International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:4, No:2, 2010

Improving Academic Performance Prediction using Voting Technique in Data Mining

Ikmal Hisyam Mohamad Paris, Lilly Suriani Affendey, and Norwati Mustapha

***Abstract*—**In this paper we compare the accuracy of data mining methods to classifying students in order to predicting student's class grade. These predictions are more useful for identifying weak students and assisting management to take remedial measures at early stages to produce excellent graduate that will graduate at least with second class upper. Firstly we examine single classifiers accuracy on our data set and choose the best one and then ensembles it with a weak classifier to produce simple voting method. We present results show that combining different classifiers outperformed other single classifiers for predicting student performance.

***Keywords*—**Classification, Data Mining, Prediction, Combination of Multiple Classifiers.

I. INTRODUCTION

ACCURATELY predicting students' performance is useful in identifying weak students who are likely to perform poorly in their studies. We have carried out some experiments in order to evaluate the performance of different prediction techniques for predicting student's CGPA class. Our main target students are weak students that achieved CGPA belonging to second class lower and third class. In addition to differentiate methods ability, we divided the class attributes into several classes.

We focused on methods that have a comprehensive visual representation, since all education models should be transparent [2] especially decision tree and naïve Bayes. These aspects are very important especially to convince lecturers to use those methods and at the same time help them easily to make decision making. In this work, decision trees and Bayesian methods that have comprehensive visual representation are considered for classification task. Results of prediction enable the management to take remedial measures at early stage to produce excellent graduates. In his work we compare our proposed voting technique accuracy with C4.5, NBTree, BayesNet, naive Bayes, hidden naive Bayes (HNB) and voting technique based on three weak classifiers (naïve Bayes, OneR and Decision stump).

In order to choose the most accurate algorithm on our classification problem, we use the simplest approach by estimating the accuracy of the candidate algorithms on our dataset and select the one that appears most accurate.

Recently, there has been increased interest in combining classifiers concept that is proposed for the improvement of the performance of individual classifiers. The approach of integration algorithms is used to making decision more reliable, precise and accurate. One of mechanism that is used to build ensemble of classifiers is using different learning methods and we use it in our experiments.

We choose combination of HNB method proposed by [12] as the best single method performs on our data set and one weak classifier that is Decision stump (DS) for voting technique. From our observation we found that HNB performed well on most of classes except for high distribution class but on the other hand decision trees like DS has high accuracy on this class. Based on that reason we choose Decision stump as a compliment method to HNB. The advantages of weak classifiers were reported in [9], [11].

Ikmal Hisyam Mohamad Paris, Universiti Putra Malaysia, Department of Computer Science (syam_paris@yahoo.com.my).

Lilly Suriani Affendey, Universiti Putra Malaysia, Department of Computer Science (suriani@fsktm.upm.edu.my).

Norwati Mustapha, Universiti Putra Malaysia, Department of Computer Science (norwati@fsktm.upm.edu.my).

II. RELATED WORK

Work by [7] have compared two classifiers (decision tree and Bayesian network) to predict students GPA at the end of the third year of undergraduate and at the end of the first year of postgraduate from two different institutes. Each data set has 20,492 and 936 complete student records respectively. The results show that the decision tree outperformed Bayesian network in all classes. The accuracy was further improved by using re-sampling technique especially for decision tree in all cases of classes. In the same time it able to reduce misclassification especially on minority class of imbalanced datasets because decision tree algorithm tends to focus on local optimum.

In the other work, [4] have compared six classification methods (Naive Bayes, decision tree, feed-forward neural network, support vector machine, 3-nearest neighbour and logistic regression) to predict drop-outs in the middle of a course. The data set contained demographic data, results of the first writing assignments and participation to group meetings. The data set contained records of 350 students. Their best classifiers, Naive Bayes and neural network, were able to predict about 80% of drop-outs. The results also showed that simple model such as naïve Bayes able to generalize well on small data set compare to other method such as decision tree and nearest neighbour that require much larger size of datasets.

The comparisons of six classifiers involved quadratic Bayesian classifier, 1-nearest neighbours, k-nearest neighbours, Parzen window, feed-forward neural network, and decision tree to predict the course final results from a learning system log data were runned by [5]. The data contained attributes concerning each task solved and other actions like participating in the communication mechanism and reading support material. The data set contained records of 250 students. Their best classifier, k-nearest neighbours, achieved over 80% accuracy, when the final results had only two classes (pass/fail).

