International Journal of Electrical Electronics & Computer Science Engineering
Special Issue - TeLMISR 2015, ISSN : 2348-2273
Available Online at www.ijeecse.com

A Comprehensive Study of Educational Data Mining

Jasvinder Kumar
Department of Computer Science, SGTBIMIT, GGSIP University, New Delhi, India
JassKumar2008@yahoo.com

Abstract: ***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. 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.***

Keywords: ***EDM, DM, PSCL, NCES, PRS, PLE ,KDD, KT, ITS, DSS.***

1. INTRODUCTION

In recent years, the ease and advancement in education domain has created a vast amount of data. As the volume of data increase so do the complexity and relationship underneath the data [1]. Exploring knowledge from large volume of data is the biggest challenge. Data mining which is also known as Knowledge Discovery in Databases (KDD) is a technology used in different disciplines to search for significant relationships among variables in large data sets. Data mining is mainly used in commercial applications. Now a days, the researcher have shown great interest in using data mining applications in the field of education to efficiently manage and extract undiscovered knowledge from the data [2]. The Educational Data Mining community website [3]defines educational data mining as follows: "Educational Data Mining is 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 an interdisplinary field which inherits the features from various domains such as Psychometrics, artificial Intelligence, Domain driven, Information retrieval, Machine learning, Learning analytics, Databases, Cognitive Psychology and so on. [4].The large repositories of data generated from different sources should be analyzed to fulfill the goals in education. The main objective of EDM viewed by different Researchers as [5][6]-

1. Student Modeling: it is related with the creation of student models that includes student behavior, learning style, performance and environment in which they can develop their skills and solve their problems.

2. Domain Modeling: it is related with the designing of methods, tools and techniques for the growth of particular branch/institution.

3. Learning System: developing the system for studying the effects of educational support. e.g. Pedagogical support.

4. Building the computational models for learning and learners that consist of students, domain.

5. Study the effects of resources related to infrastructure, human resource, and Industry-academic relationship in the organization.

To meet all the above mentioned objectives, a study of EDM is required for delivering the quality education. This objective of this paper is to provide brief knowledge of EDM to the researchers or non expert user in this field. It highlights the different modules of EDM and the opportunities for future research.

The paper is organized into Sections. Process of EDM is defined in Section 2. Different phases of EDM are described in Section 3.Future Research Directions are given in Section 4.Section 5 conclude the Introduction.

II. EDM PROCESS

The process of Educational data mining is an iterative, Knowledge discovery process which consists of Hypothesis formulation, Testing and refinement [4] (see Fig-1). Hypothesis is developed from various educational environments. It creates large volume of data. The main process of EDM starts with validating data (i.e. finding relationship between variables/parameters/data items). This is also known as preprocessing of data. After preprocessing various DM techniques, tools will be employed on processed data and final results/interpretation will be given to different user of education. Further recommendation will be suggested for the refinement of problems/task.

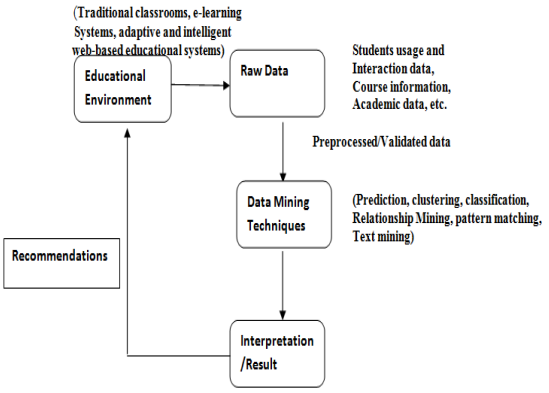

Fig-1 Process of EDM

International Journal of Electrical Electronics & Computer Science Engineering
Special Issue - TeLMISR 2015, ISSN : 2348-2273
Available Online at www.ijeecse.com

III. EDM MODULES

The main modules of EDM are User and Stake holders of Education, Tools, Techniques and Models of DM, Educational Data, Task and Results that will altogether used to achieve the objectives of EDM.

