A Review On Research Areas In Educational Data Mining And Learning Analytics

AbhilashaSankari, Dr. ShraddhaMasih, Dr. Maya Ingle

Abstract: Over the last two-decade or so, educational data mining has evolved as an emerging discipline to analyze the type of data that comes from academics. Several research studies has carried outin Intelligent Tutoring System (ITS), Difficulty Factor Assessments, Latent Knowledge Estimation, Knowledge Inferences, Recommender System and Social Network Analysis.Gathering evidence of learning from educational setup has laid the foundation of learning analytics and educational data mining. Bayesian Knowledge Tracing (BKT), Q-Metrics, Performance Factor Analysis and Latent Knowledge Estimation methods are useful for the study of student's success. Other methods like matrix factorization and knowledge components are suited for analyzing the student's knowledge and performance. On the other hand, knowledge engineering and clustering is useful to develop student models for educational software.The current scope of research areas and methods utilized in educational data mining and learning analytics has discussed in this paper.

Keywords:Educational Data Mining, Learning Analytics

1. BACKGROUND STUDY

EDM is becoming an increasingly important area of research that is apparent through the number of publications on educational data mining in recent times. Two handbooks [1] [2] haspublished exclusively for EDMresearch. There are also review papers published in these two areas of research. The first is a review from 1995 to 2005 [3] that classified the EDM research according to data mining techniques used. Second review [4] was improvement over the first. In this, the classification done by the type of data used and based on educational categories defined in [3].A review of educational data mining regarding clustering algorithms, applicability in educational data mining and their usability [5]. Otherpresentreviews on the classification of tools used in EDM according to the taskthey perform [6].

2. SOURCES OF DATA FOR RESEARCH IN EDM

Sources of data includes log files, educational computer based systems, student enrollment data, online classes, discussion forum, standardized test and intelligent tutoring systems. Popular websites that provides educational dataset include PSLC Datashop, Peerwise Learning Environment,DataZoa,Github and MOOCs.Although computational and statistical techniques are, necessary but they are not sufficient to advance a scientific domain. Then we need some basic understanding of teaching and learning process and requirement of the stakeholder involved in the process such as teachers, students and parents.

- *Mrs. AbhilashaSankari, Assistant Professor Department of Computer Science, St. Aloysius' College, Jabalpur (M.P.)*
- *Dr. ShraddhaMasih,Reader, School of Computer Science & InformationTechnology,Devi AhilyaVishwavidyalaya, Indore (M.P.)*
- *Dr. Maya Ingle, Professor, School Of Computer Science & Technology, Devi AhilyaVishwavidyalaya, Indore(M.P.)*

Machine learning and data mining techniques provide us expertise in working with massive data. Data mining is a step in the overall KDD (Knowledge Discovery with Database) process that contains preprocessing, data mining and post processing. Data mining is already successful in many domains [7] and now showing practical results in mining educational processes.

Popular Platform of Educational Data Mining

1. DataShop
2. Peerwise
3. RiPPLE (Recommendation in Personalized Peer Learning Environment)

TABLE 1: *PREVIOUS WORK ON EDM USING DIFFERENT PLATFORMS*

SNo.	Platform	Reference	Purpose
1.	DataShop	[17]	Showed improvement in cognitive models with difficulty factor assessment.
2.	PeerWise	[8]	Showed a comparison of PeerWise and Moodle, and found that PeerWise to be more effective in promoting interaction than Moodle.

3.	RiPPLE (Recommendation in Personalized Peer Learning Environment)	[10]	Designed a framework for PeerWise learning environment. It used collaborative filtering techniques upon biased matrix factorization.

