International Journal of Research In Science & Engineering e-ISSN: 2394-8299
Volume: 3 Issue: 2 March-April 2017 p-ISSN: 2394-8280

BIG DATA CLASSIFICATION OF STUDENT RESULT PREDICTION

Sourabh Sahu[1],Prof. Mayank Bhatt[2]
M.Tech Scholar, Assistant Prof. & HOD
Department of Computer Science & Engineering, LNCT Indore, India
sourabhsahu31@gmail.com ,maynkbhatt27@gmail.com

ABSTRACT

This Paper focuses on the specific problem of Big Data classification of student result. It discusses the system challenges presented by the Big Data problems associated with student result prediction. The prediction of a possible student result requires continuous collection of t data and learning of their characteristics on the fly. The continuous collection of data by the university to Big Data problems that are caused by the volume, variety and velocity properties of Big Data. The Big Data properties will lead to significant system challenges to implement machine-learning frameworks. This paper discusses the problems and challenges in handling Big Data classification using geometric representation-learning techniques and the modern Big Data technologies. In particular this synopsis discusses the issues related to combining supervised learning techniques, representation-learning techniques, machine lifelong learning techniques and Big Data technologies (e.g. Hadoop, Hive and Cloud) for solving student result classification problems

I. INTRODUCTION

Educational data mining is emerging as a research area with a suite of computational and psychological methods and research approaches for understanding how students learn. New computer-supported interactive learning methods and tools intelligent tutoring systems, simulations, games— 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 hypotheses about how students learn. Data collected from online learning systems can be aggregated over large numbers of students and can contain many variables that data mining algorithms can explore for model building. Just as with early efforts to understand online behaviors, early efforts at educational data mining involved mining website log data, but now more integrated, instrumented, and sophisticated online learning systems provide more kinds of data. Educational data mining generally emphasizes reducing learning into small components that can be analyzed and then influenced by software that adapts to the student.

Figure1. Three Tiers of Big Data

In this synopsis class X student's data of Central Board of Secondary Education in all over India is considered for input dataset. Since it contains terabytes of student data, these datasets

are considered as big data. In big data concept the traditional data mining algorithms are translated to Map Reduce algorithms for running them on Hadoop clusters by translating their data analytics logic to the Map Reduce job which is to be run over Hadoop clusters. Hadoop clusters are designed for storing and processing huge amount of data in a distributed computing environment. Map Reduce is one of the major hadoop components for distributed data processing. The clustering method is used to identify academically at- risk students and categorize the students accordingly. Since there are many algorithms for data clustering, the K-Means method is used here. The multiple regression algorithms is used for predicting student results. Both algorithms are translated to Map Reduce algorithms to run on hadoop clusters. In K- means clustering it consists of 2 parts Map and Reduce. The map function performs the procedure of assigning each sample to the closest center while the reduce function performs the procedure of updating the new centers. In order to decrease the cost of network communication, a combiner function is developed to deal with partial combination of the intermediate values with the same key within the same map task. The Map Reduce programming model consists of two separate and distinct tasks that Hadoop programs perform. The first is the map job, which takes a set of data and converts it into another set of data, where individual elements are broken down into tuples (key/value pairs). The reduce job takes the output from a map as input and combines those data tuples into a smaller set of tuples.

II. RELATED WORK

In April of 2011, Clint McElroy[2] designed a system for The Online Student Profile Learning System: a Learner-Centered Approach to Learning Analytics , solution to addressing the retention and success of at-risk students (the Online Student Profile system developed in CPCC's 2003-08) and work with partner colleges to deploy both the OSP and the related faculty and staff development activities in order to improve retention and student success .

Beth Dietz-Uhler & Janet E. Hurn [5] define learning analytics, how it has been used in educational institutions, what learning analytics tools are available, and how faculty can make use of data in their courses to monitor and predict student performance. They also provide details of several issues and concerns with the use of learning analytics in higher education. Weizhong Zhao[7] designed Parallel K- Means Clustering Based on MapReduce for clustering , Data clustering has been received considerable attention in many applications, such as data mining, document retrieval, image segmentation and pattern classification.
Learning analytics (LA) is a multi-disciplinary field involving machine learning, artificial intelligence, information retrieval, statistics, and visualization. LA is also a field in which several related areas of research in TEL converge. These include academic analytics, action research, educational data mining, recommender systems, and personalized adaptive learning. M.A. Chatti, A.L. Dyckhoff, U. Schroeder, and H[5]. These review recent publications on LA and its related fields and map them to the four dimensions of the reference model. Furthermore, we identify various challenges and research opportunities in the area of LA in relation to each dimension. Kenneth Wottrich[7] propose a research in 2010 to characterize and model the performance of MapReduce applications on typical, scalable clusters based on fundamental application data and processing metrics. He identified five fundamental characteristics which define the performance of

Map Reduce applications. Then he created five separate bench mark tests, each designed to isolate and test a single characteristic. The results of these benchmarks are helpful in constructing a model for MapReduce applications.