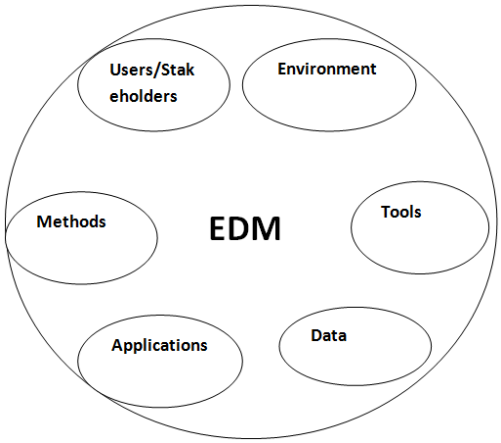

Fig- 2 Phases of EDM

A. Users and Stakeholders: According to researchers there are four main groups of users and stakeholders involved in accomplishing the specific objectives.

i) Learners (Student/Pupils): The main objective of Learners is to reply to student requisites, improve learning performance, and provide feedback or recommendations to learners.

ii) Educators (Teachers/Faculties/Tutors): Their objective is to analyze student behavior, learning, cognitive and social aspect and reflect on their own teaching methods.

iii) Educational Researchers: The main objective of researchers is to develop new tools and techniques for the growth of educational system.

iv) Administrators: The main objective is to utilize and enhance available resources (human and material) and their educational offer and so on.

B. EDM Methods: One of the essential modules of EDM is the methods of DM, used for different purpose. Romer Ventura [7] and Ryan Baker [8] categorized the methods as-

- Prediction
- Clustering
- Relationship mining

These methods are useful in mining web data and in mining other forms of educational data. Universally accepted across types of DM. Some of the methods which are acknowledged when validated relationships are applied to make predictions are-

- Distillation of data for human judgment
- Discovery with models
- Knowledge Tracing(KT)

i) Prediction: This technique is used to derive predicted variable(single variable) from predictor variables(combination of variables). Prediction is used analyze student performance and drop out.[9,10] and for detecting student behavior.[11]. It is classified into three types.

Classification: used to predict class label from (discrete or continue). Some popular classification methods include logistic regression, support vector machines and decision trees [12].

Regression: used to predict from continuous variable. Some popular regression methods within educational data mining include linear regression, neural networks [13].

Density Estimation: probability density function is used to predicted variable. Density estimator can be based on variety of kernel functions, including Gaussian function.

ii) Clustering: Clustering is an unsupervised classification process .it is used for grouping objects into classes of similar objects [14]. Data items are partitioned into groups or subsets (clusters) based on their locality and connectivity within N-dimensional space. In educational data mining, clustering has been used to group students according to their learning [15].

iii) Relationship mining: Relationship mining is used to determine relationship between variables in a data set and form rules for specific purpose. Relationship mining is classified into four types:

Association rule mining: This method is used to identify relationship between attributes in data set, extracting interesting correlations, frequent patterns among data items.[16] for finding students' mistakes often occurring together while solving exercises [17].

Correlation mining: This method is used to find Linear correlations between variables (positive or Negative). Correlation analysis is used to find the most strongly correlation attributes.

Sequential pattern mining: This method is used to find inter-session patterns such as the presence of a set of items followed by another items in a time-ordered set of sessions or episodes.[18] based on temporal relationship between variables to predict which group a learner belongs to. Wang et. al. proposes a four phase learning portfolio mining approach. [19]

Casual data mining: This method is used to find casual relationship between variables by analyzing the

covariance of two events or by using information about how one of the events was trigger.

Other Methods are:

Distillation of data for human judgment: The objective of this method is to present data in summarize and visualized way for e.g. (3D graph etc), to focus on appropriate information and support decision making. In EDM it is used for identification and classification [20].

Discovery with models: This type of model is used as component in other analysis such as relationship mining or prediction.[2].

Knowledge Tracing: This method is used to monitor student knowledge and skills over time. It is an effective method in cognitive tutor system [21]

C. EDM Application: Ryan baker[6] has divided the application areas into four major sections such as student Models, Knowledge Domain, Pedagogical support provided by learning software, Scientific discovery about learning and learners. For the development of academic and administrative sections of institutions, there are several applications or tasks that have been resolved through DM techniques. Some of them are summarized in table-1.

D. EDM Environment: These are the domains in which different users/stakeholders learn. EDM environments are classified into following three categories.

i) Traditional Class room Environment: It is a formal environment in which users of education communicate directly with each other(i.e. face-to face).for e.g. schools, colleges where lectures are delivered by teacher to students in classrooms.[4]

ii) Online/Web Based Environment: It is an informal environment in which users of education make use of internet. For e.g. e-learning [29], Web Based performance prediction [12].

iii) Computer Based Learning: It is hybrid environment of both (formal and informal interaction).In computer based Environment user can work -

Offline: Intelligent Tutoring System (ITS) [30] .learning Management [31], Online for e.g., e-learning [24, 29], Collaborative learning [32].

