



Big Data In Education Data Mining And Learning

Suwarna Nimkarde¹, Shobhana Gaikwad²

Lecturer, Computer Technology, Bharati vidyapeeth Institute of Technology, Navi Mumbai, India¹

Lecturer, Computer Technology, Bharati vidyapeeth Institute of Technology, Navi Mumbai, India²

Abstract: Education institute are the nursery for the future minds of the nation. Knowledge represents the intangible assets of the education institutes, industries and nations. With development of the information and telecommunication technology activities like commerce, communication, entertainment and learning are occurring on internet. As universities and colleges are started using an online learning platform for providing content to students and started using student management system for better management of the students personal data. Education institute have large amount of student data like basic personal information, attendance, marks, achievements etc. Online learning platform provide an opportunity to capture fine gain data about student online activities like course content he browse, time spent on each unit, post on forum, practice test, sequence of activity will generate large amount of structured and unstructured data. But it is found that educational system are notoriously poor in managing the data and taking advantage of this generated data. There are two research area for Big Data mining in education called educational data mining and learning analytics. Educational data mining is suit for the computational and psychological methods and research approach for the understanding how student learn, predict student future learning behaviour. Learning analytics is becoming defined as an area of research and application and is related to academic analytics actions and predication analytics. Recently waste amount of work has been done in other area like ecommerce portal and increase the click through rate. So now considering opportunities with data generate by the online learning platform we can mine the education data for calculating learner performance interest problem face by the learner knowledge level for different knowledge point. As recently there are lot of research in education data mining and some researcher started treating data in education system as big data problem, we have done survey of various research in education.

Keywords: —Education, Data Mining, Predictive analysis, Big Data, Hadoop, Association Rules.

I. INTRODUCTION

Recent years we have witnessed rapid growth of the Internet, which has become an important medium to deliver digital content to Web users instantaneously. E-commerce and portal site now have vast amount of data about their visitors profile, visitor behavior and action on web portal, customer care, feedbacks and search queries. Social media site such as twitter, facebook etc. provide great venues for customer to share joy and struggle, vent emotion and stress and seek social support. The recent report shows that online courses are spreading and developing around the world [1]. An educational institute and universities started providing e-learning facilities for providing contents and using ERP for better management of student data have opened up opportunities to collect and analyze student data, to discover patterns and trends in those data and to make new discoveries and test hypothesis about how students learn [6]. An educational institutes have vast amount of data like student personal information like name, student registration number, birth date, and address etc. and curriculum information like code, name, semester achievements, assignments status unit tests and semester marks etc. When students are learning online, there are multiple opportunities to exploit the power of the technology for formation of for formative assessment [1]. The same technology that supports learning activities gathers data in the course of learning that can used for assessment. An online system are collecting much more detail information about how students are learning than manual method. Now many universities and college and have started online forum where student share their day today encounter share their knowledge seek help for problem also provide opportunity to capture fine grain data about student can provide deep insight view [4]. There are two areas that are specific to the use of Data in education are educational data mining and learning analytic. Educational data mining is emerging as a research area with a suit of computational and psychological methods and research approaches for understanding how students learn. Learning analytic is becoming defined as an area of research and application and it related to academic analytics, and predictive analytics. Learning analytics draws on a broader array of academic disciplines than educational data mining incorporating concept and technique from information science and sociology in addition to computer science statistics psychology and learning sciences. Unlike educational data mining learning analytics generally does not emphasize reducing learning into components but instead seeks to understand entire systems and to support human decision making. Big Data as a term has been among the biggest trends of the last three years, leading to an upsurge of research, as well as industry and government applications. The process of handling big