Seyed Reza Pakize[6] make a study on A Comprehensive View of Hadoop MapReduce Scheduling Algorithms which helps researchers. There are three important scheduling issues in MapReduce such as locality, synchronization and fairness. The most common objective of scheduling algorithms is to minimize the completion time of a parallel application and also achieve to these issues. There are many algorithms to solve this issue with different techniques and approaches. Some of them get focus to improvement data locality and some of them implements to provide Synchronization processing.

III. EXISTING SYSTEM

The main goal of this paper is to identify academically at risk students and to develop a predictive model to predict student academic performance in educational institutions, which helps to predict their future results. Student academic performance is affected by numerous factors. The scope of this research is limited to the investigation of learning progression on their academic performance. The proposed system consists of two functionalities: a) Identifying academically at-risk students b) Prediction of student result

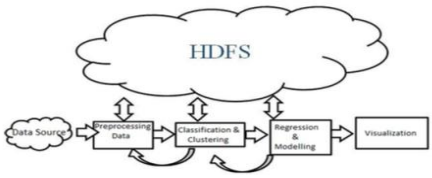

Figure2. Existing System

Algorithm MAP (key,value)

Input: Global variable centers, the offset key, the sample value Output: <key', value'> pair, where the key' is the index of the closest center point and value' is a string comprise of sample information

1. Construct the sample instance from value;

2. mindist = Double.maxvalue;

3. index = -1;

4. For i=0 to centers.length do

dist= ComputeDist(instance, centers[i]); If dist < mindist { mindist = dist; index = i;

}

5. End For

6. Take index as key';

7. Construct value' as a string comprise of the values of

different dimensions;

8. output < key, value> pair;

The data collected from different applications require proper method of extracting knowledge from large repositories for better decision-making. This makes an extreme challenge for institutions using traditional data management mechanism to store and process huge datasets. So it is required to define a new paradigm called "Big Data Analytics" to re-evaluate current system and to manage and process huge data.

IV. PROPOSED SYSTEM

Clustering analysis is the process of identifying data sets that are similar to each other to understand the differences as well as the similarities within the data. Clusters have certain traits in common that can be used to improve targeting algorithms. For example, clusters of customers with similar buying behaviour can be targeted with similar products and services in order to increase the conversation rate. A result from a clustering analysis can be the creation of personas. Personas are fictional characters created to represent the different user types within a targeted demographic, attitude and/or behaviour set that might use a site, brand or product in a similar way. The programming language R has large variety of functions to perform relevant cluster analysis and is therefore especially relevant for performing a clustering analysis.

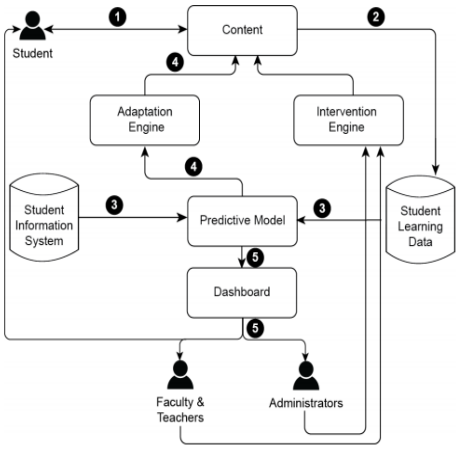

Figure3. Proposed System

Classification Analysis is a systematic process for obtaining important and relevant information about data, and metadata – data about data. The classification analysis helps identifying to which of a set of categories different types of data belong. Classification analysis is closely linked to cluster analysis as the classification can be used to cluster data.

V. CONCLUSION

This synopsis describes about the advent of Big Data, Architecture and Characteristics. Here

we discussed about the classifications of Big Data to the business needs and how for it will help us in decision making in the business environment.

REFERENCE

[1] Pool, Lorraine Dacre, Pamela Qualter, and Peter J. Sewell. "Exploring the factor structure of the CareerEDGE employability development profile." Education+ Training 56.4 (2014): 303-313.

[2] Saranya, S., R. Ayyappan, and N. Kumar. "Student Progress Analysis and Educational Institutional Growth Prognosis Using Data Mining." International Journal Of Engineering Sciences & Research Technology, 2014

[3] Hicheur Cairns, Awatef, et al. "Towards CustomDesigned Professional Training Contents and Curriculums through Educational Process Mining." IMMM 2014, The Fourth International Conference on Advances in Information Mining and Management. 2014.

[4] Archer, Elizabeth, Yuraisha Bianca Chetty, and Paul Prinsloo. "Benchmarking the habits and behaviors of successful students: A case study of academic-business collaboration." The International Review of Research in Open and Distance Learning 15.1 (2014).

[5] Arora, Rakesh Kumar, and Dharmendra Badal. "Mining Association Rules to Improve Academic Performance." (2014).

[6] Peña-Ayala, Alejandro. "Educational data mining: A survey and a data mining-based analysis of recent works." Expert systems with applications 41.4 (2014): 1432-1462.