E. EDM Data: The large volume of data gathered from distributed and diverse fields are used for decision making and learning process in educational context. Data collected depends on the EDM environment discussed in section D.

i) Private data: Direct environment generates offline or private data (related to data collected from academic institution).

ii) Public data: Indirect environment generates online or public data (generally related to e-learning, web logs, e-mail, text data etc. EDM research is more feasible with the advent of public educational data repositories such as Pittsburgh Science of Learning Center DataShop (PSLC), National Center for Education Statistics (NCES) [33, 34].

F.EDM Tools: There are various tools for mining the repositories of data based on their usage, functionality, and working environment. [42].Different tools will be used on the basis of the respective goal such as given in table-2.

Table I. EDM Applications

Application	Objective	DM techniques	Reference
Course Management	To focus on the construction/development/selection of course curriculum which helps students to increase learning outcomes and success	Clustering algorithms, naïve algo, Rough set theory	[22], [23]
Predicting Performance of Students	To focus on knowledge, grade of student various researchers predict the performance of students based on academic and other psychometric factor	Regression(Continuous Variables), Classification(Discrete Variables, Bayesian networks, Neural Network, Decision tree etc.	[9], [12]
Personal Learning Environments / Recommender System	It provides various tools and services so that the system can adapt to student learning needs and recommend students directly to their personalized activities, next task and links to visit.	Association-rule mining, Clustering(PRS)/Neural Networks, Decision tree	[24]

Student Retention	Focus on finding students which are at-risk and their success factor	Classification ,Regression trees	[10], [25]
Social Network Analysis(SNA)	Focus on studying a relationship between group of people/organization/individual, rather than attributes or variables. The individual are related to each other on the basis of like friendship, cooperative relationship or informative exchange.	Collaborative filtering	[26]
Grouping/Profiling students	To find out the group of students which will be used by instructor to enhance group learning and build a personalized learning system.	Classification, Clustering, K-means, model based clustering	[27], [28]

Table II. EDM Tools

Tool	Goal	Reference
WEKA Tool	Used for developing Machine Learning Task	[35]
Moodle Tool	Help Users in Course Management System	[41]
Rapid Miner	Identify Student Behavior model in virtual courses	[40]
KEEL	Use to assess the behavior of evolutionary learning and soft computing based techniques.	[37], [45]
TADA-ED	To Identify patterns in student's online exercise	[39]
DataShop	To store and analyze Public data	[33], [36]
Decision Tool	To analyze factors related to success and failure of student	[38]

IV. OPPORTUNITIES FOR FUTURE WORK

[4][43][44] Researchers has share their experience and research opportunities in all aspect of EDM are

- Development of tools for protecting individual privacy.
- Development of system that reduce instructor intervention. such as(DSS ,PRS).
- Developing more generalized tools that can be used by expert and non expert user easily.
- Integration with e-learning.
- Standardization of data and models.

Plagiarism is most concerned topic among research scholar. Thus there should be predictive models to detect plagiarism. As, EDM is a growing field and involved in education technology we should find best ways to discover, learn, utilize technology that can improve learning, teaching and leading in the 21st century. Emphasis should be given on-

- Integrating Pedagogy support and data mining.
- Building model to improve personalized learning. For e.g. finding association between teachers and learners.
- Building tools to enhance Open Education Resource.
- Describing more ways of Flipped Classroom Concept (Model that has changed traditional classroom concept to modern Mobile learning).
- Developing models for building User Centric EContent (i.e. EBook).

V. CONCLUSION

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 .Utilization of data mining techniques within education environment requires a joint effort by the ICT specialists, educationists and the learners.

VI. REFERENCES

[1] Xindong Wu, Xingquan Zhu, Gong-Qing Wu and Wei Ding,"Data Mining with Big Data". IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 1, JANUARY 2014, 1041-4347.

[2] Barahate Sachin R., Shelake Vijay M," A Survey and Future Vision of Data mining in Educational Field", Second International Conference on Advanced Computing & Communication Technologies.2012,IEEE, 978-0-7695-4640.

[3] www.educationaldatamining.org.

