A Systematic Review on the Educational Data Mining and its Implementation in the Applications

Dr. S. Saravana Kumar
Assistant Professor, Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore
saravanan.msc16@gmail.com

***Abstract* —** The Mining of the education data is emerging trend in the learning analytics as it is time consuming to analyse the data and to identify the hidden information automatically. In this paper, detailed investigation about the educational data mining technique is carried out. The Application of the data mining includes storing and retrieval of the student data in the large repositories such as mark sheet, attendance sheet and student profile etc. The importance analysis is carried out on the retrieval of the large data using machine learning algorithm in the data mining. Along the retrieval of the data, nowadays deep focus is made on predicting and recommendation models which provide more effectiveness to the educational applications in terms of suggestions and extracting the correlation among the students. However, handling of large data from repositories leads to performance bottleneck, hence it is resolved by employing Map Reduce Paradigm from big data analytics. Through extensive study, classification and clustering provides more value for the data management, hence semantic and opinion mining is presented as the future research solution.

***Keywords* —** Educational data mining, Clustering, Classification, Data Prediction, Recommendation, Map Reduce, Semantic and Opinion Mining

I. INTRODUCTION

Educational Data Mining (EDM) is a field which uses data mining to extract the data from large repositories using knowledge discovery model to generate information [1]. It is also used to determine the data patterns, organize the information of the hidden relationship, estimate the unknown labels, predict the value to classify the object and compose a new cluster to the homogenous objects [2]. The Major focus of the EDM is made on knowledge discovery, decision making and recommendations. EDM emerges as learning paradigm algorithms for exploring data to find out patterns and make predictions. The objective of the EDM is to develop models to upgrade the learning experience and institutional effectiveness. One of the pre-processing algorithms of EDM is known as Clustering. It is an unsupervised approach for analyzing data [3]. It refers to collecting similar objects together to form a group or cluster. Each cluster contains objects that are similar to each other but dissimilar to the objects of other groups [4]. Data clustering enables academicians to predict student performance, correlate learning styles of different learner types and their behaviors and collectively improve upon institutional performance. Though handling of large data from repositories leads to performance bottleneck which need to big data solutions, hence it is resolved by employing Map Reduce Paradigm from big data analytics [5]. The predictive and recommendation algorithms work on grade and enrolment data to suggest the courses to student based on their interest predicted [6]. The rest of the paper is organized as follows section 2 describe the review of literature related to study, it then continuous to provide the future solution in the section 3 and finally section 4 provides the conclusion.

II. REVIEW OF LITERATURE

2.1. Educational data mining and its role in determining factors affecting students' academic performance

In this literature, various supervised data mining techniques used in problem solving and decision making skills of individuals. The Supervised algorithms are to improve the overall quality and effectiveness of education. Clustering techniques have broadly been classified into two types, hierarchical and partitional. Centroid-based clustering, Graph-based clustering, Grid-based clustering, Density-based clustering, neural networks-based clustering can be used to predict the student's characteristics on various on various aspects. Clustering algorithms are also applied to voluminous data sizes of the student data if it exceeds the particular limit. The concept of big data refers to voluminous, enormous quantities of data [7].

2.2. Clustering Methods in Educational Data Mining

The literature discusses various clustering technique using hierarchical and non-hierarchical methods to differentiate the groups of students, according to their interaction and performance characteristics. There are different learning models to perceive the groups obtained, to determine a similarity between the results,

confirming the acquired knowledge from the clustering and demonstrating that the choice of method on the knowledge obtained from interactions and students performance on the course [8].

2.3. A Data Prediction model to predict school Failure and Dropout

This literature applies data prediction algorithms to predict school failure and drop outs. Decision Tree Algorithms have employed to predict the student has possibility to fail in the central examinations and drop out due to non performing attitude in their studies. The high dimensional data is taken for processing; in this feature reduction algorithm is also employed to remove the unwanted attributes in the dataset, after which class label employed to that category of data. The data normalization is also employed on the feature reduced attributes in the resultant dataset [9].

