data, and finally how can we benefited from the discovered knowledge. There are many kind of knowledge can be discovered from data.

In this study, university students were predicated his/her final grade by using SSVM classification and grouped according to their similar characteristics, forming clusters. The clustering process was carried out using Kernel K-Means algorithm technique.

The rest of the paper is organized as follows: Section 2 summaries related work in an application of data mining techniques in educational environments. Section 3 describes the data mining task on educational system and brief review the methods of SSVM classification and kernel k-means clustering. Section 4 gives a general description of data we used in our case study and describes the preprocess stage and transformation of the used data. Section 5 reports our experiments about applying data mining techniques on the educational data. Finally, section 6 we conclude this paper with a summary and section 7 describes an outlook for future work.

2. Related Work

Data mining in higher education is a new emerging field, called Educational Data Mining [4],[5]. There are many works in this area, because of its potentials to educational institutes. Romero.C and Ventura.S, have a survey on educational data mining between 1995 to 2005[5]. They conclude that educational Data mining is promising area of research and it has a specific requirements not presented in other domains. Merceron.A and Yacep.K, gave a case study that used data mining to identify behavior of failing students to warn students at risk before final exam [4]. Also, Data mining in educational area used by Beikzadeh,M and Delavari N, to identify and then enhance educational process in higher educational system., which can improve their decision making process [7]. Finally, Waiyamai, K. used data mining to assist in development of new curricula and to help engineering students to select an appropriate major [3].

3. Data mining Task in Educational System

Today, data collecting and storing are no longer expensive and difficult task. As a result, datasets are growing explosively. To extract the knowledge and information from these massive dataset has attracted a great deal of scientific attention and has become an important research area [2],[6]. Data mining is a flourishing research field and has become a synonym for the process of extracting hidden and useful information from datasets.[10]
Data mining used advanced techniques to discover pattern from data. The data mining task are the kinds of patterns that can be mined. In this study we focused to use both Smooth Support Vector Machine (SSVM) classification and kernel k-means clustering algorithms. In the following sections describes the brief review and the results of applying data mining techniques to the data of our case study for each of algorithms.

3.1. Multiclass SSVM for Classification

Classification is data mining task that predicts group memberships for data instances [9]. In educational area application of the classification method, given works of a student, one may predicate his/her final grade.
The SSVM is further development of Support Vector Machine (SVM) [9,],[11][17]. The SSVM generated and solve an unconstrained smooth reformulation of the SVM for pattern classification using completely arbitrary kernel [11]. SSVM is solved by a very fast Newton-Armijo algorithm and has been extended to non linear separation surfaces by using non linear kernel techniques. The numerical results show that SSVM is faster than other methods and has better generalization ability [9].
The fast Newton-Armijo Algorithm for SSVM [9],[11],[17] is described in the C/C++ style pseudo-code.

Start with any $(w^0,\gamma^0) \in R^{n+1}$.
Having (w^i,γ^i), *stop if the gradient objective function*

$$\min_{w,\gamma} \quad \frac{\nu}{2}\left\|\left(e-D(Aw-e\gamma)\right)_+\right\|_2^2+\frac{1}{2}(w'w+\gamma^2) \quad \textit{is zero}$$

That is $\nabla\Phi_\alpha\left(w^i,\gamma^i\right)=0$. *Else compute* $\left(w^{i+1},\gamma^{i+1}\right)$ *as follows:*

i). ***Newton Direction*** *: Determine direction* $d^i \in R^{n+1}$ *by setting equal to zero the linearization of* $\nabla\Phi_\alpha\left(w\ ,\gamma\ \right)$ *around* $\left(w^i,\gamma^i\right)$ *which gives n +1 linear equations in n+1 variables :*

$$\nabla^2\Phi_\alpha\left(w^i,\gamma^i\right)d^i=-\nabla\Phi_\alpha\left(w^i,\gamma^i\right)'$$

ii). ***Armijo Stepsize*** *: Choose a step size* $\lambda_i \in R$ *such that :*

$$\left(w^{i+1},\gamma^{i+1}\right)=\left(w^i,\gamma^i\right)+\lambda_i d^i$$

Where $\lambda_i=\max\left\{1,\frac{1}{2},\frac{1}{4},\ldots\right\}$ *such that:*

$$\Phi_\alpha\left(w^i,\gamma^i\right)-\Phi_\alpha\left(\left(w^i,\gamma^i\right)+\lambda_i d^i\right)\geq-\delta\lambda_i\nabla\Phi_\alpha\left(w^i,\gamma^i\right)d^i$$

Where $\delta\in\left(0,\frac{1}{2}\right)$.