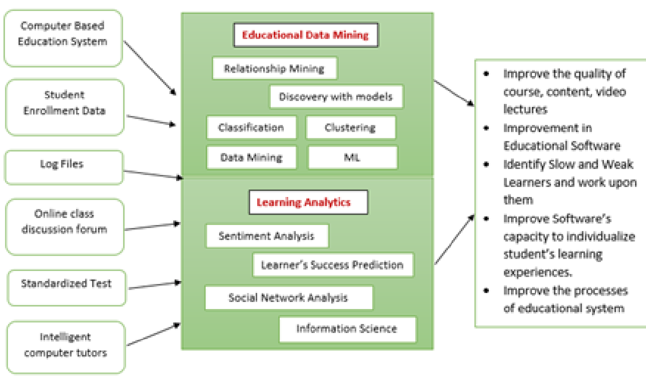


Figure 1: *Distinction between educational datamining and learning analytics*

Educational Data mining and Learning Analytics
Distinguishing educational data mining with learning analytics, is quiet entwine, since both defined in relatively similar ways.Learning analytics and educational data mining both need similar datasets and research skills with overlapping research areas, but their techniques and methods are different. Learning analytics used techniques such as sentiment analysis, social network analysis, and learner's success predictions for its research. On the other hand, an educational data mining utilizes methods of relationship mining, discovery with models, classification and clustering. The techniques has laid its foundation from automated method for discovery using educational data [12].Figure 1: Distinction between educational datamining and learning analyticsTwo board approaches of educational data mining are the classification [13] of student model. Italso suggest some methods for educational data mining. Another way is the classification of educational system. [3].

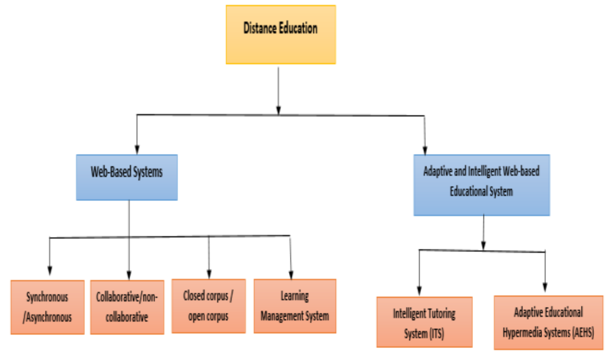


Figure 2: *Categories of distance education*

Face-to-face contact is the traditional way of educational system. On the other hand, distance education can provides methods and techniques to access educational programs and lectures from any remote area. In traditional classroom learning environment, monitoring the learners' behavior is straightforward. While in distance educational system, we need some methods to study the student's behavior. Another domain of research in educational data mining is web usage mining, whichcan be done on adaptive and intelligent web-based educational system (AIWBES). Web usage mining is the study of user's navigation behavior while using the web log files. Several research in the past has beenperformedon learning management systems. Adaptive and intelligent web-based educational system (AIWBES) is a combination of both, an intelligent tutoring system (ITS) and adaptive educational hypermedia systems (AEHS). Intelligent tutoring systemsare computer-based instructional systems that attempt to determine information about a student's learning status, and use that information to dynamically adapt the instruction to fit the student's needs. ITSs, often known as knowledge based tutors, because they have separate knowledge bases for different domain knowledge. The knowledge bases specify what to teach and different instructional strategies specify how to teach.In ITS research studies, researchers use measures such as gaming the system, off-task behavior and carelessness to analyze the disengagement behavior of students. Engagement on the other hand, detected with three aspects behavioral, affective and cognitive. To detect the off-task behavior of student [15] a machine-learned model that used just log files of student interactions with the tutoring system. Conclusion is that the off-task behavior is the result of student's lack of educational drive, passive aggressiveness and disliking of the subject.

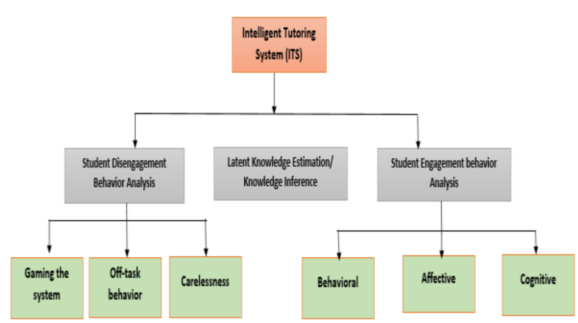


Figure 3: *Classification of ITS*

Adaptive nature of AEHS makes it useful for learners. That means learners differ in their learning styles, cultural background, knowledge level and preferences. The adaptive systems make the learning process easier for them, since it can change according to their need and preferences. Rather than applying same approach for all learning, different approaches are for different learnersare adopted.