2.4. Mutual reinforcement of prediction and recommendation on academic performance

This literature identifies the importance of prediction and recommendation in the academic performance. It is one of the most important tasks in educational data mining and has been widely used in intelligent tutoring systems. Academic performance could be affected with factors like personality, skills, social environment and the use of library books and so on. The supervised content-aware matrix factorization for mutual reinforcement of academic performance prediction and library book recommendation is analysed. This model not only addresses the sparsity challenge by employing dimension reduction techniques, but also promotes recommendation by recommendation using frequent item mining algorithm named as Apriori algorithm for students based on their performance levels and book meta information.

2.5. Curriculum Generation based on Student Preference and profiles – Recommendation Model

The Learning objects classification and prediction a challenging topic because of its direct application to curriculum generation based on Student Preference and profiles. It is recommended by employing the FP growth Algorithm by incorporating personalization of contents, the pedagogical requirements, and specific necessities of each student. This literature presents a general and effective approach to extract information from the e-learning contents, a form of reusable learning objects, to generate a planning domain in a simple, automated way. Such a domain is used by an intelligent planner that provides an integrated recommendation system, which adapts, stores, and reuses the best learning routes according to the students' profiles and course objectives.

3. Future Solutions

Enabling the semantic mining and opinion mining on the current learning model improves the adaptive learning system. Education content representation as semantics provides better retrieval efficiency. The semantic can be presented in terms of conceptual metadata representation. Automated acquisition of hierarchical relationships between relevant domain terms from educational content, which constitutes a fundamental step in the semantic composition of an educational course. It improves the speed of the domain model creation. Also it reduces the overall difficulty of the task and finally it has significant improvement in the quality of resulting domain models.

4. Conclusion

This paper has presented the detailed review of the Educational data mining carried out in the areas of classification, clustering, prediction and recommendation models. In Summary, key advantage of the clustering provides the grouping of heterogeneous objects using the learning models. The Classification algorithms are used to categorize the objects of the cluster. Prediction and recommendation algorithm are used to suggest the students with future outcomes in order to improve overall performance and enhance the student retention effort. This paper also highlights the future solutions of the EDM in terms of employing semantic mining and opinion mining to improve the performance of present solutions.

REFERENCES

[1] J. Vandamme, -P.,Meskens, N., Superby, F.-, J, "Predicting Academic Performance by Data Mining Methods," Education Economics, vol. 15, pp. 405-419, 2007.

[2] R. S. Baker and K. Yacef, "The state of educational data mining in 2009: A review and future visions," JEDM-Journal of Educational Data Mining, 2009.

[3] O. R. Zaıane, "Building a recommender agent for e-learning systems," in Computers in Education, 2002. Proceedings. International Conference on, 2002, pp. 55-59.

[4] T. S. Madhulatha, "An overview on clustering methods," arXiv preprint arXiv:1205.1117, 2012.

[5] J. Manyika, M. Chui, B. Brown, and J. Bughin, "Big data: The next frontier for innovation, competition, and productivity," 2011.

[6] S. Parack, Z. Zahid, and F. Merchant, "Application of data mining in educational databases for

predicting academic trends and patterns," in Technology Enhanced Education (ICTEE), 2012 IEEE International Conference on, 2012, pp. 1-4.

[7] Leena Khanna et.al "Educational data mining and its role in determining factors affecting students academic performance: A systematic review" in Information Processing (IICIP), 2016 1st India International Conferenc

[8] Jorge Luis Cavalcanti Ramos et.al "A Comparative Study between Clustering Methods in Educational Data Mining"in IEEE Latin America Transactions vol.14, issue ;8, Aug 2016

[9] Carlos Marquez-Vera et.al "Predicting School Failure and Dropout by Using Data Mining Techniques" in IEEE Latin American journal of learning technologies, vol: 8, issue: 1, Feb2013

[10] Antonio Garrido et.al " E-Learning and Intelligent Planning: Improving Content Personalization" in IEEE Ibero-American Review of Learning Technologies, Vol: 9, issue: 1, feb 2014.