

Clustering Students of Computer in Terms of Level of Programming

Shirin Esmailbegi Kermani

Department of Computer, Kerman Branch, Islamic
Azad University, Kerman, Iran

Abbas Rezaie

Department of Computer, Kerman Branch, Islamic
Azad University, Kerman, Iran

Abstract: Educational data mining (EDM) is one of the applications of data mining. In educational data mining, there are two key domains, i.e. student domain and faculty domain. Different type of research work has been done in both domains.

In existing system the faculty performance has calculated on the basis of two parameters i.e. Student feedback and the result of student in that subject. In existing system we define two approaches one is multiple classifier approach and the other is a single classifier approach and comparing them, for relative evaluation of faculty performance using data mining

Techniques. In multiple classifier approach K-nearest neighbor (KNN) is used in first step and Rule based classification is used in the second step of classification while in single classifier approach only KNN is used in both steps of classification.

But in proposed system, I will analyse the faculty performance using 4 parameters i.e., student complaint about faculty, Student review feedback for faculty, students feedback, and students result etc.

For this proposed system I will be going to use opinion mining technique for analyzing performance of faculty and calculating score of each faculty.

Keywords: clustering algorithm, Data mining, Educational data mining, classification

1. INTRODUCTION

1. Data Mining is the process of analyzing data from different perspectives and summarizing the result as useful information. Educational Data mining is an emerging research field in data mining and continuously growing with a boom. Incalculable data mining techniques have been applied to the variety of educational data [15].
2. For the better decision making in the learning environment, EDM consists of four phases: Data Preprocessing Data Validation Data Prediction Decision making process.
3. This study investigates the education domain for relative evaluation of faculty performance on the basis of two parameters student feedback and result in a course taught by faculty. Classification technique of data mining is applied to fulfill the objective. The goal of classification technique is to place an object into a category based on the characteristics of the object.
4. The whole paper is organized as: In section 1 Introduction about DM, EDM and some idea of relative problem is mentioned. In section 2 overview of related work in which educational data mining on student domain and faculty domain is discussed from previous research. In section 3 proposed brings up where educational data mining is presented in faculty domain. The proposed framework is explained in section 4.

For the better decision making in the learning environment, EDM consists of four phases:

Data Pre-processing

Data Validation

Data Prediction

Decision making process

2. LITERATURE REVIEW

2.2 A Comprehensive Study of Educational Data Mining:

Author Jasvinder Kumar suggests that, educational data mining is a new discipline in research community that applies various tools and techniques of data mining (DM) to explore data in the field of education [2]. This discipline helps to learn and develop models for the growth of education environment. It provides decision makers a better understanding of student learning and the environment setting in as of EDM. It also highlights the opportunities for future research. Educational data Mining (EDM) has been evolved as multidisciplinary scientific learning area, rich in data, methods, tools and techniques used to provide better learning environment for educational users in educational context. This paper integrates all the modules of EDM required to facilitate the objectives of educational research. Lastly it shows that, there are many more research topics that exist in this domain.[2] Utilization of data mining techniques within education environment requires a joint effort by the ICT specialists, educationists and the learners.

2.3 Educational Data Mining: a Case Study-

Author Kalina suggest in this paper, Author show how using data mining algorithms can help discovering pedagogically relevant knowledge contained in databases obtained from Web-based educational systems. These findings can be used both to help teachers with managing their class, understand their students learning and reflect on their teaching and to support learner reflection and provide proactive feedback to learners. In this paper, Author has shown how the discovery of different patterns through different data mining algorithms and visualization techniques suggests to us a simple pedagogical policy. Data exploration focused on the number of attempted exercises combined with classification led us to identify

students at risk, those who have not trained enough. Clustering and cluster visualization led us to identify a particular behavior among failing students, when students try out the logic rules of the pop-up menu of the tool. As in [5], a timely and appropriate warning to students at risk could help preventing failing in the final exam. Therefore it seems to us that data mining has a lot of potential for education, and can bring a lot of benefits in the form of sensible, easy to implement pedagogical policies as above. The way Author have performed clustering may seem rough, as only few variables, namely the number and type of mistakes, the number of exercises have been used to cluster students in homogeneous groups. This is due to our particular data. All exercises are about formal proofs. Even if they differ in their difficulty, they do not fundamentally differ in the concepts students have to grasp. Author have discovered a behavior rather than particular abilities. In a different context, clustering students to find homogeneous groups regarding skills should take into account answers to a particular set of exercises. Currently, Author are doing research work along these lines.

2.4 Extraction of rules based on students questionnaires:

Author Manolis Chalaris; suggest there are many students in the Greek Higher Education that are still “lingering” in their Departments beyond the six years. The length of studies beyond 6 years has not been justified, and this study focuses on this problem. Author also studies another problem: The percentage of graduates scoring about 8.5/10 or more is extremely low [8]. Association rules mining is a well known data analysis method for extracting associations between data in a wide range of different fields. In this paper, Author focus on the generation of the appropriate association rules based on students’ questionnaires in Higher Education. A sample of 50.000 questionnaires was filled by 10.000 students in the TEI of Athens. Various interesting rules could be extracted related to learning goals, practices, years required for graduation, etc. These rules and clustering techniques could be used for solving the problem of the students that are still “lingering”, and the problem of the low “scoring” of the graduates. In this section, Author use the simple example of the section 3 in order to present the potential impact of these results in Higher Education, and how the described technique could be competitive.