[7] Vanhercke, Dorien, Nele De Cuyper, Ellen Peeters, and Hans De Witte. "Defining perceived employability: a psychological approach." Personnel Review 43, no. 4 (2014): 592-605.

[8] Potgieter, Ingrid, and Melinde Coetzee. "Employability attributes and personality preferences of postgraduate business management students." SA Journal of Industrial Psychology 39.1 (2013): 01-10.

[9] Jantawan, Bangsuk, and Cheng-Fa Tsai. "The Application of Data Mining to Build Classification Model for Predicting Graduate Employment." International Journal Of Computer Science And Information Security (2013).

[10] Bakar, Noor Aieda Abu, Aida Mustapha, and Kamariah Md Nasir. "Clustering Analysis for Empowering Skills in Graduate Employability Model." Australian Journal of Basic and Applied Sciences 7.14 (2013): 21-28.

[11] Singh, Samrat, and Vikesh Kumar. "Performance Analysis of Engineering Students for Recruitment Using Classification Data Mining Techniques." International Journal of Computer Science & Engineering Technology, (2013).

[12] Finch, David J., Leah K. Hamilton, Riley Baldwin, and Mark Zehner. "An exploratory study of factors affecting undergraduate employability." Education+ Training 55, no. 7 (2013): 681-704.

[13] Jackson, Denise, and Elaine Chapman. "Non-technical skill gaps in Australian business graduates." Education+ Training 54.2/3 (2012): 95-113.

[14] Dejaeger, Karel, et al. "Gaining insight into student satisfaction using comprehensible data mining techniques." European Journal of Operational Research 218.2 (2012): 548-562.

[15] Padhy, Neelamadhab, Dr Mishra, and Rasmita Panigrahi. "The survey of data mining applications and feature scope." Asian Journal Of Computer Science And Information Technology (2012).

[16] Osmanbegović, Edin, and Mirza Suljić. "Data mining approach for predicting student performance." Economic Review 10.1 (2012).

[17] Şen, Baha, Emine Uçar, and Dursun Delen. "Predicting and analyzing secondary education placement-test scores: A data mining approach." Expert Systems with Applications 39.10 (2012): 9468-9476.
[18] Agrewal, S., G. Pandey, and M. Tiwari. "Data mining in education: data classification and decision tree approach." International Journal of e-Education, e- International Journal of Computer Applications (0975 – 8887) Volume 110 – No. 15, January 2015 66 Business, e-Management and e-Learning, 2 (2) (2012): 140-144.
[19] Pandey, Umesh Kumar, and Brijesh Kumar Bhardwaj. "Data Mining as a Torch Bearer in Education Sector." Technical Journal of LBSIMDS (2012).
[20] Srimani, P. K., and Malini M. Patil. "A Classification Model for Edu-Mining." PSRC-ICICS Conference Proceedings. 2012.
[21] Sukanya, M., S. Biruntha, Dr S. Karthik, and T. Kalaikumaran. "Data mining: Performance improvement in education sector using classification and clustering algorithm." In International conference on computing and control engineering,(ICCCE 2012), vol. 12. 2012.
[22] Yadav, Surjeet Kumar, and Saurabh Pal. "Data Mining Application in Enrollment Management: A Case Study." International Journal of Computer Applications (IJCA) 41.5 (2012): 1-6.
[23] Torenbeek, M., E. P. W. A. Jansen, and W. H. A. Hofman. "Predicting first-year achievement by pedagogy and skill development in the first weeks at university." Teaching in Higher Education 16.6 (2011): 655-668.
[24] Gokuladas, V. K. "Predictors of Employability of Engineering Graduates in Campus Recruitment Drives of Indian Software Services Companies." International Journal of Selection and Assessment 19.3 (2011): 313- 319.
[25] Yongqiang, He, and Zhang Shunli. "Application of Data Mining on Students' Quality Evaluation." Intelligent Systems and Applications (ISA), 2011 3rd International Workshop on. IEEE, 2011.
[26] Sakurai, Yoshitaka, Setsuo Tsuruta, and Rainer Knauf. "Success Chances Estimation of University Curricula Based on Educational History, Self-Estimated Intellectual Traits and Vocational Ambitions." Advanced Learning Technologies (ICALT), 2011 11th IEEE International Conference on. IEEE, 2011.
[27] Pandey, Umesh Kumar, and Saurabh Pal. "A Data mining view on class room teaching language." International Journal Of Computer Science Issues (2011).
[28] Arjariya, Tripti, et al. "Data Mining and It's Approaches towards Higher Education Solutions." International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307. [29] Aher, Sunita B., and L. M. R. J. Lobo. "Data mining in educational system using Weka." IJCA Proceedings on International Conference on Emerging Technology Trends (ICETT). Vol. 3. 2011.
[30] Suthan, G. Paul, and S. Baboo. "Hybrid chaid a key for mustas framework in educational data mining." IJCSI International Journal of Computer Science Issues 8 (2016): 356-360.