The work of [6] have applied data mining classifiers as a means of analyzing and comparing use and performance of students who have taken a technical course via the web. Their results show that combination of multiple classifiers leads to a significant accuracy improvement in the given data set. Often prediction performance of combining classifiers is better than a single classifier because the decision is relying on collective output of several models.

Weak classifiers are linear classifiers which less likely to suffer from over-fitting problems. Combination of weak classifiers based on boosting approach was used by [11] to predict the final score. Each weak classifier used only one of 74 attributes to predict the course score. The combination achieved only 69% accuracy but the boosting revealed the most influencing factors for the course success. In other work, [9] propose a technique of localized voting of weak classifiers and achieved great accuracy because it does not overlook local singularities as what happened to global learning methods.

III. Dataset, Class Labels, Features

A. Selecting a Data Mining Tool

A detailed comparison of data mining tools that appropriate to predict academic performance was conducted in [7]. They have chosen Weka [10] in term of computational perspective, wider range of algorithms, better data preparation tools and its support for very large data sets. In our experiment we used classifiers provided in WEKA software to predict students' academic performance.

B. Preparing the Data

We were collected 2427 complete records for Bachelor of Computer Science students at University Putra Malaysia (UPM) admitted from 2000 to 2004. Students that failed to complete their studies are not included in our records.

The Bachelor of Computer Science students are required to take a total of 102 credits of subjects which comprised of compulsory and elective courses. The compulsory courses are divided into two components, namely university courses (Public Speaking, Management, Malaysian Nationhood, etc.) and main core courses (Computer Science, Mathematics). Elective courses can be any courses offered by any faculties.

In our records, there are 396 attributes or subjects that registered by previous students. Without attributes selection it makes our data set too large for predicting purpose. An attribute importance analysis was performed in order to rank the attributes by significance in determining the target values as well as to reduce the size of a prediction. Furthermore it helps to increase speed and accuracy of methods in predicting task.

We used algorithm Minimum Description Length (MDL) to rank and we prefer to select courses that have significant contribution to the academic performance. Attributes that have importance value lower than 0.01 were eliminated from dataset. The results of this analysis demonstrate subjects that have strong correlation with student graduated class as shows in Table I. This information is very useful for the Faculty's management to monitor the deliverables of the top ranking courses.

TABLE I
Subject Attributes and Attributes Importance

Rank	Att	Imp	Rank	Att	Imp
1	SMM3001	0.136790	23	MTH3002	0.025108
2	SAK3103	0.136049	24	SMM3312	0.024562
3	MTH3100	0.124986	25	SIM3202	0.024350
4	MGM2111	0.113718	26	SIM3303	0.022526
5	SAK3309	0.112922	27	SMM3311	0.022392
6	SAK3408	0.109821	28	SKR3303	0.021949
7	SAK3117	0.107712	29	SKP2201	0.021595
8	SAK3207	0.102695	30	SKR4401	0.021592
9	SKR3200	0.099232	31	EDU3616	0.021139
10	SAK3101	0.086601	32	SKR4301	0.019533
11	SAK3109	0.083313	33	SIM4306	0.019364
12	SAK3100	0.063167	34	SKR4402	0.017337
13	SIM3302	0.052145	35	SIM4307	0.016743
14	SKP2202	0.048784	36	BBI2409	0.015445
15	SKP2101	0.043316	37	SIM4300	0.015088
16	SMM3111	0.031687	38	SKR3201	0.013643
17	SKR3504	0.031185	39	SMM3112	0.013100
18	BBI2410	0.030800	40	SAK4801	0.012184
19	SMM4302	0.026182	41	SAK4610	0.011858
20	SKR3202	0.026018	42	SMM4301	0.011363
21	SKR4200	0.025702	43	KOC3433	0.011127
22	SAK4401	0.025218			

From these attributes, we can grouping the subjects based on course code that are Computer Science department course (SAK), Information System department course (SIM), Communication Technology and Network department course (SKR), Multimedia department course (SMM), university course (SKP), Mathematic course (MTH), English course (BBI), Communication course (KOC), Educational course (EDU) and Management course (MGM).

From the Table I, we know that main university courses, main core courses especially Mathematic and Programming subjects (SAK3100, SAK3101 and SAK3109) have important role compare to elective courses. Certain subjects such as programming, English and Mathematic subjects were reported before as important roles that influence students in academic achievements.