2.5 Predicting Students academic Performance using Artificial Neural Network: A Case Study of an Engineering Course:

Author V.O. Oladokun suggest that , The observed poor quality of graduates of some Nigerian Universities in recent times has been partly traced to inadequacies of the National University Admission Examination System. In this study an Artificial Neural Network (ANN) model, for predicting the likely performance of a candidate being considered for admission into the university was developed and tested. Various factors that may likely influence the performance of a student were identified. Such factors as ordinary level subjects scores and subjects’ combination, matriculation examination scores, age on admission, parental background, types and location of

secondary school attended and gender, among others, were then used as input variables for the ANN model. A model based on the Multilayer Preceptor Topology was developed and trained using data spanning five generations of graduates from an Engineering Department of University of Ibadan, Nigeria’s first University. Test data evaluation shows that the ANN model is able to correctly predict the performance of more than 70% of prospective students. This study has shown the potential of the artificial neural network for enhancing the effectiveness of a university admission system. The model was developed based on some selected input variables from the pre admission data of five different sets of university graduates. It achieved an accuracy of over 74%, which shows the potential efficacy of Artificial Neural Network as a prediction tool and a selection criterion for candidates seeking admission into a university. One limitation of this model stems from the fact that not all the relevant performance influencing factors are obtainable from the pre-admission record forms filled by the students. A model incorporating the use of results from a carefully designed oral interview administered to the students may likely be an improvement over the present model. Also the extension this research to non-engineering departments is recommended. The current admissions system should be reviewed in order to improve the standard of candidates being admitted into the institution. A more adequate ANN may be very useful for such an exercise.[11]

2.6 Survey on Decision Tree Classification algorithms for the Evaluation of Student Performance:

Author Anjou Rathee, suggest that Now days, the amount of data stored in educational database is increasing rapidly. These databases contain hidden information for improvement of student’s performance. Classification of data objects is a data mining and knowledge management technique used in grouping similar data objects together. There are many classification algorithms available in literature but decision tree is the most commonly used because of its ease of implementation and easier to understand compared to other classification algorithms. The ID3, C4.5 and CART decision tree algorithms has been applied on the data of students to predict their performance. In this paper, all the algorithms are explained one by one. Performance and results are compared of all algorithms and evaluation is done by already existing datasets. All the algorithms have a satisfactory performance but accuracy is more witnessed in case of C4.5 algorithm. In this paper, three existing decision tree algorithms have been applied on the educational data for predicting the student’s performance in examination. All the algorithms are applied on student’s internal assessment data to predict their performance in the final exam. The efficiency of various decision tree algorithms can be analyzed based on their accuracy and time taken to derive the tree. The predictions obtained from the system have helped the tutor to identify the weak students and improve their performance. [14] C4.5 is the best algorithm among all the three because it provides better accuracy and efficiency than the other algorithms. Still effective algorithms for decision tree should be developed.

2.7 Survey on Decision Tree Classification algorithms for the Evaluation of Student Performance:

Author Anjou Rathee, suggest that Now days, the amount of data stored in educational database is increasing rapidly. These databases contain hidden information for improvement of student’s performance. Classification of data objects is a data mining and knowledge management technique used in grouping similar data objects together. There are many classification algorithms available in literature but decision tree is the most commonly used because of its ease of implementation and easier to understand compared to other classification algorithms. The ID3, C4.5 and CART decision tree algorithms has been applied on the data of students to predict their performance. In this paper, all the algorithms are explained one by one. Performance and results are compared of all algorithms and evaluation is done by already existing datasets. All the algorithms has a satisfactory performance but accuracy is more witnessed in case of C4.5 algorithm. In this paper, three existing decision tree algorithms have been applied on the educational data for predicting the student’s performance in examination. All the algorithms are applied on student’s internal assessment data to predict their performance in the final exam. The efficiency of various decision tree algorithms can be analyzed based on their accuracy and time taken to derive the tree. The predictions obtained from the system have helped the tutor to identify the weak students and improve their performance.[14] C4.5 is the best algorithm among all the three because it provides better accuracy and efficiency than the other algorithms. Still effective algorithms for decision tree should be developed.

a. Rule Mining Framework for Students Performance Evaluation:

