GNITED MINDS
Journals

EDUCATIONAL DATA MINING: PREDICTION OF STUDENTS' PERFORMANCE TO THEIR IMPROVEMENT

Educational Data Mining: Prediction of Students' Performance to Their Improvement

Shweta Singh[1] Dr. Pankaj Kumar[2]

[1]Research Scholar, Shridhar University, Rajasthan

[2]Asst. Prof. (Reader), Dept. Of Computer Science & Engineering, Shri Ram College of Engineering & Management (SRMCEM), Lucknow

Abstract – Educational data mining concerns with developing methods for discovering knowledge from data that come from educational domain In this paper we used educational data mining to improve graduate students' performance, and overcome the problem of low grades of graduate students. After preprocessing the data, we applied data mining techniques to discover association, classification, and clustering and outlier detection rules. In each of these four tasks, we present the extracted knowledge and describe its importance in educational domain.

Keywords: Educational, Data Mining, Developing, Techniques, Improve, Students, Performance, etc.

- - - - - - - - - - - - - - X - - - - - - - - - - - - - -

1. INTRODUCTION

Data mining showed what kind of data could be collected, how could we preprocess the data, how to apply mining methods on the data, and finally how can we benefited from the discovered knowledge. There are many kinds of knowledge can be discovered from the data. In this work we investigated the most common ones which are association rules, classification, and clustering and outlier detection (Kotas, 2000). The Rapid Miner software is used for applying the methods on the graduate student's data set. From The discovered knowledge, we need to provide a college management with a helpful and constructive recommendation to overcome the problem of low grades of graduate students, and to improve students' academic performance. Official approval from the college of science and technology was obtained to have an access to the related databases for the sole use of analysis and knowledge discovery purposes. To achieve privacy, all individual and personal data are extracted from the database before applying the data mining methods (Pascarella, 2004). Although, using data mining in higher education is a recent research field, there are many works in this area. That is because of its potentials to educational institutes. Romero and Ventura, have a survey on educational data mining between 1995 and 2005. They concluded that educational data mining is a promising area of research and it has a specific requirements not presented in other domains. Thus, work should be oriented towards educational domain of data mining. El-Halees, gave a case study that used educational data mining to analyze students' learning behavior. The goal of his study is to show how useful data mining can be used in higher education to improve student' performance He used students' data from database course and collected all available data including personal records and academic records of students, course records and data came from e-learning system (Albertelli, *et. al.,* 2002). Then, he applied data mining techniques to discover many kinds of knowledge such as association rules and classification rules using decision tree. Also he clustered the student into groups using EM clustering, and detected all outliers in the data using outlier analysis. Finally, he presented how can we benefited from the discovered knowledge to improve the performance of student applied the data mining techniques, particularly classification to help in improving the quality of the higher educational system by evaluating student data to study the main attributes that may affect the student performance in courses (Kashy, *et. al.,* 1995).

2. REVIEW OF LITERATURE:

The extracted classification rules are based on the decision tree as a classification method; the extracted classification rules are studied and evaluated. It allows students to predict the final grade in a course under study. Applied the classification as data mining technique to evaluate student' performance, they used decision tree method for classification. The goal of their study is to extract knowledge that describes students' performance in end semester examination (Woods, *et. al.,* 1995). They used students' data from the student' previous database including Attendance,