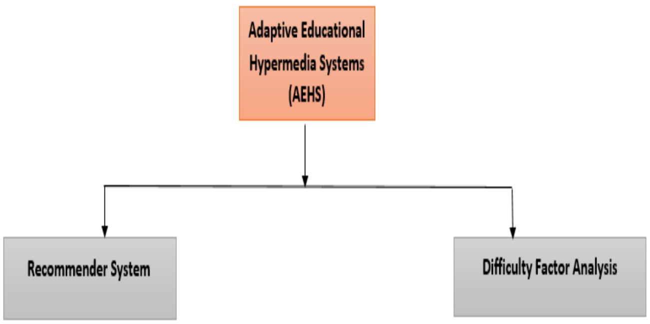


Figure 4: *Categories of AEHS*

Recommended System for Technology Enhanced Learning [16] is useful for learners to find help on materials or content. RecSystel used filtering techniques followed byRiPLE(Recommendation in Peer-Learning Environment) [10] that used collaborative filtering algorithms. Another area of research under AEHS is difficulty factor assessment. The idea of difficulty factors assessment to presents improvement on Cognitive model.
Educational data mining classifies into five broad categories [13] prediction, clustering, relationship mining, distillation of data for human judgment and discovery with models. Apart from these categories, method like knowledge engineering also exist and works on human judgment.
In prediction, some aspect of the data called predictor variable are used to find some data called the predicted variable. Prediction requires labels. Labels with noise, called bronze standard. Kappa [18] a measure to infer the degree of noise in label. The four type of prediction methods are classification, regression, latent knowledge estimation and density estimation.
Apart from other methods, knowledge engineering works for development of student models. Knowledge Engineering plays a prominent role in reasoning, problem solving and decision-making.

Table 2: *Outcomes using knowledge engineering and classification*

S.No	Method	Reference	Purpose
1.	Knowledge Engineering	[19]	Model gaming the system
2.	Knowledge Engineering	[20]	Mathematical model of self-explanation detection and prominent use of bottom-out hints.
3.	Knowledge Engineering	[21]	Student systematicity level assessment in interactive simulation
4.	Classification	[22] [23]	Discusses the detectors of student affect such as boredom , joy, frustration , anxiety etc.
5.	Classification	[24]	Develop Detectors of Off-Task behavior
6.	Classification	[25]	Develop detectors of off-task behavior, predicting learning difference among students.

Table 3: *Methods used in educational context through educational data mining and machine learning*

S. No.	Educational System	Method
	Web-based System	Log files , Association and sequential pattern mining
	LMS	Technology Acceptance Model
	ITS	Q-Matrix , Machine Learned Models , Latent Response Model
	AEHS	Decision Trees

Tools and Techniques
Data mining tools and techniques are applied to improve students success prediction, student retention and recommender systems. In recent years, many tools developed for EDM research [1]. These tools fall into four different categories namely statistical and visualization tools, clustering and classification tools, association rule mining and sequential pattern mining tools and text mining.