Author P. Ajith, B. Tejaswi, M.S.S.Sai suggest that Academic Data Mining used many techniques such as Decision Trees, Neural Networks, Naïve Bayes, K- Nearest neighbor, and many others. Using these techniques many kinds of knowledge can be discovered such as association rules, classifications and clustering. The discovered knowledge can be used for prediction and analysis purposes of student patterns. Prior approaches used decision tree classifications optimized with ID3 algorithms to obtain such patterns. Among sets of items in transaction databases, Association Rules aims at discovering implicative tendencies that can be valuable information for the decision-maker which is absent in tree based classifications. So we propose a new interactive approach to prune and filter discovered rules. First, Author propose to integrate user knowledge in the post processing task. Second, Author propose a Rule Schema formalism extending the specifications to obtain association rules from knowledge base. Furthermore, an interactive framework is designed to assist the user throughout the analyzing task. Applying our new approach to discover the likelihood of students deviations / requiring special attention is organized and efficient providing more insight by considering more information.[12]

3. PROPOSED WORK:

Similar to this, the Data Mining technique can also help the education industry to achieve better goals. Educational institutes can apply data mining techniques for analyzing student performance, faculty performances and the overall performance of institutes. Institutes can study comparative performances with other institutes to compete with and lead other institutes and improve their ratings.

Data Collection
Data Preprocessing
Refined Data
KNN Algorithm
4 Parameters used(Calculate result for each parameter)
Sum of all parameters
Rule based classification algorithm
Analysis of all subject and class of each faculty result

Figure 1 Flow of System

So, By using multiple classifier approach and cure clustering algorithm the performance of the system is better than the existing one. Here KNN is used for mapping purpose. And rule based algorithm is used for deciding the rules on the data. Following figure shows the flow of the proposed system that will be implemented.

4. CONCLUSIONS

The overall performance of multiple classifier approach is better than the single classifier approach. In the second step of multiple classifier approach we Rule-based classification have been used where the authors define their own rules for classification, which make the difference from single classifier approach. In single classifier approach we sum up the both parameters scores which may restrict the performance of this approach. Future work can enhance the performance of the approach by considering some more parameters according to the Requirement of the organization. Secondly, in the present work, author has taken only Computer Science and engineering department faculties. Future work can be enhanced for all branches of the college for overall performance enhancement of the college.

5. REFERENCES

[1] N.Aggrowal et.al,"Analysis the effect of data mmmg techniques on Database" Elsevier, Advances in Engineering Software 47 (2012) 164-169.
 [2] N.Ahmad and S.Shamsuddin," A Comparative Analysis of Mining Techniques for Automatic Detection of Students Learning Styles," IEEE, 978-1-4244-8136-1 (2010).
 [3] Z.Abdullah et.al," Extracting highly positive association rules from students enrollment data", Elsevier, Procedia - Social and Behavioral Science 28 (2011) 107-111.
 [4] K.Bunkar et.al," Data Mining: Prediction for

Performance Improvement of Graduation Student using Classification", IEEE, 978-4673-1989-8 (2012).

[5] B.Bidgoli et.al,"Predicting Student Performance: An Application of Data Mining Methods with an Educational Web-Based System", 33rd ASEE/IEEE Frontiers in Education Conference T2A-13, November 5-8, 2003.

[6] N.Delavari et.al,"A New Model for Using Data Mining Technology in Higher Education Systems", IEEE, 0-7803-8596-9 (2004).

[7] E.Garcia and P.Mora,"Model Prediction of academic performance for first year students", IEEE, 978-0-7695-4605-6 (2011).

[8] S.Liao et.al,"Data Mining techniques and applicationsA decade review from 2000-2011",Elsevier, Expert System with Application 39 (2012) 11303-11311.

[9] LiPing and DuanFu,"The Disposal of Incomplete Classification Data in Teaching Evaluation System", IEEE, 2009 Third International Symposium on Intelligent Information Technology Application, 978-7695-3859-4

[10] X. Long and y.wu,"Application of decision tree in student achievement evaluation", IEEE, 2012

International Conference on Computer Science and Electronics Engineering, 978-7695-4647-6.

[11] S. Mardikyan and B. Badur,"Analysis Teaching Performance of Instructors Using Data Mining Techniques" ,Information in Education, 2011, Vol. 10, No.2,245-257.

[12] P.Qingxian et.al,"data mining and application of teaching evaluation based on association rules", IEEE, Proceeding of 2009 4th International Conference on Computer Science and Education, 978-4244-3521-0

[13] Al.Radaishel et.al,"Mining student data using decision tree", in the proceeding of the International Arab conference on information technology, Jordan, 2006.

[14] C.Romero and S.Ventura", Educational data mining: A survey from 1995 to 2005". Elsevier, Expert System with Application 33 (2007) 135-146.

[15] C.Singh et.al,"Extraction and Analysis of Faculty Performance of Management Discipline from Student Feedback Using Clustering and Association rule mining techniques", IEEE, 978-4244-8679-3 (2011).

[16] K. Shaban and M. Abdulwahed,"Research-Based Learning in Computing Courses for Senior Engineering Students", IEEE International Conference on Teaching, assessment and Learning for Engineering (TALE) 2012, 978-1-4673-2418-2.

[17] H.Wang and R.Zhong,"Application of Association Rules in College Teaching",IEEE, 2011 International Conference on Internet Computing and Information Services, 978-0-7695-4539-4.