

Intelligent Report and Predicting Student's Performance using Classification via clustering in e-learning System

P.Gopinadh¹, Mr. K. Palanivel²

Department of Computer Science¹, Computer Centre Pondicherry University², Pondicherry.

gopi24g@gmail.com¹, kpalani@yhaoo.com²

Abstract-Here we are developing a enhanced learning management system (LMS) which gives a intelligent report as a result of all aspect in the e-learning system and in this paper we discuss about the prediction of students final performance based on their activities which they doing in the online learning system by the effective and efficient algorithms which gives classification accuracy by classification via clustering method followed a prediction by clustering plus class association rules mining at the middle and end of the course. Our ultimate aim is to predict the student mark at the middle of the course duration and with the students performance it will be insisted by the instructor to the students for a better improvement at the final exams by this we can get a high percentage of student have a chance to get pass in the final exams this work is the new and improved advancement in EDM. The future work may be contain this prediction in the field of Business Management Process which includes improvement in online shopping, prediction of weather forecasting, etc.

Keywords: Educational Data Mining(EDM), Learning Management System(LMS), Business Management Process(BMP)

INTRODUCTION:

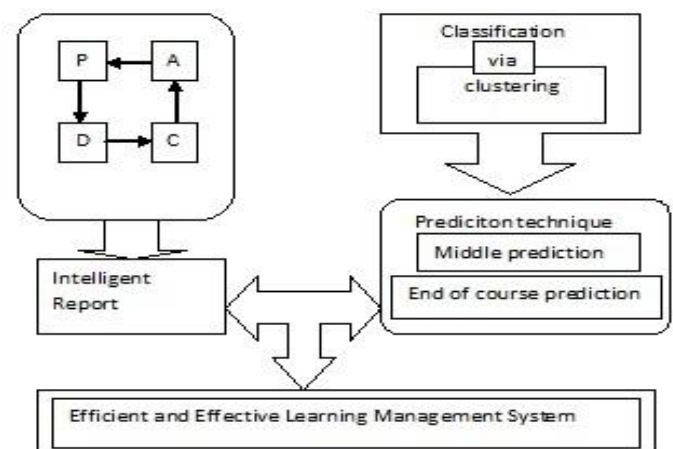
This learning management system was developing today in a tremendous way of new technologies and techniques in this proposal the e-learning system has the feature of predicting the students result in middle of the course by the instructors prediction policy of algorithm of classification and it defines the student result will obtain more in numbers of possibility in passing the final exams and also producing a intelligent report with the existing work PDCA (Plan, Do, Check, Act) [5] method of improving web-based intelligent reports of an e-learning system as intelligent system, The focus is on improving LMSs[1] (Learning Management Systems) or e-learning systems by predicting patterns for intelligent report and final grades of students by efficient algorithms, classification via clustering including improving classification accuracy and clustering plus class association rules mining for earlier and final prediction at the end of the course.

New learning management system is improved by using data mining techniques and increasing the efficiency of the result of passing percentage. Our objective is to predict at the middle of the course to find possibility of the passing students and then it decides the ratio of the passing students with the classification algorithm which is a predetermined existing algorithm used in data mining tools to find the results, by this result the instructor will give a message to students you may get pass or fail by the performance given by the students so it is a good opportunity to the students and also the LMS to develop the passing number of possibility

may be increase at the final exams and the final prediction may increase the number of students passing is more than the middle of the course. This proposed module is a very small idea but it may develop the e learning system much more well in this prediction based method and improve the LMS.

Architecture of efficient and effective LMS.

Fig1



OBJECTIVE OF PROBLEM DEFINITION:

Gives an intelligent report and predicting students final performance middle and end of the course says more effective and efficient in educational data mining environment how the selection of instances and attributes,

the use of different classification algorithms and comprehensibility of the prediction with an intelligent report.

The use of these systems has resulted in a need for Educational data mining to predict the student performance at the middle and causing the result more at the final prediction gives the possibility of higher number of percentage in passing the exams performs a new developed feature of this learning management system. A universal solution for this problem lies in the use of data mining techniques, with the possibility of improving LMS [2]. Data mining or knowledge discovery in databases (KDDs) is the automatic extraction of implicit and interesting patterns from large data collections [3]. Part of the LMS report also requires improvement in terms of including web intelligence to detect prediction by classification technique and clustering plus class association rule algorithm [4]. The need for such a solution includes the prediction of student's performance in the middle and end of the course.

RELATED WORK

Existing Module:

In the architecture of the system the PDCA describes as from the existing module for obtaining the intelligent report [1].