Basically, the theory of SVM classification for two class or binary classification [16]. A multi-class classification system can be obtained by combining two class SSVM. Usually there are two schemes for this purpose. One is the one versus others or the one-against-all strategy to classify between each class and all remaining; the other is one versus one or the one –against-one strategy to classify between each pair.
In this study we used one versus others strategy caused by the training effort side the results is better than one versus one strategy.

3.2. An Effective Kernel K-Means for Clustering

Clustering is finding groups of objects such that the objects in one group will be similar to one another and different the objects in another group [8]. In educational area, clustering will be used to grouping students according to their behavior and performance.
In this study we used Kernel K-means algorithm to cluster the given data.
A drawback to original K-means is that it cannot separate cluster that are non-linearly separable input space. Kernel K-Means is one approaches have emerged for tackling such a problem. Kernel K-means, where, before clustering, points mapped to a higher dimensional feature space using a non-linear function, and then Kernel K-means partitions the points by linear separator in new space[13][14].
Kernel K-means has been extended to efficient and effective large scale clustering [8], since the original Kernel K-means had serious problems, such as the high clustering cost due to the repeated calculations of kernel values, or insufficient memory to store the kernel matrix, that make it unsuitable for large corpora. The new clustering scheme is a large scale clustering for Kernel K-means algorithm [8].

1. *Assign* $\delta(x_i, C_k)$ *(*$1 \le i \le N$, $1 \le k \le K$*) with initial value, forming K initial cluster* $C_1, C_2, \ldots, C_K$
2. *For each cluster* C_k*, compute* $|C_k|$ *and* $g(C_k)$*.*

$$IC_kI = \sum_{i=1}^{N} \delta(x_i, C_k)$$
$$g(C_k) = \frac{1}{|C_k|^2} \sum_{j=1}^{N} \sum_{i=1}^{N} \delta(u_j, C_k)\delta(u_i, C_k)H(x_j, x_i)$$

3. *For each training sample* x_i *and cluster* C_k*, compute* $f(x_i, C_k)$*. And then assign* x_i *to the closest cluster*

$$f(x_i, C_k) = \frac{2}{|C_k|} \sum_{j=1}^{N} \delta(u_j, C_k)H(x_i, x_j)$$

$$\delta(x_i, C_k) = \begin{cases} 1, & f(x_i, C_k) + g(C_k) < f(x_i, C_j) + g(C_j) \\ & \text{for all } j \neq k \\ 0, & otherwise \end{cases}$$

4. *Repeat step 2 and 3 until converge.*
5. *For each cluster* C_k*, select the sample that is closest to the centre as the representative of* C_k*,* $m_k = Arg\ min\ D(\Phi(x_i), z_k)$*.*

4. Data Preparation and Transformation

In our case study we collected the student data from database management system course held at the Universiti Malaysia Pahang (UMP) in third semester of 2007/2008. The number of student was 5568. The sources of collected data were: personal records, academic record of students and course records.
To get better input data for data mining technique, we did some preprocessing for collected data. The data was maintained in different tables was joined in a single table. After we integrated the data into one files, to increase interpretation and comprehensibility, we discretized the attributes to categorical ones. For examples, we grouped all grades into four groups e*xcellent*, *good*, *average*, and *poor*. In this step the fields used in the study were determined and transformed if necessary.

5. Experiment

5.1.Classification

In our case study, by the training effort outside, we used J48 decision tree to represent logical rules of student final grade. The represented tree is large, we generated some of the strong rule in tree are:

if attendance=good,Homework=four,Midterm=good, and Lab=pass
then finalgrade=Excellent
if attendance=good,Homework=three,Midterm=good, and Lab=pass
then finalgrade=Good
if attendance=good,Homework=three,Midterm=average,and Lab=pass
then finalgrade=Average
if attendance=average,Homework=two,Midterm=average,and Lab=fail
then finalgrade=Poor

Based on the logical rules, we implemented multiclass SSVM binary classification. It was trained to distinguish between predicate from final grade (labeled +1) and the others (labeled-1).
Experiment carried out in two steps: training and testing. Training conducted for each grade 80 and 90 of percentage from total number of data set. The testing conducted for each grade using all the data set. The number of value labeled 1 interpreted as the prediction of student final grade.

