

Chapter 4

Big Data Analytics: Educational Data Classification Using Hadoop–Inspired MapReduce Framework

Pratiyush Guleria

Himachal Pradesh University, India

Manu Sood

Himachal Pradesh University, India

ABSTRACT

Due to an increase in the number of digital transactions and data sources, a huge amount of unstructured data is generated by every interaction. In such a scenario, the concepts of data mining assume great significance as useful information/trends/predictions can be retrieved from this large amount of data, known as big data. Big data predictive analytics are making big inroads into the educational field because with the adoption of new technologies, new academic trends are being introduced into educational systems. This accumulation of large data of different varieties throws a new set of challenges to the learners as well as educational institutions in ensuring the quality of their education by improving strategic/operational decision-making capabilities. Therefore, the authors address this issue by proposing a support system that can guide the student to choose and to focus on the right course(s) based on their personal preferences. This chapter provides the readers with the requisite information about educational frameworks and related data mining.

INTRODUCTION

Educational data mining is one of the most promising areas for getting new insight into the trends and predictions in our educational systems supporting continuous integration of newer technologies and corresponding transformations. The inclusion of various modes of e-learning and other online educational resources into the teacher-taught paradigm, in the formal as well as informal education sectors, results into a collection of huge volumes of data. For this structured, semi-structured or unstructured data to make reasonable sense to the stakeholders of the systems, the emerging trends of data mining need to be explored for processing this data available on distributed systems with parallel computations.

DOI: 10.4018/978-1-5225-6210-8.ch004

With the adoption of these new mining techniques, educational sector is the beneficiary because of faster decision-making with a support from analyses of data fetched from students. The data from students and other stakeholders may include: a) preferences for the courses, course outcomes, trainings especially vocational trainings, industry, industry oriented courses as optional subjects, job profiles, etc.; b) choices of the appropriate existing subjects; c) available options at the national and international levels, and; d) in-house training needs for the employees and management and so on.

Big data is the emerging field that uses data mining and provides answers to resolve the problems arising due to accumulation of large amount of data obtained from academia. In big data, mining techniques process the data in the form of small chunks and distribute it on multiple machines for processing and finally aggregate it to present the results. Big data perform these operations with the help of its programming paradigm i.e. Hadoop. It is a framework for distributed processing of datasets so large and/or complex where traditional data processing applications are incompetent to deal with them. Big data is the term for collection of datasets which are large, complex and becomes difficult to process using on-hand database management tools. Big data includes gathering of data for storage and analysis purpose which gain control over operations like searching, sharing, visualization of data, query processing, updation and maintain privacy of information. In Big data, there is extremely large dataset that is analyzed computationally to reveal patterns, trends and associations. It deals with unstructured data which may include Microsoft Office files, PDF, Text etc whereas structured data may be the relational data. Hadoop is one technique of big data and answer to problems related to handling of unstructured and massive data.

In Educational sector, data mining is the area for getting new insight into educational system where with increase in technologies as well with transformation of class room teaching to online learning and other educational resources, it results into collection of huge volume of data which may be structured, semi-structured or completely unstructured. With data-processing and decision driven technologies like “Big Data”, educational sector will be the beneficiary as decision-making will be faster with analyses of data fetched from students in terms of their feedback for the courses, syllabi curriculum in preference for industry oriented courses as optional subjects, in-house trainings for the staff members etc.

Educational System Design

With the rapid development of technologies (Begona Gros, 2016), flexible and efficient learning methods for learners are being developed. The students usually acquire basic knowledge and core skills in the classroom. Learning goals and processes always are the same for each student in traditional classroom. But students (Tomlinson & McTighe, 2006) with different backgrounds have different needs. The interactions in the classrooms should therefore, be differentiable and responsive enough to accommodate the variations according to the learners readiness levels, interests and learning profiles (Tomlinson & Kalbfleisch, 1998). In a traditional classroom, the teacher is the main source of information and students are required to stay in the same place and participate simultaneously in the same set of activities, whereas in a situation of ubiquitous learning, activities can be conducted in a different space and time for each student. In addition, integrated teaching aids are also available to them all the time and are accessible from any device (Begona Gros, 2016). The paradigm shift in educational system from traditional classroom teaching to smart learning environment is shown in Figure 1(a) and 1(b).

30 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/big-data-analytics/219118

Related Content

A Role of Artificial Intelligence and Machine Learning Algorithms for Energy Efficiency Applications

M. Senbagavalli, Shekarappa G. Swetha and R. Rajagopal (2023). *Handbook of Research on Network-Enabled IoT Applications for Smart City Services* (pp. 115-134).

www.irma-international.org/chapter/a-role-of-artificial-intelligence-and-machine-learning-algorithms-for-energy-efficiency-applications/331329

Powering Healthcare IoT Sensors-Based Triboelectric Nanogenerator

Saeed Ahmed Khan, Shamsuddin Lakho, Ahmed Ali, Abdul Qadir Rahimoon, Izhar Hussain Memon and Ahsanullah Abro (2020). *IoT Architectures, Models, and Platforms for Smart City Applications* (pp. 29-51).

www.irma-international.org/chapter/powering-healthcare-iot-sensors-based-triboelectric-nanogenerator/243908

Automatic Address Scheduling and Management for Broadband IP Networks

Jun Bi, Chongfeng Xie, Chen Li and Qiong Sun (2019). *Emerging Automation Techniques for the Future Internet* (pp. 262-281).

www.irma-international.org/chapter/automatic-address-scheduling-and-management-for-broadband-ip-networks/214436

Reliability of IoT-Aware BPMN Healthcare Processes

Dulce Domingos, Ana Respício and Ricardo Martinho (2017). *Internet of Things and Advanced Application in Healthcare* (pp. 214-248).

www.irma-international.org/chapter/reliability-of-iot-aware-bpmn-healthcare-processes/170242

Green Energy Supply to Pre-School Cluster for Sustainable Anganwadi Educational Transformation

Dharmbir Prasad, Rudra Pratap Singh, Jatin Anand, Ranadip Roy and Md. Irfan Khan (2024). *Harnessing Green and Circular Skills for Digital Transformation* (pp. 86-107).

www.irma-international.org/chapter/green-energy-supply-to-pre-school-cluster-for-sustainable-anganwadi-educational-transformation/349632