© 2018 IJSRSET | Volume 4 | Issue 4 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section : Engineering and Technology

A Systematic Review on Educational Data Mining

M. Manjula

P. G. Student, Department of Computer Science and Engineering, Adiyamaan College of Engineering, Hosur, Tamil Nadu, India

ABSTRACT

In this paper, implementing K-Means clustering algorithm for analyzing the particular dataset and data mining. The main purpose is WEKA process. In Weka process we can get perfect graph, accuracy and random process. The Pre-processing was important concept it may clear a null values, removes a unwanted data and unwanted memory space. In Data mining analyzing data set. In Data mining implementing two methods classification, clustering process. By using classification, clustering we get flexible result and large amount of database. Here, weka process and K-means algorithm going to compare whether both graphs are accurate manner.

Keyword : K-means clustering, Weka process, Classification, Cluster process.

I. INTRODUCTION

Educational Data Mining (EDM) applies machine-learning, statistics, Data Mining (DM), psycho-pedagogy, information retrieval, cognitive psychology, and recommender systems methods and techniques to various educational data sets so as to resolve educational issues. The International Educational Data Mining Society defines EDM as “an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in” EDM is concerned with analysing data generated in an educational setup using disparate systems. Its aim is to develop models to improve learning experience and institutional effectiveness. While DM, also referred to as Knowledge Discovery in Databases (KDDs), is a known field of study in life sciences and commerce, yet, the application of DM to educational context is limited One of the pre-processing algorithms of EDM is known as Clustering. It is an unsupervised approach for analysing data in statistics, machine learning, pattern recognition, DM, and bioinformatics. 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.

This approach when applied to analyse the dataset derived from educational system is termed as Educational Data Clustering (EDC). An educational institution environment broadly involves three types of actors namely teacher, student and the environment. Interaction between these three actors generates voluminous data that can systematically be clustered to mine invaluable information. Data clustering enables academicians to predict student performance, associate learning styles of different learner types and their behaviours and collectively improve upon institutional performance. Researchers, in the past have conducted studies on educational datasets and have been able to cluster students based on academic performance in examinations various methods have been proposed, applied and tested in the field of EDM. It is argued that these generic methods or algorithms are not suitable to be applied to this emerging discipline. It is proposed that EDM methods must be different from the standard DM methods due to the hierarchical and non-independent nature of educational data . Educational institutions are increasingly being held accountable for the academic success of their students Notable research in student retention and attrition rates has been conducted by Luan. For instance, Lin applied predictive modelling technique to enhance student retention efforts. There exist various software's like Weak, Rapid Miner, etc. that apply a combination of DM algorithms to help researchers and stakeholders find answers to specific problems.

II. LITERATURE REVIEW

1. J. Luan, "Data mining applications in higher education," SPSS Executive (2004): In this paper, most challenge is higher learning institute, it is improve quality Managerial decision. By using this Managerial decision making they increased Educational entities and more complex. In Educational entities more efficient technology, Managerial decision making support new strategies current process. By addressing the new knowledge it can improve the educational process and managerial. In Data Mining technique used analysis tool so they extract the knowledge of large dataset. 2. C. Romero and S. Ventura "Educational data mining: A survey from 1995 to 2005" (2007): Today most important role is higher education for human begin. Educational data mining have certain development for unique types of data education setting, whether the student have proper education. Educational data mining methods are Prediction, Clustering, Relationship Mining, and Discovery with model, Distillation of data for human judgement. Even though they have some Drawbacks: Course outline formation, teacher student understanding and high output. 3. M.F.M. Mohsen, N.M. Norwawi, C.F. Hibadullah, and M.H.A. Wahhabi "Mining the student programming performance using roughset" (2010): In this they analysis the programming data set by using Roughset, they investigate student programming data based on previous student performance. The results were compare to data using statistic, clustering and association rules. Rough set concept is defined by topologies operation, interior and closure. 4. J. E. Beck and B. P. Woolf "High-level student modelling with machine learning" (2000): In this paper they constructed the student of behaviour high-level of mathematic tutor and going to analysis whether students have particular knowledge, learning about student can answer a problem correctly. To construct this model they used machine learning agent. This agent is used for gather about student information and current topics. It is very good for offline learning. The main large drawback is online learning.