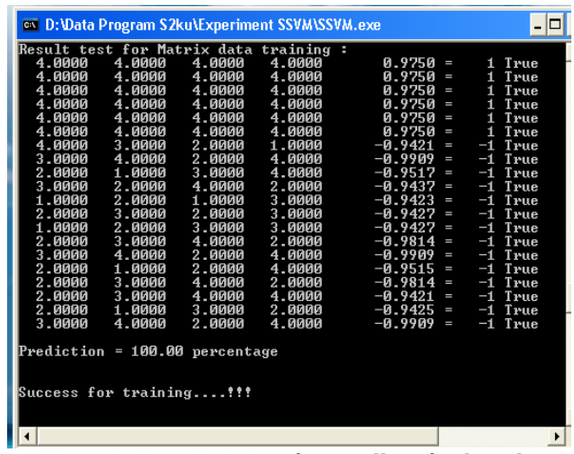

Figure1. Data Training, for excellent final grade

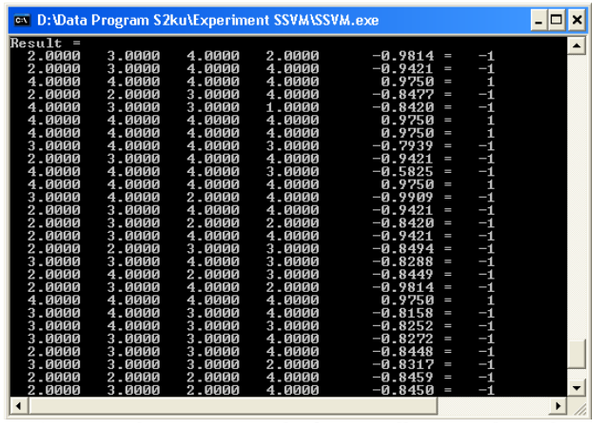

Figure2. The testing result, for Excellent Final Grade

Tabel.1. Final Grade Prediction UMP's student Faculty of computer System and Software engineering third semester of 2007/2008;

N0	Final Grade Prediction	Percentage
1	Excellent	14,70 %
2	Good	35,30 %
3	Average	46,25 %
4	Poor	3,75 %

The benefit of this method is that it can predict the grade of each student on time. For example the teacher can predict *poor* or *average* students before the end of the semester and he/she may work on them to improve their performance before the final..

The different of classification rules and the rules generated from association it is important to know. Association rules are characteristics rules (it describes current situation), but classification rules are prediction rules (it describes future situation).

5.2.Clustering

After the data preparation, the data selection and transformation process was performed. The prepared data was then put through the data mining process. The Kernel K-Means algorithm was used in this step. The number of clusters was determined as an external parameter. Different cluster numbers were tried, and successful partitioning was achieved with 5 clusters. In our case study, the cluster centroids on figure 3 gives mean of each cluster for each attribute. For graphs, the MapToolbox plug-in of MATLAB software was used. The graphs are given in figure 4 as a picture of students group according to their performance. Using these results we can divide students into five groups and guide them according to their behavior.

Attribute	Cluster_1	Cluster_2	Cluster_3	Cluster_4	Cluster_5
Attendence	87.0563	65.0879	39.7925	15.1285	67.9645
gpa	77.6896	70.3137	65.3750	67.8410	71.5248
hours	88.0806	95.6241	64.3750	57.1243	90.8782
midterm	14.7148	12.5206	5.0000	1.6923	13.3454
lab	16.3502	15.3346	5.0000	4.7692	12.9627
final	37.1029	30.7416	14.8750	29.7281	30.2987
grade	77.6682	65.9394	26.8750	68.3721	65.2804

Cluster Centroids.