data encompasses collection, storage, transportation and exploitation. “Big data” has come to be defined by the four V’s— Volume, Velocity, Veracity, and Variety [2]. Volume indicates the size of the data, which might be too big to be handled by the current state of algorithms and/or systems. Velocity implies data are streaming at rates faster than that can be handled by traditional algorithms and systems. Veracity suggests that despite the data being available, the quality of data is still a major concern. That is, we cannot assume that with big data comes higher quality. In fact, with size comes quality issues, which needs to be either tackled at the data pre-processing stage or by the learning algorithm. Variety is the most compelling of all V’s as it is presenting data of different types and modalities for a given object in consideration [2]. Data in education is big as compared to common data sets in many domains. It is not human genome project or Google big but bigger than data in most domains. If we consider the number of students in institute and university it will produce data at very large rate because educational software provides an opportunity to capture data at fine scale (can log behaviour at second by seconds levels). If we consider forum set by colleges produce unstructured data form posts log file. To get hostile knowledge about student, require continue analysis of data generated because student knowledge level is continuously changing student performance and behaviour also change frequently [9]. Recently there is massive research in field of education data mining.

II. EDUCATION DATA MINING

Educational data mining is emerging as a research area with a suit of computational and psychological methods and research approaches for understanding how students learn. With the rapid development of computer technology and the increasing popularity of the Internet, the e-learning system ushered a new development opportunities, and worldwide to be extensively adopted. Now days there is a vast amount of research in education. Education data mining have targeted various issues like Understanding real learning behaviour of a students, the detection of students who show signs of frustration, students' academic achievements (successes and failures), students' drop out, and students' financial behaviour. We have done the survey of various research in education domain. The technology of data mining could discover useful knowledge from mass data, so it has been broadly applied in various fields. There are various research to analyse the exam grade of college students. There is attempt to find the weighted association rules to analyse grades of college wide examination course [6]. Instead of directly applying the Apriori algorithm, more valuable correlations have been obtained between the chapters, chapters and scores, colleges and chapters at the same threshold values. Association rule is an implication of the form $A \Rightarrow B$, where A and B are the sets of items (for short item sets) A and B are called antecedent (Left-hand-side or LHS) and consequent (right-hand-side or RHS) of the rule respectively. Let I be a set of m binary attributes called items. Let D be a set of transactions called the database. Each transaction in D has a unique transaction ID and contains a subset of the items in I . The support of an item set A is defined as the proportion of transactions in the data set which contain the item set. Confidence of the rule is interpreted as the percentage of transactions that contain B among transaction that contain A , association rule mining is the discovery of all association rules that are above a user-specified minimum support and minimum confidence. Online discussion forums have emerged as a popular Web application to build and support online communities for numerous engineering interest areas and practice. However, a review of engineering education literature reveals scant research on the use of online discussion forums for engineering learning beyond the classroom [4]. In case of online discussion there is an various ongoing research to find what student’s sought help for and for what purpose, which words are frequently used by students who seek homework help?, what are the major themes of discussion?. To answer these questions there are various attempt to use keyword analysis, topic modeling, linguistic inquiry and word count analysis. There is another stream of research in application of system based on emotion recognition of learner’s the teaching. In the traditional E-learning the learner's emotional information cannot be extracted effectively, so that the study results are not ideal. The E-learning system based on emotional recognition will play a greater role in E-learning system, which has become an important research topic. There are three main facial expression methods used 1) Geometric feature-based recognition 2) Based on the overall identification method 3) The model-based on identification method. There is an also attempt to mining social media data for understanding students’ learning experiences. Students are more active on social media sites like facebook, twitter, LinkedIn where they share there day today encounters, experiences, emotions, stress, achievements [3]. Analysis of these informal discussions on social media can give vast amount of information about student learning experience, opinions, feeling, and concerns. Analysis of such data will provide key information for education decision makers. There is an attempt to mine twitter data, have collected about 28500 tweets using the hashtag #engineeringProblems over a period of 14 months, and a second data set of 39,095 tweets using the geo-code (longitude and latitude) of Purdue University, West Lafayette. This research proposes a workflow to bridge and integrate a qualitative research methodology and large-scale data mining techniques. Research base their data-mining algorithm on qualitative insights resulting from human interpretation, so that gain deeper understanding of the data. Research apply the algorithm to another large-scale and unexplored data set, so that the manual method is augmented. Research can keep refining the model based on further human feedback like the cycle [3]. Another area of research is to predict the student learning outcome. There is an attempt to improve engineering education in the Taiwan. In a successful engineering education, continuous improvement should be fulfilled through all phases: curriculum, speech or talk listening, practical training, industrial visiting, contest attending and undergraduate research [1]. Once a student is