Class test, Seminar and Assignment marks (Duda, *et. al.,* 2001). This study helps earlier in identifying the dropouts and students who need special attention and allow the teacher to provide appropriate advising. Applied the classification as data mining technique to predict the numbers of enrolled students by evaluating academic data from enrolled students to study the main attributes that may affect the students' loyalty (number of enrolled students). The extracted classification rules are based on the decision tree as a classification method, the extracted classification rules are studied and evaluated using different evaluation methods. It allows the University management to prepare necessary resources for the new enrolled students and indicates at an early stage which type of students will potentially be enrolled and what areas to concentrate upon in higher education systems for support, applied the association rule mining analysis based on students' failed courses to identify students' failure patterns (Kohavi, 1995). The goal of their study is to identify hidden relationship between the failed courses and suggests relevant causes of the failure to improve the low capacity students' performances. The extracted association rules reveal some hidden patterns of students' failed courses which could serve as a foundation stone for academic planners in making academic decisions and an aid in the curriculum re-structuring and modification with a view to improving students' performance and reducing failure rate (Lim, *et. al.,* 2000).Used k-means clustering algorithm as a data mining technique to predict students' learning activities in a students' database including class quizzes, mid and final exam and assignments. This correlated information will be conveyed to the class teacher before the conduction of final exam. This study helps the teachers to reduce the failing ratio by taking appropriate steps at right time and improve the performance of students [Quinlan, 1986. Freitas, 2002).

3. THE GRADUATE STUDENTS DATA SET AND PREPROCESSING:

The data set used in this paper contains graduate students information collected from the college of Science and Technology – Khanyounis for a period of fifteen years in period from 1993 to 2007. The graduate student's data set consists of 3360 record and 18 attribute. Table presents the attributes and their description that exists in the data set as taken from the source database. The college of Science and Technology in Khanyounis grants their graduates a bachelor degree and diploma in twenty four technical scientific specialty, including three specialties for bachelor degree in Information Technology, Engineering of Buildings, and Medical Laboratory Sciences, and twenty one specialty for diploma which are English Language, Medical Laboratories, Pharmacy, Health monitoring, Secretarial and medical record, Software and databases, Programming and Systems Analysis, Computer networks and the Internet, Architecture, Civil Engineering, Computer Engineering, Technology of Office Equipment's,

Technology of radio and television, Business Administration, Accounting, Financial management and banking, Marketing and Management Products, and Management and office automation, Office Management and Secretarial, Interior Design and Graphical Design. As part of the data preparation and preprocessing of the data set and to get better input data for data mining techniques, we did some preprocessing for the collected data before loading the data set to the data mining software, irrelevant attributes should be removed [Ching, *et. al.,* 1995. Dan, Colla, 1998. Breiman, *et. al.,* 1984. Jain, Zongker, 1997).

4. APPLICATION OF DATA MINING TECHNIQUES TO GRADUATE STUDENT'S DATASET:

We applied data mining techniques to discover knowledge. Particularly we discovered association rules and we sorted the rules using lift metric. Then we used two classification methods which are Rule Induction and Naïve Bayesian classifier to predict the Grade of the graduate student. Also we clustered the students into groups using K-Means clustering algorithm (Loh, Shih, 1997). Finally, we used outlier detection to detect all outliers in the data, two outlier methods are used which are Distance-based Approach and Density-Based Approach. Each one of these tasks can be used to improve the performance of graduate student. Our future work include applying data mining techniques on an expanded data set with more distinctive attributes to get more accurate results (Shih, 1999). Also, experiments could be done using more data mining techniques such as neural nets, genetic algorithms, k-nearest Neighbor, and others. Finally, the used preprocess and data mining algorithms could be embedded into the college system so that anyone using the system can benefit from the data mining techniques (Clear Software, Inc. 1996. Falkenauer, 1998. Park, Song, 1998. Michalewicz, 1996. De Jong, *et. al.,* 1993. Murthy, 1998).

5. CONCLUSION:

In this paper we found that how useful data mining can be used in education particularly to improve graduate students' performance. We clustered the students into groups using K-Means clustering algorithm. Finally, we used outlier detection to detect all outliers in the data, two outlier methods are used which are Distance-based Approach and Density-Based Approach. Each one of these tasks can be used to improve the performance of graduate student. Our future work include applying data mining techniques on an expanded data set with more distinctive attributes to get more accurate results.