- In phase P (Plan), four tasks are defined: establishment of participants, design of questionnaires, locating passive participants, and pilot tests.
- Phase D (Do) provides information on completing the questionnaire, forwarded schools, and specificity.
- In phase C (Check), analysis gives concrete results in regards to the percentage of completed questionnaires.
- Finally, phase A (Act) gives identified measures to improve the obtained results.

1. Defining the first formal structure of the environment and relationships within the e-learning system.
2. Creating appropriate data mining models.
3. Writing Data Mining Extensions (DMXs) to obtain predictive patterns.
4. Creating a web-based application that has the ability to interact with the user.
5. OLAP analysis: the designing of a multidimensional structure in which the main factors under analysis (i.e., year, month, day, time, minutes, course, activity, and module) will be taken as dimensions and will be used later to build an OLAP cube in order to analyze the recorded data.[6]
6. Data pre-processing to clean and prepare the forum data.
7. Describing the process and object models.
8. Selecting key resources for the Moodle LMS.
9. Creating links between applications and a DMX-over-Adomdnet.

10. Defining educational elements

11. Visualizing the results.

The Modules and features of the existing and proposed system

Existing

User's activity report
 Visualization of results
 Real time data analysis
 Interactivity
 Existence of web intelligence in the system
 Prediction of user behavior Patterns
 Numerical results for the approach probability

Proposed features:

Classification via clustering algorithm clustering and class association rules
 Prediction of students grade result
 Middle prediction to improve the LMS
 EDM (Educational Data Mining) has enrichment in Data Mining technique with classification accuracy

Proposed Work:

A representative set of traditional classification algorithms [7] have been used and compared versus classification via clustering algorithms for predicting whether students will pass or fail the course on the basis of data about their performance database usage of the e-learning system periodically and results obtained. The results obtained indicate the suitability of performing a middle prediction at the middle of the course duration and final prediction end of the course by applying clustering plus class association rules mining instead of traditional classification for obtaining highly importance in this result examining of student performance based with the collection of attributes of mark of various subjects conducted by the online educational system. By this entire database we are going to do the middle prediction at the time of the half state.

Middle Prediction:

Periodical time calculation of middle prediction occurrence:
 $Mp = t/2$

Where

Mp – middle prediction of n datasets.
 n data sets refers an individual student attends how many number of test conducted by LMS.
 t - Time (Course duration) total no. of days in a course selected by the student.

End of Course/ Final Prediction:

Periodical time calculation of final/end prediction occurrence:

$$Fp = t - d$$

Where

Fp – Final prediction of n datasets

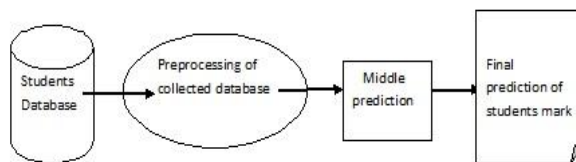
t – Total time

d – Days are reduced from the course duration days.

d can be modified or set by the administrator for predicting the final prediction whenever the instructor needed or by the LMS.

Methods:**Collecting students database [8]:**

Fig 2

**Pre-processing data:**

The first thing is data pre-processing is an important step for preparing and filtering data before applying data mining algorithms. In our case, three main pre-processing tasks have been applied to the previous summarization excel file (i.e.): instance selection, attribute selection and data transformation [9].

Instance selection:

Many different approaches have been used to address the data explosion issue, such as algorithm scale-up and data reduction. Instance, example, or tuple selection pertains to methods or algorithms that select or search for a representative portion of data that can fulfill a KDD task as if the whole data is used. Instance selection is directly related to data reduction and becomes increasingly important in many KDD applications due to the need for processing efficiency and/or storage efficiency. One of the major means of instance selection is sampling whereby a sample is selected for testing and analysis, and randomness is a key element in the process. Instance selection also covers methods that require search. This is a data reduction task by choosing just one subset of data instances. In our case, we used two types of values filtering criteria. The first one selects marks obtained at different exams, moments/dates during the course, in our case in the middle and at the end of the course. The second one allows us to select all the marks or to exclude the invalid attributes and messages, i.e., (the

entire test obtained mark) to select only the needed data and ignore other data to be specific.

Attribute selection:

This is a feature or variable selection task for reducing the data dimensionality by selecting a subset of relevant attributes. In our case, although we do not have a great number of available attributes/indicators, some of them may be irrelevant for predicting students' performance. We have used a ranking approach to select attributes given that there exist a wide range of attribute selection algorithms.