III. EXISTING SYSTEM

In an Existing system using many algorithm concept to perform the mining concept. It was easy to perform the mining operations. The Accessing speed is high in Educational Data Mining applies machine-learning, statistics, Data Mining(DM), psycho-pedagogy, information retrieval, cognitive psychology, and recommender systems methods and techniques to various educational data sets so as to resolve educational issues. The International Educational Data Mining Society defines EDM as "an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in". EDM is concerned with analyzing data generated in an educational setup using disparate systems. Its aim is to develop models to improve learning experience and institutional effectiveness. While DM, also referred to as Knowledge Discovery in Databases (KDDs), is a known field of study in life sciences and commerce, yet, the application of DM to educational context is limited.

One of the pre-processing algorithms of EDM is known as Clustering. It is an unsupervised approach for analyzing data in statistics, machine learning, pattern recognition, DM, and bioinformatics. 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. This approach when applied to analyze the dataset derived from educational system is termed as Educational Data Clustering (EDC). An educational institution environment broadly involves three types of actors namely Teacher, student and the environment. Interaction between these three actors generates voluminous data that can systematically be clustered to mine invaluable information. Data clustering enables academicians to predict student performance; associate learning styles of different learner types and their behaviors and collectively improves upon Institutional performance. Researchers, in the past have conducted studies on educational datasets and have been able to cluster students based on academic performance in examinations.

Disadvantages

- ✓ In a Data mining concept particularly any algorithms not using at educational purpose. So, it was one of the drawback of to accessing educational data's.
- ✓ To accessing educational data's the mining process is slow.
- ✓ The basic concepts and basic algorithms only using the educational mining process.
- ✓ Traditional data mining algorithms cannot be directly applied to educational problems.

IV. PROPOSED SYSTEM

This process is implement that a preprocessing algorithm has to be enforced first and only then some specific data mining methods can be applied to the problems. One such preprocessing algorithm in EDM is Clustering. Many studies on EDM have focused on the application of various data mining algorithms to educational attributes. Therefore, this paper provides over three decades long (1983-2016) systematic literature review on clustering algorithm and its applicability and usability in the context of EDM. Future insights are outlined based on the literature reviewed, and avenues for further research are identified. The EDM process converts raw data coming from educational systems into useful information that could potentially have a greater impact on educational research and practice" Traditionally, researchers applied DM methods like clustering, classification, association rule mining, and text mining to educational context. A survey conducted in 2007, provided a comprehensive resource of papers published between 1995 and 2005 on EDM by Romero & Ventura. This survey covers the application of DM from traditional educational institutions to web-based learning management system and intelligently adaptive educational hypermedia systems. In another prominent EDM survey by Pena-Ayala, about sample works published between 2010 and 2013 were analyzed. One of the key findings of this survey was that most of the EDM research works focused on three kinds of educational systems, namely, educational tasks, methods, and algorithms.

Application of DM techniques to study on-line courses was suggested by the proposed a non-parametric clustering technique to mine offline web activity data of learners. Application of association rules and clustering to support collaborative filtering for the development of more sensitive and effective e-learning systems was studied by the researchers Baker, Corbett & Wagner conducted a case study and used prediction methods in scientific study to game the interactive learning environment by exploiting the properties of the system rather than learning the system. Similarly, Brusilovsky Pylon provided tools that can be used to support EDM. In their study Beck & Woolf showed how EDM prediction methods can be used to develop student models. It must be noted that student modeling is an emerging research discipline in the field of EDM While another group of researchers, Garcia at al devised a toolkit that operates within the course management systems and is able to provide extracted mined information to non-expert users. DM techniques have been used to create dynamic learning.