Table 4: *Educational Data Mining and Machine Learning Tools*

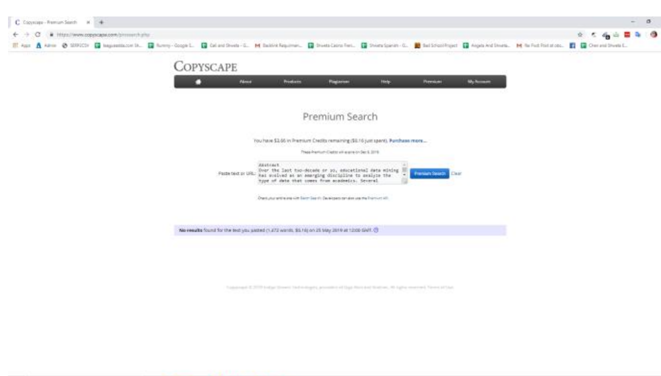
S. No.	Tool	Purpose
1.	WEKA (Waikato Environment for Knowledge Assessment)	Free software contains large collections of machine learning and data mining algorithm
2.	Simulog (Simulation of User Logfiles)	Tool that simulate log files [26]
3.	ASquare(Author Assistant)	Includes data mining methods[26]
4.	GISMO	Visualization Tool
5.	CourseVis	Visualization Tool
6.	Listen Tool	Visualization Tool
7.	AccessWatch	Analyze web server logs
8.	Gwstat	Analyze web server logs
9.	Webstat	Analyze web server logs
10.	Multistar	Association and Classification
11.	Synergo	Visualization Tool
12.	Co1AT	Visualization Tool