Shweta Singh[1] Dr. Pankaj Kumar[2]

REFERENCES:

Albertelli, G., Minaei-Bigdoli, B., Punch, W.F., Kortemeyer, G., and Kashy, E., (2002). "Concept Feedback In Computer-Assisted Assignments", Proceedings of the (IEEE/ASEE) Frontiers in Education conference.

Breiman, L., Freidman, J.H., Olshen, R. A., and Stone, P. J. (1984). Classification and Regression Trees. Belmont, CA: Wadsworth International Group.

Ching, J.Y. Wong, A.K.C. and Chan, C.C. (1995). "Class Dependent Discretisation for Inductive Learning from Continuous and Mixed-mode Data". IEEE Transaction. PAMI, 17(7), pp. 641 - 645.

CLEAR Software, Inc. (1996). allCLEAR User's Guide, CLEAR Software, Inc, 1996. Wells Avenue, Newton, MA.

Dan, S.; and Colla. P., (1998) CART--Classification and Regression Trees. San Diego, CA: Salford Systems.

De Jong K.A., Spears W.M. and Gordon D.F. (1993). Using genetic algorithms for concept learning. Machine Learning 13, pp. 161-188.

Duda, R.O., Hart, P.E., and Stork, D.G. (2001). Pattern Classification. 2nd Edition, John Wiley & Sons, Inc., New York NY.

Falkenauer E. (1998). Genetic Algorithms and Grouping Problems. John Wiley & Sons.

Freitas, A.A. (2002) "A survey of Evolutionary Algorithms for Data Mining and Knowledge Discovery", In: A. Ghosh and S. Tsutsui. (Eds.) Advances in Evolutionary Computation, pp. 819-845. Springer-Verlag.

Jain, A.K.; Zongker, D.; (February 1997). "Feature Selection: Evaluation, Application, and Small Sample Performance" IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 19, No. 2.

Kashy, E, Gaff, S, J, Pawley, N, H, Stretch, W, L., Wolfe, S, L., Morrissey, D.J., Tsai, Y., (1995) "Conceptual Questions in Computer-Assisted Assignments", American Journal of Physics, Vol, No 63, pp. 1000-1005.

Kohavi, Ron (1995) "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection", Proceeding of International Joint Conference on Artificial intelligence (IJCIA).

Kotas, P, (2000). "Homework Behavior in an Introductory Physics Course", Master's Thesis (Physics), Central Michigan University (2000).

Lim, T.-S., Loh, W.-Y. & Shih, Y.-S. (2000). "A Comparison of Prediction Accuracy, Complexity, and Training Time of Thirty-three Old and New Classification Algorithms". Machine Learning, Vol. 40, pp. 203—228. See http://www.stat.wisc.edu/~limt/mach1317.pdf)

Loh, W.-Y. & Shih, Y.- S. (1997). Split Selection Methods for Classification Trees, Statistica Sinica 7: pp. 815-840.

Michalewicz Z. (1996). Genetic Algorithms + Data Structures = Evolution Programs. 3rd Ed. Springer-Verlag.

Murthy, S. K. (1998). "Automatic construction of decision trees from data: A multidisciplinary survey", Data Mining and Knowledge Discovery, vol. 4, pp. 345—389.

Park Y and Song M. (1998). A genetic algorithm for clustering problems. Genetic Programming 1998: Proceeding of 3rd Annual Conference, pp. 568-575. Morgan Kaufmann.

Pascarella, A, M, (2004). "The Influence of Web-Based Homework on Quantitative Problem-Solving in a University Physics Class", NARST Annual Meeting Proceedings.

Quinlan, J. R. (1986), Induction of decision trees. Machine Learning, 1: pp. 81-106.

Shih, Y.-S. (1999). "Families of splitting criteria for classification trees", Statistics and Computing, Vol. 9, pp. 309-315.

Woods, K.; Kegelmeyer Jr., W.F.; Bowyer, K. (1995) "Combination of Multiple Classifiers Using Local Area Estimates"; IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 4.

Shweta Singh[1] Dr. Pankaj Kumar[2]