Advantages

- ✓ Mining process speed is high.
- ✓ Easily access the educational and educational types of all data's.
- ✓ This process may fully concentrate the education purpose.
- ✓ Using (EDM) process.
- ✓ So the pre-processing speed is high.

V. SYSTEM ARCHITECTURE

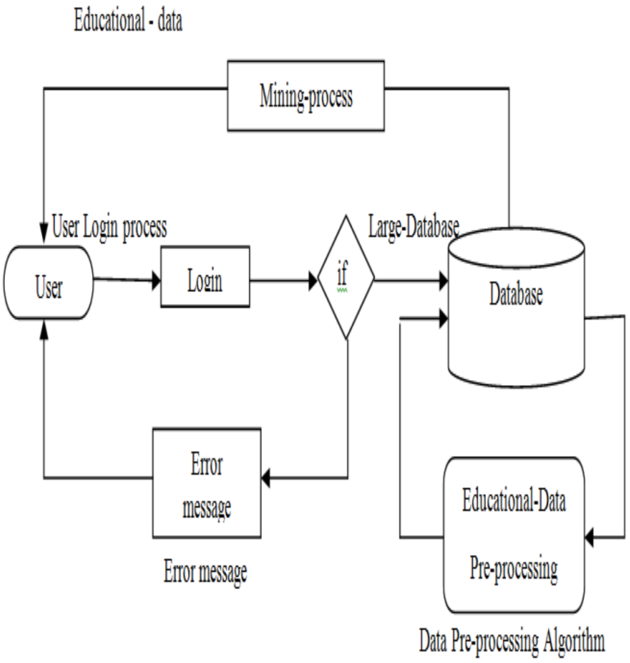

Figure 1. System Architecture

VI. MODULES

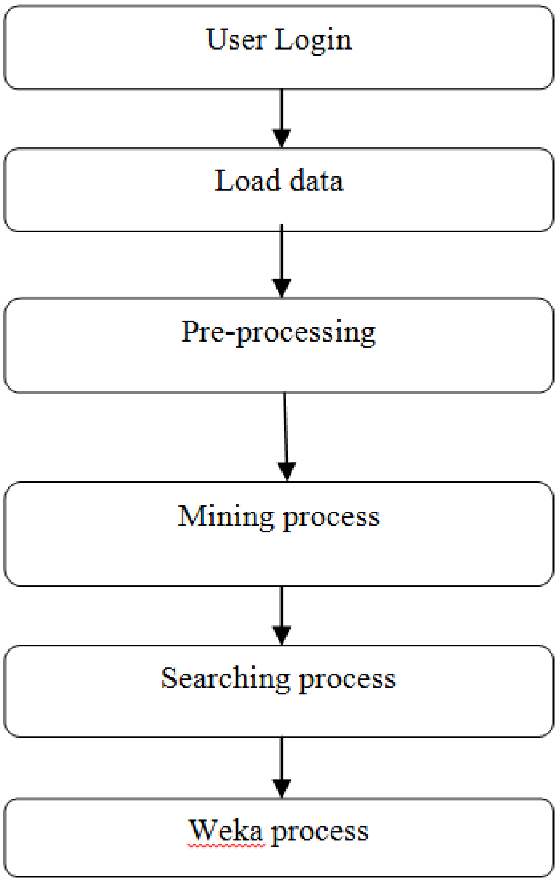

Figure 2. Flow Diagram

6.1 User Login
6.2 Load Data
6.3 Pre-processing
6.4 Mining Process
6.5 Searching Process
6.6 Weka Process

6.1 USER LOGIN-PROCESS

This module was important for user to access the database. This process was most secure process the authentication person only access the data's. In this process the user give the correct username and password. If the name and password was wrong the user does not access his/he own data's. In large database the security purpose was necessary and more important.

6.2 LOAD DATA

In this process load the education relevant data's. The data's may upload the large amount of database.