[4] Romero Cristobal & Ventura Sebastian,"Data mining in education", *WIREs Data Mining Knowl Discov* 2013, 3: 12–27 doi: 10.1002/widm.1075.

[5] Baker, R. S. J. d.. "Data Mining for Education." In International Encyclopedia of Education, 3rd ed., edited by B. McGaw, P. Peterson, and E. Baker. Oxford, 2011,UK: Elsevier.

[6] Baker, R. S. J. D., and K. Yacef. "The State of Educational Data Mining in 2009: A Review and Future Visions." Journal of Educational Data Mining, 2009, 1 (1): 3–17.

[7] C. Romero *, S. Ventura," Educational data mining: A survey from 1995 to 2005", Expert Systems with Applications 33, 2007 ,135–146.

[8] Baker, R.S.J.D. in press." Data Mining For Education", In *International Encyclopedia of Education (3rd edition)*, B. MCGAW, PETERSON, P., BAKER Ed. Elsevier, Oxford, UK. 2009.

[9] P.V.Praveen Sundar." A Comparative Study For Predicting Student's Academic Performance Using Bayesian Network Classifiers", IOSR Journal of Engineering (IOSRJEN) e-ISSN: 2250-3021, p-ISSN: 2278-8719 Vol. 3, Issue 2 (Feb. 2013), ||V1|| PP 37-42

[10] Dekker, G., Pechenizkiy, M., and Vleeshouwers J."Predicting Students Drop Out:A Case Study"In Proceeding of the 2nd International Conference on Educational Data Mining,2009,pp.41-50

[11] Baker R.S.J.d, Gowda SM, Corbett AT. "Automatically detecting a student's preparation for future learning: help use is key". In: Fourth International Conference on Educational Data Mining. Eindhoven, The Netherlands, 2011, 179–188.

[12] Behrouz Minaei-Bidgoli **I ,** Deborah **A.** Kashy ', Gerd Kortemeyer and William F. Punch,"Predicting Student Performance: *An* Application of Data Mining Methods With *an* Educational Web-Based System", 33'd ASEEIIEEE Frontiers in Education Conference, 2003, 0-7803-7961.

[13] Ahmad A. kardan , Hamid Sadeghi, "A Decision Support System for Course Offering in Online Higher Education System" Int. Journal of Computational Intelligence Systems Vol. 6 No.5 , Sep 2013, 928-942.

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

[15] Amershi, S., Conati, C., "Automatic Recognition of Learner Groups in Exploratory Learning Environments," Proceedings of ITS 2006,8th International Conference on Intelligent Tutoring Systems, 2006.

[16] Dai Shangping , Zhang Ping." A Data Mining Algorithm In Distance Learning". IEEE, 2008, 978 - 1-4244-1651.

[17] Merceron, A., & Yacef, K., "Mining student data captured from a web-based tutoring tool: Initial exploration and results," Journal of Interactive Learning Research, 15(4), 2004, (pp. 319–346).

[18] Agarwal, R., & Srikant, R., "Mining sequential patterns," In Proceedings of the eleventh international conference on data engineering, Taipei, Taiwan, 2005, (pp. 3–14).

[19] Wang, W., Weng, J., Su, J., & Tseng, S., "Learning portfolio analysis and mining in SCORM compliant environment," In ASEE/ IEEE frontiers in education conference, 2004, (pp. 17–24).

[20] Scheuer, O. & McLaren, B.M., " Educational Data Mining". In the Encyclopedia of the Sciences of Learning, 2011,Springer.

[21] Corbett A, Anderson J. Knowledge tracing: modeling the acquisition of procedural knowledge. User Model User-Adapted Interact 1995, 4:253–278.

[22] Zocco D." Risk Theory and Student Course Selection", Research in Higher Education Journal ,Vol. 3.

[23] Anthony G. Greenwald and Gerald M. Gillrnore," No Pain, No Gain? The Importance of Measuring Course Workload in Student Ratings of Instruction", Journal of Educational Psychology,1997, Vol. 89, No. 4.743-751.

[24] F. Lu, X. Li, Q. Liu, Z. Yang, G. Tan, and T. He, "Research on personalized e-learning system using fuzzy set based clustering algorithm," in Proc. Int. Conf. Comput. Sci., Beijing, China, 2007, pp. 587–590.