Data transformation:

This task converts a set of data values from the data format of a source data system into the data format of a destination data system. In our case, data is transformed from the Excel format provided by our Moodle's module to the .ARFF Weka format

Predicting final mark:

Our objective is to predict the mark of the student in the middle of the course and by which the students can improve the level of learning will improve accordingly with some higher in a short period of time and students can get passed in more number of percentage while comparing to the middle prediction of the course. Which method are used for middle prediction and the final prediction are applied by various data mining algorithms which are so efficient and giving the nearest prediction values.

So, this is a classification problem and not a regression problem. We propose to do it by using different DM (Data Mining) methods. On the one hand, we propose to use a traditional classifier given that it is the traditional data mining method used for solving the task. On the other hand, we propose to use classification via clustering as an alternative method.

Classification is a supervised method for identifying to which set of categories (labels) a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known. A classifier is a mapping from a (discrete or continuous) feature space X to a discrete set of labels Y of the class to be predicted. In our case, the exploratory variables or feature space are the forum interaction/participation data, and the outcome or class to be predicted is the final grade in the course.

Clustering is a process of grouping objects into groups of similar objects or subsets or clusters. Clustering and classification are both classification methods, although clustering is an unsupervised method and classification is a supervised method. Classification via/based on clustering is an approach

for using clustering as a classifier based on the assumption that each cluster corresponds to a class [10]. A clustering algorithm is firstly executed using the training data, after removal of the class attribute, and the mapping between classes and clusters is determined. This mapping is then used to predict class labels for unseen instances in the test data. In other words, the class attribute is not used in clustering but it is used to evaluate the obtained clusters as classifiers. So, it is important to ensure that the number of clusters generated is the same as the number of class labels in the dataset in order to obtain a useful model that relates each cluster with one class. The advantage of using classification via clustering is the ability to obtain a general display of the two groups or clusters which are generated are students that pass and the students that fail.

CLASSIFICATION VIA CLUSTERING:

Here I applied Weka tool for k-fold cross validation it will work in combination with classification via clustering. Some machine learning tools which uses the algorithms of classification in predefined manner are listed here as Bioconductor, GGobi, R, Yale, Tanagra, Weka and orange etc, are used worldwide to do the data mining works for the implementation. The interpretation and employment of the previously obtained classification models can be very useful for instructors to detect the final performance of students in time and to make decisions about helping students predicted to fail.

DATABASES GATHERED:

Middle of the course, when the summary information was included in a data file called Dataset 1 which contains n number of database of test attended by the candidate. The result which obtains the student result will be imported in a database for the future reference for the mark list which collected by the prediction made by the middle of the course.

Then the next step is to intimate the student who are all failed by the instructor of the e-learning system because of the student should be get alerted if they are not attending the next test and not giving a good performance in the upcoming test and showing involvement to the e-learning course.

End of the course, by the LMS the next dataset is collected called dataset2 which contains the database of all the test event conducted by the online learning system also it includes the database of appeared retest which missed at the first cycle of test upto the middle of the course duration and by this again the prediction is made by the instructor to say what will be prediction result of the student before the final exam by concluding with the middle prediction and final prediction or end prediction the passing criteria

must be improved with this two predictions made by the learning system.

Applying some traditional classification algorithms provided by Weka:

1. Rule-based algorithms that reveal rules
2. Tree-based algorithms that reveal a decision tree.
3. Function-based algorithms that reveal a function
4. Bayes-based algorithms that reveal a probabilistic classifier based on Bayes theorem

Using classification via clustering approach in which we executed clustering algorithms by setting them to generate 2 clusters in order to predict the two classes (Pass/Fail) correctly on the basis of the obtained clusters. Executing the following clustering algorithms provided by Weka is:

- SimpleKMeans[11] (Kanungo et al., 2000) is the most widely-used, simple and well-known clustering algorithm, which aims to partition instances into k clusters in which each instance belongs to the cluster with the nearest mean.
- Xmeans[12] (Pelleg & Moore, 2000) is an extension of the SimpleKMeans algorithm that efficiently searches the space of cluster locations to optimize the Bayesian Information Criterion (BIC) measure.
- EM [13] (Expectation Maximization) algorithm (Moon, 1996) that assigns a probability distribution to each instance which indicates the probability of it belonging to each of the clusters.

OBJECTIVE OF EXPERIMENT AND RESULTS:

While comparing the traditional classification versus classification via clustering algorithms we obtain the how optimal accuracy we obtained in this module

If we compare Datasets 1 and 2 **versus** Datasets 3 and 4, where dataset 3 is the result obtained with the middle prediction and dataset 4 is the result obtained by final prediction which student will be pass or gets fail will be saved as a database for the instructor to future references. It has more possible to get high number of students will get pass after the middle of the prediction made by the instructor it helps the students and also develop a high feature of efficient module which is delivered as a intelligent report and effective Learning Management System. [14] Dataset 1 and 2 gives the details of marks obtained by the students and the database collected after the prediction of result is named as database 3 and 4. assert that Dataset 4 behaved statistically better than Dataset 2 in both measures. We conclude that

the database 3 and 4 may gives best accuracy and were obtained when predicting at the end of the course, the average values obtained in the middle of the course (Dataset 1) are between 70 and 80% accurate, which is a good accuracy rate for making an early prediction compared to the final average accuracy values of between 80 and 90% (Dataset 2).