6.3 PRE-PROCESSING

Pre-processing is one of main modules for data mining system. Here we are removing unwanted data or null values and unstructured data. So when we remove unstructured data's then only we get accurate results for given dataset. It is particularly applicable to data mining and machine learning projects. Data-gathering methods are often loosely controlled, resulting in out-of-range values. Impossible data combinations, missing values, etc. Analysing data that has not been carefully screened for such problems can produce misleading results.

6.4 MINING PROCESS

Data mining is the computing process of discovering pattern in large data set involving methods at the intersection of machine learning, statistic and database system. The goal of data mining process is extract information from dataset.

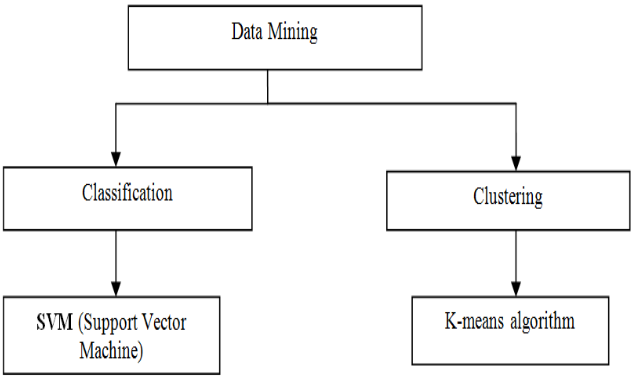

Figure 3. Data Distribution

Data mining have two methods

1. Classification
2. Clustering

6.1.1 CLASSIFICATION

Classification is the data mining (machine learning) technique used for group of membership.

SVM (SUPPORT VECTOR MACHINE)

Support Vector machine (SVM) is a machine learning algorithm and supervised learning algorithm that analyze data used for classification and regression analysis. SVM is powerful, strong theoretical and strong regularization properties.

6.1.2 CLUSTERING PROCESS

Clustering can be considered the most important unsupervised learning problem so, as every other problem of this kind; it deals with finding a structure in a collection of unlabeled data. In this mining process may using some clustering algorithm's to may using this process. Example for clustering algorithms (wards method).

WARDS METHOD

Wards method is a criterion applied in hierarchical cluster analysis.

6.5 SEARCHING PROCESS

In this module the user may searching the important data's. This module may gives the correct formatted data's and cleaned data's information's. The user viewing process may include in this process.

6.6 WEKA (Waikato Environment for Knowledge Analysis) PROCESS

Weka is the collection of machine language algorithm for data mining. The algorithm can either be applied directly to a dataset or own java code. Weka contains tools for Pre-processing, Classification, Clustering, Association rules and Visualization.

Advantage

- ✓ Weka fully implemented in java programming language, protable and platform independent.
- ✓ It is freely available under GUI General Public License.
- ✓ Weka software contain very graphical user interface, so system can easily access.

VII. CONCLUSION

This paper has presented over three decade's systematic review on clustering algorithm and its applicability and usability in the context of EDM.

This paper has also outlined several future insights on educational data clustering based on the existing literatures reviewed, and further avenues for further research are identified. In summary, the key advantage of the application of clustering algorithm to data analysis is that it provides relatively an unambiguous schema of learning style of students given a number of variables like time spent on completing learning tasks, learning in groups, learner behaviour in class, classroom decoration and student motivation towards learning. Clustering can provide pertinent insights to variables that are relevant in separating the clusters. Educational data is typically multi-level hierarchical and non-independent in nature, as suggested by Baker &Yosef [6] therefore a researcher must carefully choose the clustering algorithm that justifies the research question to obtain valid and reliable results.

VIII. REFERENCES

[1]. C.Romero and S.Ventura, "Educational data mining: a review of the state of the art," Systems,Man,and Cybernetics,Part C: Applications and Reviews, IEEE Transactions on, vol.40, pp.601-618, 2010.

[2]. (2011,01 July). International Educational Data Mining Society. Available: http://www.educationaldatamining.org/

[3]. J. Ranjan and K. Malik,"Effective educational process: a data-mining approach," Vine, vol.37, pp.502-515, 2007.