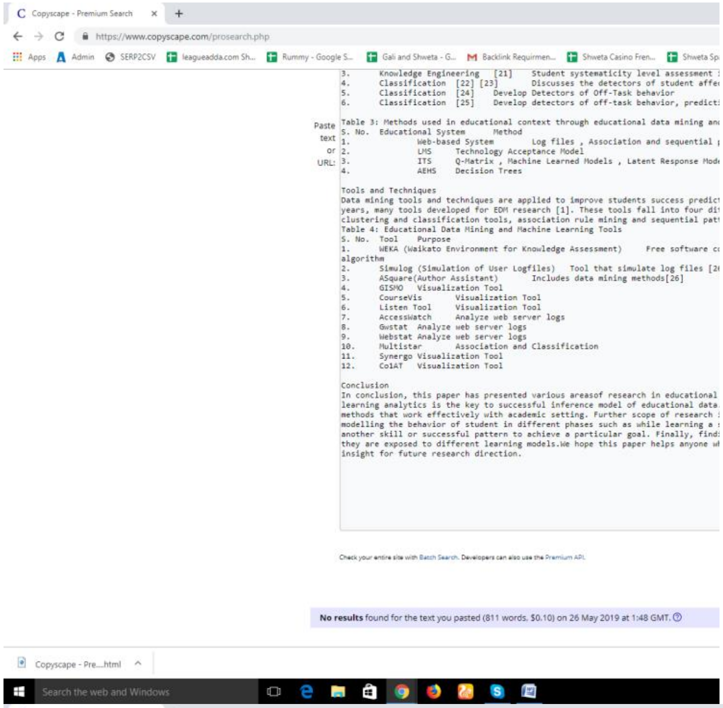
CONCLUSION

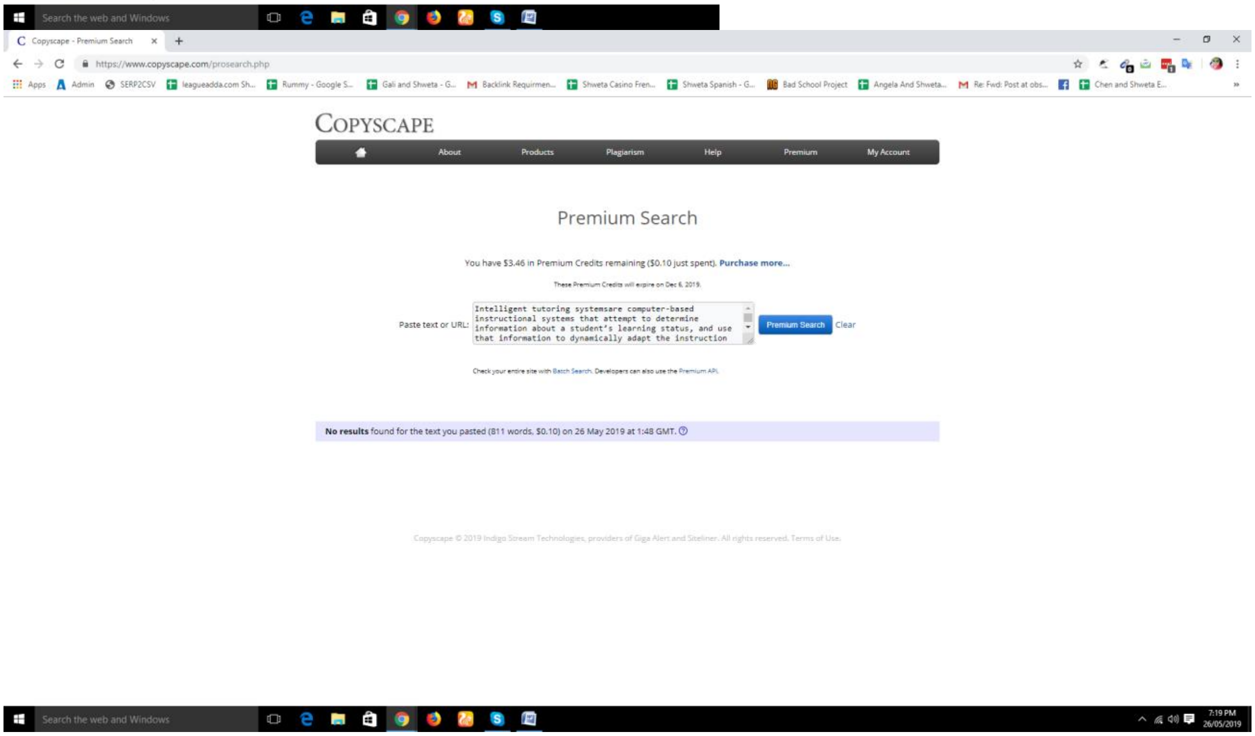
In conclusion, this paper has presented various areasof research in educational data mining and learning analytics. The use of EDM and learning analytics is the key to successful inference model of educational data. In addition to basic approaches in EDMthe paper presents methods that work effectively with academic setting. Further scope of research in learning analytics and educational data mining includes modelling the behavior of student in different phases such as while learning a skill and prediction of their performances while learning another skill or successful pattern to achieve a particular goal. Finally, finding correlation between student learning performances when they are exposed to different learning models.We hope this paper helps anyone who is interested in educational datamining research with insight for future research direction.