CONCLUSIONS AND FUTURE WORK:

In this research we analyze an intelligent report e-learning system created with data mining techniques leads to a modern and intelligent way of reporting user activity. Compared with the existing reporting system the proposed system presents an improvement since it [15] predicts the students results at the middle of the course and gives a result by which we can get a report whom might get a passed and fail, with this middle of the prediction the student gets a alert from the instructor and the students might do the performance better after the middle prediction and there are many chance of getting pass at the final prediction and also the student have more chance to get pass in the final exams, not only getting the intelligent report with this feature the prediction is added so the learning management system might get more and more efficient. This paper has investigated how different classification data mining approaches can be used to improve the prediction of students' final performance on the basis of their performance in an on-line exams and marks obtained.

We have shown how a classification via clustering approach can be used instead of traditional classification algorithms and also applying clustering plus class association rules mining for earlier and final prediction at the end of the course. Intelligent report and prediction of students grade by different classification technique was a good approach is developed in the field of prediction based data mining says a well developed model of a EDM in e-learning system.

The future work is to develop the prediction in high modification of implementing in the new fields of online shopping that will improve the Business management Process and also many researches should implement this prediction in the real time projects like day to day using projects like weather forecasting for much effective, predictive of the disease in the Hospital medicine description analysis etc.,

REFERENCES:

[1] Castro F, Vellido A, Nebot A, Mugica F. Applying data mining techniques to e-learning problems.<<http://www.sci2s.ugr.es/keel/pdf/specific/capitulo/ApplyingDataMiningTechniques.pdf>> [accessed 10.01.12].

[2] Zaiane O, Luo J. Web usage mining for a better web-based learning environment. In:Proceedings of conference on advanced technology for education.Banff, Alberta; 2001.p.60-4.

[3]Klosgen W, Zitkow J. Handbook of data mining and knowledge discovery. New York: Oxford University press, 2002

[4]Zorrilla ME, Marin D, Alvarez E. Towards virtual course evaluation using web intelligence.<<http://www.personales.unican.es/ZORRILLM/PDFs/ME%20Zorrilla%20-%20Towards%20Virtual%20Course.pdf>> [accessed 15.01.12].

[5]Tanigawa T, Yamakita T, Fujii T, Maeda M, Fuji T. Construction of driving model with faculty perspectives of ePortfolio for Improving University Education in Japan.<<http://www.editlib.org/p/35651>> [accessed 08.02.12].

[6]IBM Redbooks. DB2 Cube Views: A Primer. Durham, NC, USA: IBM, 2003. ebrarycollections.SanJoseStateUniversity. <<http://site.ebrary.com/lib/sjsu/Doc?id=10113016&ppg=43>>

[7]Duda, R. O., Hart, P. E., & Stork, D. G. (2000).Pattern classification. Wiley Interscience

[8]Cole, J., & Foster, H. (2007).Using Moodle: Teaching with the popular open source course management system. O'Reilly Media, Inc

[9] Klosgen, W., & Zytow, J. (2002). Handbook of data mining and knowledge discovery. New York: Oxford University Press

[10] Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. ACM Computing Surveys, 31(3), 264–323.

[11]Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C., Silverman, R., & Wu, A. Y. (2000, May). The analysis of a simple k-means clustering algorithm. In Proceedings of thesixteenth annual symposium on computational geometry (pp. 100–109).

[12]Pelleg, D., & Moore, A. (2000). X-means: extending k-means with efficient estimation of the number of clusters. In Proceedings of the seventeenth international conference on machine learning, 1 (pp. 727–734).

[13]Moon, T. K. (1996). The expectation-maximization algorithm. Signal Processing Magazine, IEEE, 13(6), 47–60.

[14]Romero, C., Espejo, P., Romero, R., & Ventura, S. (2013).Web usage mining for predicting final marks of students that use Moodle courses. Computer Applications in Engineering Education, 21(1), 135–146.

[15] Cristóbal Romero*, Manuel-Ignacio López, Jose-María Luna, Sebastián Ventura Department of Computer Sciences and Numerical Analysis, University of Córdoba, 14071 Córdoba, Spain