C. Modeling the Academic Performance Classification Problem

Table II shows the grouping of data into various categories. The values (A, A-, B+, B, B-, etc.) are actual grades obtained by students for each subjects. In our case we prefer to represent not taken subject with a specific value instead of mode value. The CGPA for first class is 4.0-3.75, second class upper is 3.74- 3.5, second class lower is 3.4 – 3.0, third class is 2.75 – 2.00. Other experimented to group the data result in a slightly lower accuracy for classification.

TABLE II
ATTRIBUTE VALUES

Values	Categories / groups
A, A-	A
B+, B, B-	B
C+, C, C-	C
D+, D	D
F	E
Not taken	N

IV. Voting Technique

Voting is an aggregation technique used to combine decisions of multiple classifiers. In its simplest form that based on plurality or majority voting, each individual classifier contributes a single vote [1]. The aggregation prediction is decided by the majority of the votes, i.e., the class with the most votes is the final prediction. The final prediction is decided by summing up all votes and by choosing the class with the highest aggregate.

A voting technique that used three most common weak machine learning algorithm OneR, Decisionstump and naïve Bayes as learner have proposed by [4]. The advantages of weak classifiers were reported in [9], [11]. The weak classifiers are less likely to suffer from over-fitting problem, since they avoid learning outliers, or quite possibly a noisy decision boundary and the training time is often less for generating an ensemble classifier as what reported previously.

Based on our experiments we found that HNB performed well on most of classes except for high distribution class but on the other hand decision trees have high accuracy on this class. On another experiments, we found that weak classifier from decision trees such as DS has better accuracy in high distribution class compare to other decision tree methods such as C4.5, ID3 and simpleCart.

Based on that reason, we plan to choose combination of the best classifier on our dataset that is HNB and Decisonstump to form a voting technique as shows in Fig. 1.

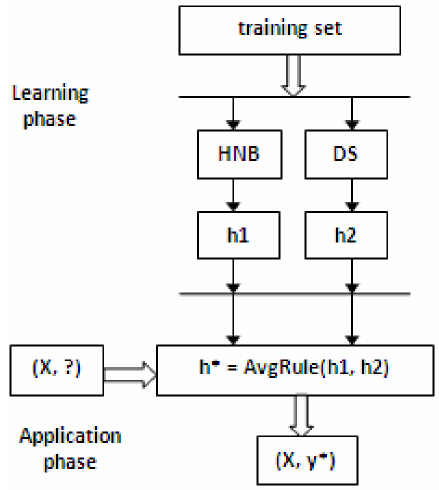

Fig. 1 Model of the proposed voting technique

In our case, each individual classifier (HNB and DS) generate their hypothesis respectively (h1, h2). For each output class, a-posteriori probabilities are generated by the individual classifiers. Next, the class represented by the maximum average value of the a-posteriori probabilities is selected to be the voting hypothesis (h*) to determine a decision, hence reducing the generalization error of prediction.

V. Results and Analysis

We perform three experiments with different number of classes to classify. The first experiment was to predict classification of students into 2 classes. The first class is CGPA of first class and second class upper, the second class is CGPA of second class lower and third class. The second experiment was to predict classification of students into 3 classes. The first class is CGPA of first class, the second class is CGPA of second class upper, and the third class is second class lower and third class and the third experiment was to predict classification of students into 4 classes. The first class is CGPA of first class, the second class is CGPA of second class upper, the third class is second class lower and the fourth class is third class.

We applied different classifiers on our data set and obtained the following results for all experiments on different classes as shown in Table III. A Cross-Validation with 10 folds are carried out to evaluate the prediction accuracy. The results show that the ensemble method performed best on our data set compare to other single methods and the accuracy of classifiers decreased when the numbers of classes become bigger.

It is also important to note that our attempt to resample the data set in order to create a more balanced distribution for training the algorithms has not improved the accuracy of prediction in all cases. Only certain algorithms improve their classification performance when apply preprocessing tasks as rebalancing data as showed in [8]. It might caused by adding weight to features not containing the target concept that appear in positive class will affect the results.

Based on results in Table III it shows that HNB performed best in all cases and important to state that HNB performed well on most of classes except for high distribution class but on the other hand decision trees have high accuracy on this class. It also shows that a simple combination of HNB and DS achieved best on our dataset and has only slight decrease on one class compare to previous voting method.

