Implementation of Learning Analytics with Deduplication

I. BHUVANESHWARRI
Assistant Professor (Senior), Department of Information Technology, Institute of Road and Transport Technology, Erode, INDIA.

Abstract- ***Learning Analytics (LA) provides useful insights to tutors on the behavior of students in an online learning environment. The information collected for LA is treated as big data used to customize the educational environment, optimize the learning resources and activities, and personalize the student experience. The major three steps of LA are information gathering, information processing and information presentation. The benefit related with big data is scalability, it can be collected automatically in a computer-enabled learning environment and analyzed in bulk without requiring additional time and resources even though the data size grows significantly. However, the analysis of big data is not easy and requires suspicious investigations before it can be useful to researchers, teachers and students, and curriculum and technology developers. The duplication of big data impact results of analysis. In this paper, I have proposed deduplication technique on big data in information gathering step of LA. In this paper, I have taken real world data set and apply the deduplication technique and then compare the results. From the result, it shows that after applying the proposed method the duplicated data is removed from large data set and the size of big data taken for LA is optimized to reasonable size and provides better result.***

Indexed Terms- ***Learning Analytics, Big Data, Deduplication, optimization.***

I. INTRODUCTION

Learning analytics is an emerging field in which sophisticated analytic tools are used to improve learning and education. It draws from, and is closely tied to, a series of other fields of study including business intelligence, web analytics, academic analytics, educational data mining, predictive analytics and action analytics. The goal of this emerging field as the ability to scale real-time use of learning analytics by students, instructors, and academics advisors to improve student success.

Moreover, learning analytics is focused on building systems able to adjust content, levels of support and other personalized services by capturing, processing, reporting and acting on data on an ongoing basis in a way that minimizes the time delay between the capture and use of data. Thus, in contrast to current evaluation processes which use the results from one semester to inform improvements in the next, learning analytics seeks to combine historical and current user data to predict what services specific users may find useful now [1].

Thus, learning analytics seeks to capitalize on the modelling capacity of analytics: to predict behavior, act on predictions, and then feed those results back into the process in order to improve the predictions over time [2] as it relates to teaching and learning practices. Currently however, the built-in student tracking functionality in most Content Management System (CMS) / Learning Management System (LMS) are far from satisfactory [3] and do not offer sufficient learning activity reports for instructors to effectively tailor learning plan that meet the needs of their students [4].

Thus, the study and advancement of learning analytics involves: (1) the development of new processes and tools aimed at improving learning and teaching for individual students and instructors, and (2) the integration of these tools and processes into the practice of teaching and learning.

This research work will be concentrated on LA and provide optimized performance and solutions based technique on Big data. In this paper Section – II discuss about Big data analysis and deduplication of Big data with its issues, Section –III discuss about proposed system, Section –IV covers problem definition, Section-V describes the contribution of the

proposed work and in Section – VI concludes the paper.

II. BIG DATA ANALYSIS AND DEDUPLICATION OF BIG DATA WITH ITS ISSUES

Big Data LA gives access to wider perspectives of information. Especially it allows processing unstructured and structured data together. Hence, there is need to integrate the data from the different sources and analyze the data which can be given efficient result.

However lots of data sources do not mean that the quality of data is enough to provide reliable results [4]. There are several different quality indicators related to Big Data LA. After integration the data becomes large, duplicated and has critical issues related in the first step (information gathering) of LA such as ambiguousness and duplicates. The goal of this paper is to present the proposal of the framework used to eliminate duplicates in large datasets acquired with Big Data LA.

The duplication of Big Data reduces the efficiency of result analysis. Thus, a de-duplication technique has been introduced by integrating heterogeneous data in same format. Further, to analyze large amount of data, record level data duplication is being applied on integrated data that is also used for removing similar type of data. Finally, the experimental results validate the efficiency in terms of execution time, storage space and provide better analysis results for LA.

III. PROPOSED SYSTEM

In the proposed system, I have introduced the concept of deduplication to overcome the drawbacks with Big Data LA. For this work, I have implemented record level deduplication algorithm on learning data set to improve the efficiency of result for Big Data LA. It removes duplicates and reduce the size of data set to analyze.