BIBLIOGRAPHY

[1] Romero, C., & Ventura, S. (Eds.). (2006). Data mining in e-learning (Vol. 4). WIT press.
[2] Romero, C., Ventura, S., Pechenizkiy, M., & Baker, R. S. (Eds.). (2010). Handbook of educational data mining. CRC press.
[3] Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. Expert systems with applications, 33(1), 135-146.
[4] Romero, C., & Ventura, S. (2010). Educational data mining: a review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40(6), 601-618.
[5] Dutt, A., Ismail, M. A., &Herawan, T. (2017). A systematic review on educational data mining. IEEE Access, 5, 15991-16005.
[6] Venkatachalapathy, K., Vijayalakshmi, V., &Ohmprakash, V. (2017, February). Educational Data Mining Tools: A Survey from 2001 to 2016. In 2017 Second International Conference on Recent Trends and Challenges in Computational Models (ICRTCCM) (pp. 67-72). IEEE.
[7] Srivastava, J., Cooley, R., Deshpande, M., & Tan, P. N. (2000). Web usage mining: Discovery and applications of usage patterns from web data. AcmSigkdd Explorations Newsletter, 1(2), 12-23.a
[8] McKenzie, W., &Roodenburg, J. (2017). Using PeerWise to develop a contributing student pedagogy for postgraduate psychology. Australasian Journal of Educational Technology, 33(1).
[9] Khosravi, H., Cooper, K., &Kitto, K. (2017). Riple: Recommendation in peer-learning environments based on knowledge gaps and interests. arXiv preprint arXiv:1704.00556.
[10] Botelho, A. F., Baker, R. S., & Heffernan, N. T. (2017, June). Improving sensor-free affect detection using deep learning. In International Conference on Artificial Intelligence in Education(pp. 40-51). Springer, Cham.
[11] Siemens, G., & d Baker, R. S. (2012, April). Learning analytics and educational data mining: towards communication and collaboration. In Proceedings of the 2nd international conference on learning analytics and knowledge (pp. 252-254). ACM.
[12] d Baker, R. S. (2010). Mining data for student models. In Advances in intelligent tutoring systems (pp. 323-337). Springer, Berlin, Heidelberg.
[13] Findik-Coşkunçay, D., Alkiş, N., &Özkan-Yildirim, S. (2018). A Structural Model for Students' Adoption of Learning Management Systems: An Empirical Investigation in the Higher Education Context. Journal of Educational Technology & Society, 21(2), 13-27.
[14] Baker, R. S. (2007, April). Modeling and understanding students' off-task behavior in intelligent tutoring systems. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 1059-1068). ACM.
[15] Manouselis, N., Drachsler, H., Vuorikari, R., Hummel, H., & Koper, R. (2011). Recommender systems in technology enhanced learning. In Recommender systems handbook (pp. 387-415). Springer, Boston, MA.
[16] Stamper, J. C., & Koedinger, K. R. (2011, June). Human-machine student model discovery and improvement using DataShop. In International Conference on Artificial Intelligence in Education (pp. 353-360). Springer, Berlin, Heidelberg.
[17] Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and psychological measurement, 20(1), 37-46.
[18] Beal, C. R., Qu, L., & Lee, H. (2006, July). Classifying learner engagement through integration of multiple data sources. In Proceedings of the National Conference on Artificial Intelligence (Vol. 21, No. 1, p. 151). Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999.
[19] Shih, B., Koedinger, K. R., &Scheines, R. (2011). A response time model for bottom-out hints as worked examples. Handbook of educational data mining, 201-212.
[20] Buckley, B. C., Gobert, J. D., &Horwitz, P. (2006, June). Using log files to track students' model-based inquiry. In Proceedings of the 7th international conference on Learning sciences (pp. 57-63). International Society of the Learning Sciences.
[21] Conati, C., & Maclaren, H. (2009). Empirically building and evaluating a probabilistic model of user affect. User Modeling and User-Adapted Interaction, 19(3), 267-303.
[22] D’mello, S. K., Craig, S. D., Witherspoon, A., Mcdaniel, B., &Graesser, A. (2008). Automatic detection of learner’s affect from conversational cues. User modeling and user-adapted interaction, 18(1-2), 45-80.
[23] Baker, R. S. (2007, April). Modeling and understanding students' off-task behavior in intelligent tutoring systems. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 1059-1068). ACM.
[24] Cetintas, S., Si, L., Xin, Y. P. P., &Hord, C. (2009). Automatic detection of off-task behaviors in intelligent tutoring systems with machine learning techniques. IEEE Transactions on Learning Technologies, 3(3), 228-236.
[25] Bravo, J., Vialardi, C., &Ortigosa, A. (2008, June). Asquare: a powerful evaluation tool for adaptive hypermedia course system. In Proceedings of the nineteenth ACM conference on Hypertext and hypermedia (pp. 219-220). ACM.




Copyscape - Premium Search
https://www.copyscape.com/prosearch.php
Tools and Techniques
Conclusion
No results found for the text you pasted (811 words, $0.10) on 26 May 2019 at 1:48 GMT.


COPYSCAPE
About
Products
Plagiarism
Help
Premium
My Account
Premium Search
You have $3.46 in Premium Credits remaining ($0.10 just spent). Purchase more...
Paste text or URL:
Intelligent tutoring systemsare computer-based instructional systems that attempt to determine information about a student's learning status, and use that information to dynamically adapt the instruction
Premium Search
Clear
No results found for the text you pasted (811 words, $0.10) on 26 May 2019 at 1:48 GMT.