Figure.3. Clustering Student into five Groups

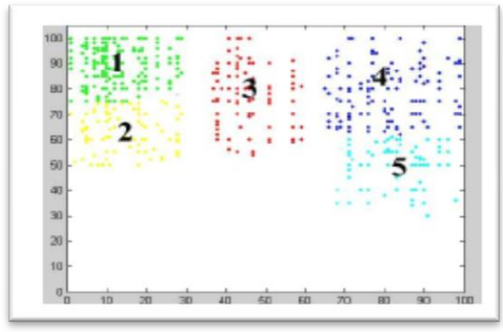

Figure.4.Each Number Represent different Clusters

6. Conclusions and Future Work

The result of this study indicates that Data Mining Techniques (DMT) capabilities provided effective improving tools for student performance. It showed how useful data mining can be in higher education in particularly to improve student performance. We used student database course. We applied data mining techniques to discover knowledge. Particularly we discovered classification rules using decision tree. We applied the rules from decision tree using multi class SSVM to predicate the students' final grade. Also we clustered the student into group using kernel k-means clustering. Each one of these knowledge can be used to improve the performance of student. For future work, application of data mining techniques in educational field can be used to develop performance monitoring and evaluation tools system. DMT has a potential in performance monitoring of High school and other levels education offering historical perspectives of students' performances. The results may both complement and supplement tertiary education performance monitoring and assessment implementations.

7. References

[1]. Kash Barker, Theodore Trafalis, and Teri Reed Rhoads "Learning From Student Data". Proceedings of the 2004 Systems and Information Engineering Design Symposium. Mathew H. Jones, Stephen D. Patek, and Barbara E. Towney eds. 2004. pp79-86

[2]. Han,J. and Kamber,M. "Data mining: Concepts and Techniques", 2nd edition. The Morgan Kaufmann series in Data Management System, Jim Grey, series Editor. 2006.

[3]. Waiyamai,K "Improving Quality Graduate Student by Data Mining". Departement of Computer engineering. Faculty of Engineering. Kasetsart University, Bangkok Thailand. 2003

[4]. Merceron, A and Ycef, K.,"Educational Data mining: A case study". In proceedings of the 12th International Conference on Artificial Intelligence in Education AIED 2005, Amsterdam, The Netherlands, IOS Press.2005

[5]. Romero, C. and Ventura, S.," Educational Data mining: A survey from 1995 to 2005", Expert systems With Application" (33) 135-146. 2007

[6]. Luan, J. "Chapter 2: Data Mining and Its Application in Higher Education. Knowledge Management – Building a Competitive Advantage in Higher Education." Serban, A. & Luan, J. (eds.) Jossey-Bass. 2002.

[7]. Naeimeh Delavari and Mohammad Reza Beikzadeh and Somnuk Phon-Amnuaisuk, "Application of Enhanced Analysis Model for Data Mining Processes in Higher Educational System". ITHET 6th Annual International Conference. Juan Dolio, Dominican Republic. July 7 – 9, 2005,pp F4B-

[8]. Rong Zhang and Alexander I. Rudnicky," A large Scale Clustering Scheme for kernel-K-Means"School of Computer Science, Carnegie Mellon University 5000 Forbes Avenue, Pittsburgh, PA 15213, USA.2006

[9]. Y.J. Lee. And O.L Mangasarian, " A Smooth Support Vector Machine for classification",Journal of Computational Optimization and Applications.20, 2001, pp.5-22

[10].. Ogor Emmanuel. N, " Student Academic Performance: Monitoring and Evaluation Using Data Mining Techniques". Fourth Congress of Electronics, Robotics and Automotive Mechanics. 2007. I EEE Computer Society.

[11].Santi W.P, A.Embong., "Smooth Support Vector Machine for Breast Cancer Classification", IMT-GT Conference on Mathematics, Statistics and Applications(ICMSA), 2008

[12]. Christoper Burges. "A Tutorial on support vector Machines for Pattern Recognition", Data Mining and Knowledge Discovery, 2(2), 1998

[13]. Mark Girolami." Mercer Kernel Based Clustering in Feature Space" I EEE Trans. On Newral Networks.

[14]. L.S Dhillon, Y.Guan and B.Kullis." A unified view of kernel k-means spectral clustering and graph partitioning. Technical Report. Departement of Computer Science. University of Texas Austin. 2005

[15]. Cristianini N, Taylor,J.S., " An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods" Camberidge Press University, 2000.

[16]. Nugroho,A.S, Witarto, A.B, Handoko,D., "Application of Support Vector Machine in Bioinformatics", Proceeding of Indonesian Scientific Meeting in Central japan, Dec,20, 2003, Gifu-Japan.

[17]. Furqan,M.,A.Embong, Suryanti,A, Santi W.P., Sajadin,S.,"Smooth Support Vector Machine For Face Recognition Using Principal Componen Analysis". Proceeding 2nd