[25] Jing Luan, "Data Mining and Its Applications in Higher Education", NEW DIRECTIONS FOR INSTITUTIONAL RESEARCH, no. 113, Spring 2002 © Wiley Periodicals, Inc.

[26] J. Herlocker, J. Konstan, L. G. Tervin, and J. Riedl, "Evaluating collaborative filtering recommender systems," ACM Trans. Inf. Syst. J., vol. 22,no. 1, pp. 5–53, 2004.

[27] D. Zakrzewska, "Cluster analysis for user's modeling in intelligent elearning systems," in Proc. Int. Conf. Ind., Eng. Other Appl. Appl. Intell.Syst., Poland, 2008, pp. 209–214.

[28] Suhem Parack#, Zain Zahid and Fatima Merchant. "Application of Data Mining in Educational Databases for Predicting Academic Trends and Patterns".

[29] Chellatamilan T#,M.Ravichandran,Suresh R M and Dr. G.Kulanthaivel ." Eff ect of Mining educational Data to improve Adaptation of learning in e-Learning System". Second International Conference on Sustainable Energy and Intelligent System (SEISCON 2011), Dr. M.G.R. University, Maduravoyal, Chennai, Tamil Nadu,2011.

[30] Murray, T.(1999), "Authoring Intelligent Tutoring Systems: An Analysis of the State of the Art", International Journal of Artificial Intelligence in Education.Vol.10,pp.98-129.

[31] Mahdi Nasiri and Minaei B.," Predicting GPA and Academic Dismissal in LMS Using Educational Data Mining: A Case Mining", 6th National and 3rd International conference of e-Learning and e-Teaching(ICELET),2012.

[32] Tchounikine, P., Rummel, N., McLaren, M. (2010), "Computer Supported Collaborative Learning and Intelligent Tutoring Systems", Advances in Intelligent Tutoring Systems, SCI No.308,pp.447-463.

[33] Linh Bao Ngo, Vijay Dantuluri, Michael Stealey, Stan Ahalt, and Amy Apon.."An Architecture for Mining and Visualization of U.S. Higher Educational Data". Ninth International Conference on Information Technology- New Generations, IEEE,2012, 978-0-7695-4654.

[34] Ryan S. Baker, "Educational Data Mining:An Advance for Intelligent Systems in Education", IEEE Society, 2014.

[35] Nor Bahiah Hj Ahmad, Siti Mariyam Shamsuddin,"A Comparative Analysis of Mining Techniques for Automatic Detection of Student's Learning Style", IEEE, 2010 , 978-1-4244-8136.

[36] Koedinger K, Cunningham K, Skogsholm A, LeberB. An open repository and analysis tools for finegrained,longitudinal learner data. In: First InternationalConference on Educational Data Mining. Montreal, Canada; 2008, 157–166.

[37] J. Derrac, J. Luengo, J. Alcal´a-Fdez, A. Fern´andez and S. Garcıa,"Using KEEL Software as a Educational Tool: A Case of Study Teaching Data Mining", IEEE, 2011, 978-1-4577-1127.

[38] Selmoune N, Alimazighi Z. A decisional tool for quality improvement in higher education. In: InternationalConference on Information and Communication Technologies.Damascus, Syria; 2008, 1–6.

[39] Merceron A.,Yacef K.,"TADA-ED for educational Data Mining",Interactive Multimedia Electronic Journal of Computer-Enhanced Learning,2005.

[40] Rapid Miner: Available at http:// rapid-i.com,2009.

[41] C. Romero ,S. Ventura,and Gracia E."Data Mining in Course Managemnt System :Moodle Case Study and Tutorial",Computer and Education,2007.

[42] C. Romero ,S. Ventura, Gracia E,Gea M.,Carlos De Castro,"Collabrative Data Mining Tool for Education",Educational Data Mining,2009.

[43] Huebner, R., "A.Educational data-mining research", Research in Higher Edu. Journal,2012 .

[44] Romero Cristobal & Ventura Sebastian." Educational Data Mining: A Review of the State of the Art". IEEE TRANSACTIONS ON SYSTEMS MAN,AND CYBERNETICS—PART C: Appl Rev 2010, 40, NOV 2010, 1094-6977.

[45] KEEL: http://sci2s.ugr.es/keel/description.php

[46] Jindal R. and Dutta M."A survey on Educational data Mining and Research Trends", International Journal of Database Management System(IJDMS), Vol.5,No.3, June 2013,.