IV. PROBLEM DEFINITION

The rise of education calls for basic research that can advance understanding about how students learn. A major research focus is on the process of LA. Because of the open-ended, project-based nature of education, students can produce a large quantity of data and artifacts while solving a complex design challenge, making it difficult to discern their learning.

To analyze the big data of users' behaviors that are recorded in the on line learning application logs (Eg.Moodle). Currently, the Moodle logs manager suffers from functional limitations and uninspiring visualizations. This method utilizes the possibility of downloading the logs in Microsoft Excel format and provides a simple and effective offline solution. The method I have developed is based on Excel macros.
Tutors can experiment with different combinations of metrics such as total page views, unique users, unique actions, IP addresses, unique pages, average session length and bounce rate. Furthermore, the software allows the definition of date ranges and the selection of individual or groups of students.

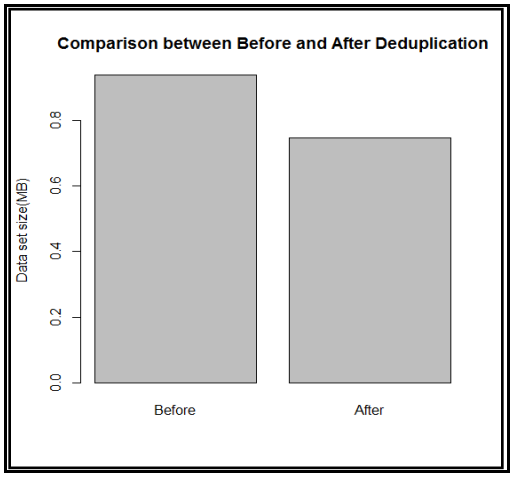
The complicated processes of analyzing and combining data are carried out in the background, enabling tutors to focus on the pedagogic implications and invest in practical, realistic scenarios through informed decision-making. Future work includes transferring the offline functionality to an online application plugin and increasing system intelligence to allow the production of meaningful and actionable suggestions with regards to set target goals [5].

V. SYSTEM DESIGN AND IMPLEMENTATION

Design is a multi-step process that focuses on data structure, software architecture, procedural details, (algorithms etc...) and interface between the modules. The design process also translates the requirements into the representation of software that can be accessed for quality before coding begins.

In this system, I have taken real world data set. The size of data set before (0.937MB) and after (0.745MB) applying record level deduplication was found and compared during the information gathering step.

The duplicated data was completely eliminated by using record level deduplication technique. The experiment evaluation is designed by using RStudio (RGUI (32-bit)) and comparison graph (R Graphics) for the size of data set before and after deduplication is shown below.



CONCLUSION

In this paper, I have implemented record level deduplication technique on big data in information gathering step of LA. In this paper, I have taken real world data set and apply record level deduplication technique and then compare the results. From the result, it shows that after applying the proposed method the duplicated data is removed and the size of big data taken for LA is reduced to reasonable size and may provide better result during information processing and visualization. The future research work towards to use deduplicated data set for LA.

REFERENCES

[1] Jacek Maślankowski, Learning Analytics: The Definitions, the Processes, and the Potential, Tanya Elias, January, 2011.

[2] Eckerson, W. W., (2006) , Performance dashboards: Measuring, monitoring, and managing your business. Hoboken, New Jersey: John Wiley & Sons.

[3] Hijon R. and Carlos, R. (2006). E-learning platforms analysis and development of students tracking functionality, in Proceedings of the 18th World Conference on Educational Multimedia, Hypermedia & Tele comunications, pp. 2823-2828.

[4] Towards de-duplication framework in big data analysis. A case study, Euro Symposium on Systems Analysis and Design, SIGSAND/PLAIS 2016: Information Systems: Development, Research, Applications, Education pp 104-113.

[5] Zhang, H., Almeroth, K., Knight, A., Bulger, M., and Mayer, R. (2007). Moodog: Tracking Students' Online Learning Activities in C. Montgomerie and J. Seale (Eds.), Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications, Chesapeake, VA: AACE, 2007, pp. 4415-4422. DOI:10.1.1.82.951.