[4]. V. P. Bresfelean,M. Bresfelean,N. Ghisoiu,and C. A. Comes,"Determining students' academic failure profile founded on data mining methods," presented at the ITI 2008 - 30th International Conference on Information Technology Interfaces, 2008.

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

[6]. 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.

[7]. J. P. Campbell,P. B. DeBlois,and D. G. Oblinger,"Academic analytics: A new tool for a new era," Educause Review, vol.42, p.40,2007.

[8]. J. Luan,"Data mining applications in higher education," SPSS Executive, vol.7, 2004.

[9]. S.H.Lin, "Data mining for student retention management," Journal of Computing Sciences in Colleges, vol.27, pp.92-99, 2012.

[10]. T. Denley,"Austin Peay State University: Degree Compass," EDUCAUSE Review Online., 2012.

[11]. M. F. M. Mohsin,N. M. Norwawi,C. F. Hibadullah,and M. H. A. Wahab,"Mining the student programming performance using rough set," presented at the Intelligent Systems and Knowledge Engineering (ISKE), 2010 International Conference on, 2010.

[12]. C. Romero and S. Ventura,"Educational data mining: A survey from 1995 to 2005," Expert Systems with Applications, vol.33, pp.135-146,7/ 2007.

[13]. A.Pena-Ayala, "Educational data mining: A survey and a data mining-based analysis of recent works," Expert Systems with Applications, vol. 41, pp.1432-1462, 3/ 2014.

[14]. O.R.Zaıane and J.Luo, "Web usage mining for a better web-based learning environment," in Proceedings of conference on advanced technology for education, 2001, pp.60-64.

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

[16]. R.S.Baker, A.T.Corbett, and A.Z.Wagner, "Off-task behavior in the cognitive tutor classroom: when students game the system," in Proceedings of the SIGCHI conference on Human factors in computing systems, 2004, pp.383-390.

[17]. P. Brusilovsky and C. Peylo,"Adaptive and intelligent web-based educational systems," International Journal of Artificial Intelligence in Education, vol.13, pp.159-172, 2003.

[18]. J.E.Beck and B.P.Woolf, "High-level student modeling with machine learning," Intelligent tutoring systems, pp.584-593, 2000.

[19]. E. Garcia,C. Romero,S. Ventura,and C. de Castro,"A collaborative educational association rule mining tool," The Internet and Higher Education, vol.14, pp.77-88, 2011.

[20]. Y. H. Wang and H. C. Liao,"Data mining for adaptive learning in a TESL-based e-learning system," Expert Systems with Applications, vol. 38, pp.6480-6485, 2011.

[21]. M. E. Zorrilla,E. Menasalvas,D. Marin,E. Mora,and J. Segovia,"Web usage mining project for improving web-based learning sites," in Computer Aided Systems Theory–EUROCAST 2005,ed: Springer,2005, pp.205-210.

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

[23]. A.K.Jain and R.C.Dubes, Algorithms for clustering data.Englewood Cliffs,NJ, USA: Prentice-Hall, Inc., 1988.

[24]. S.Sagiroglu and D.Sinanc, "Big Data: A Review," Proceedings of the 2013

[25]. B. Kitchenham,O. Pearl Brereton,D. Budgen,M. Turner,J. Bailey,and S. Linkman,"Systematic literature reviews in software engineering–a systematic literature review," Information and software technology, vol.51, pp.7-15, 2009.

[26]. H. M. Chen and M. D. Cooper,"Using clustering techniques to detect usage patterns in a Web-based information system," Journal of the American Society for Information Science and Technology, vol.52, pp. 888-904, 2001.

[27]. N.A.Rashid, M.N.Taib, S.Lias, N.Sulaiman, Z.H.Murat, and R.S. S.A.Kadir, "Learners' Learning Style Classification related to IQ and Stress based on EEG," Procedia - Social and Behavioral Sciences, vol. 29, pp.1061-1070, / 2011.