TABLE III
COMPARING THE ACCURACY OF ALL CLASSIFIERS IN ALL CASES OF CLASSES

Classifier	Performance %		
	2-Classes	3-Classes	4-Classes
C4.5	90.5	84.3	82.7
NBTree	89.7	84.6	82.7
BayesNet	87.9	83.3	81.5
NB	88.3	83.1	81.8
HNB	90.7	86.7	85.3
Voting	**94.9**	84.6	82.8
Proposed	93.8	**91.6**	**89.5**

Although ensemble methods perform well compare to single method these methods have weaknesses in term of comprehensibility because it is not easy to understand the underlying reasoning process leading to a decision, and the other weakness is increased computation because the number of individual models involved in learning process**.**

VI. CONCLUSION AND FUTURE WORK

Identifying the attributes that contribute the most significant to the student's academic performance can help to improve the intervention strategies and support services for students who perform poorly in their studies, at an earlier stage. Since educational data is normally skewed as well as sparse, a lot of effort must be put into the preprocessing steps to ensure the filtering process gives a good model. The additional work need to be taken to model the prediction outputs as useful information for identifying weak students.

Our results show that combining different classifiers improved the prediction accuracy compare to single classifiers as disadvantages of one method might be compensated by others. The results also show that the HNB method consistently outperformed other single methods on our educational dataset.

For the future research we plan to build more comprehensively voting technique that completely cover important issues especially to handle imbalanced dataset and to find suitable feature weighting scheme for our dataset using appropriate techniques in order to improve the prediction performance especially on low and high distribution classes..

REFERENCES

[1] L. Hall, K. Bowyer, W. Kegelmeyer, T. Moore, C. Chao (2000) "Distributed learning on very large data sets". In: Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 79–84.

[2] W. Hamalainen, "Descriptive and Predictive Modelling Techniques for Educational Technology", Thesis Department of Computer Science, University of Joensuu, Finland.

[3] W. Hamalainen, and M. Vinni, "Comparison of machine learning methods for intelligent tutoring systems", In proceedings of the 8th International Conference on Intelligent Tutoring Systems, Jhongli, Taiwan, pp. 525-534, June 2006.

[4] S. B. Kotsiantis, C. J. Pierrakeas, and P. E. Pintelas, "Preventing student dropout in distance learning using machine learning tachniques", In proceedings of 7th International Conference on Knowledge-Based Intelligent Information and Engineering Systems (KES 2003), pp. 267-274, 2003. ISBN 3-540-40803-7.

[5] B. Minaei-Bidgoli, Kashy, D. A. Kortemeyer, G. and Punch, W. F. "Predicting student performance: an application of data mining methods with an educational web-based system", *33rd Annual Conference on Frontiers in Education (FIE 2003)*, vol. 1, pp. 13-18, 2003. DOI: 10.1109/FIE.2003.1264654.

[6] B. Minaei-Bidgoli, G. Kortemeyer, and W. F. Punch, "Enhancing Online Learning Performance: An Application of Data Mining Method", In proceedings of The 7th IASTED International Conference on Computers and Advanced Technology in Education (CATE 2004), Kauai, Hawaii, USA, pp. 173-8, August 2004.

[7] Nguyen Thai Nghe, P. Janecek, and P. Haddawy, "A comparative analysis of techniques for predicting academic performance", ASEE/IEEE Frontiers in Education Conference, pp. T2G7-T2G12, 2007.

[8] C. Romero, S. Ventura, P. G. Espejo, and C. Hervas, "Data Mining Algorithms to Classify Students", 1st International Conference on Educational Data Mining, Montreal, Quebec, Canada, 2008. ISBN-13:9780615306292.

[9] S. B. Kotsiantis and P. E. Pintelas, "Local voting of weak classifiers", KES Journal. 3(9):pp. 239-248, 2005.

[10] Weka, University of Waikato, New Zealand, http://www.cs.waikato.ac.nz/ml/weka.

[11] W. Zang, and F. Lin, "Investigation of web-based teaching and learning by boosting algorithms". In *Proceedings of IEEE International Conference on Information Technology: Research and Education (ITRE 2003)*, pp. 445–449, 2003.

[12] H. Zhang, L. Jiang, J. Su, "Hidden naive Bayes", American Association for Artificial Intelligence. AAAI pp. 919-924, 2005.