predicted to get bad grade, its teacher can strengthen the student's learning difficulty at the beginning of a course. So research focusing on developing data mining tool which can forecast students' performance to the instructor. In order to solve this issue, research proposes a decision support of integrating data mining, fuzzy logic and Bayesian networks to qualitatively predict students' learning. The research have proposed four steps: 1) using fuzzy theory to identify the influence factors on learning outcomes through related work analysis and questionnaires; 2) Using data mining to analyse the relationships among the factors and then to construct influence diagram; (3) using machine learning to establish the fuzzy inference relations through questionnaires and parameter learning; (4) using the model to predict the exam scores at the beginning of course and thereby to help students enhance their scores according to their weakness [1]. One of the promising solution to deals with big data problem is Hadoop distributed processing platform. At the centre of the Hadoop is Map Reduce framework and Hadoop distributed file system. MapReduce is also a data processing model. Its greatest advantage is the easy scaling of data processing over multiple computing nodes. Under the MapReduce model, the data processing primitives are called mappers and reducers. Decomposing a data processing application into mappers and reducers is sometimes nontrivial. But, once you write an application in the MapReduce form, scaling the application to run over hundreds, thousands, or even tens of thousands of machines in a cluster is merely a configuration change [2]. Now if we consider number student pursuing the bachelor degree in universities like Mumbai, Pune, Solapur, and Shivaji in Maharashtra, there is a huge amount of student academic data in universities [7]. And now days various universities have started online course so these can produce huge amount of data and mostly that data are in various forms like relation data in tabular format, web site logs and forum data are semi structure format. So various research like research centre of Microvity private ltd. have started treating education data as big data problem to improve the education quality. Below is the system architecture of Microvity education analysis platform.

CONCLUSION

Thus we can done the survey of various research in big data in education data mining. We have survey system will help the student to find solution to their problems. We have survey systems that will serve student with improvement and suggestion will help to keep student interested in learning and explaining new topic. We have survey systems that will also help the education decision maker by providing key information about student so they can take appropriate measure to improve education quality.

REFERENCES

- [1]. Alana M. de Morais, Joseana M.F.R. Araujio and Evandro B. Costa, "Monitoring Student Performance Using Data Clustering and Predictive Modeling", IEEE 2014. J. Breckling, Ed., The Analysis of Directional Time Series: Applications to Wind Speed and Direction, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61. J. Breckling, Ed., The Analysis of Directional Time Series: Applications to Wind Speed and Direction, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.
- [2]. Zhi-Hua Zhou, Nitesh V. Chawla, Yaochu Jin, and Graham J. Williams, "Big Data Opportunity And Challenges: Discussion From Data Analysis perspective", IEEE Compitational intelligence magazine, Nov. 2014.
- [3]. Xin Chen, Mihaela Vorvoreanu and Krishna Madhavan, "Mining Social Media for Understanding Students Learning Experience", IEEE Transaction on Learning Technologies, Vol. 7, No. 3, Jul.-Sep., 2014.
- [4]. "Towards on Understanding of ECE Students Online Home Help Forum", IEEE, 2014.
- [5]. Hadautho Roberto Barros da Silva and Paulo Jorge Leitao Adeodato, "A Data Mining Approach for Undergraduate Students Retention", IEEE WCCI, Jun. 2012.
- [6]. Xin-hua Zhu, Ya-qiong Deng, Qing-ling Zeng, "Analysis on Course Grade of College-wide Examination Based on Mixed Weighted Association Rules Mining Algorithm", IEEE International Conference on Computer Application and System Modeling, 2010.