Chapter 13
Using Big Data in Collaborative Learning

Liz Sokolowski
University of West London, UK

Samia Oussena
University of West London, UK

ABSTRACT

Big data emerged as a dominant trend for predictive analytics in many areas of industry and commerce. The study aimed to explore whether similar trends and benefits have been observed in the area of collaborative learning. The study looked at the domains in which the collaborative learning was undertaken. The results of the review found that the majority of the studies were undertaken in the Computing and Engineering or Social Science domains, primarily at undergraduate level. The results indicate that the data collection focus is on interaction data to describe the process of the collaboration itself, rather than on the end product of the collaboration. The student interaction data came from various sources, but with a notable concentration on data obtained from discussion forums and virtual learning environment logs. The review highlighted some challenges; the noisy nature of this data and the need for manual pre-processing of textual data currently renders much of it unsuitable for automated ‘big data’ analytical approaches.

INTRODUCTION

With the rise of pervasive computing and the internet of things (IoT), data of a quantity inconceivable just a decade ago is being generated and logged daily by machines and sensors, without the need for human intervention. With it has come the requirement for new methods to process and understand this data. Big data has emerged as a dominant trend for predictive analytics in many areas of industry and commerce, fuelling applications that have automated decision making and allowing timely interventions to be made based on patterns discovered through techniques such as data mining. From systems that calculate insurance premiums based on claims history, to fraud detection systems that automatically block transactions that are outside the limits of what a system considers to be a customer’s normal purchas-
Using Big Data in Collaborative Learning

ing behaviour, examples of the practical use of big data analytics abound: “Although largely unseen, it drives millions of decisions, determining who to call, mail, investigate, incarcerate, set up on a date, or medicate” (Ryu, 2013, p64).

This study aimed to explore whether similar trends, benefits and practical uses have been observed in Big Data generated in the course of collaborative working, by carrying out a systematic literature review of case studies. Employers are increasingly requiring graduates to enter the job market with a raft of soft skills that allow them to slip seamlessly and productively into the working environment. One of these is the ability to engage in team working (Robles, 2012) and to be able to collaborate effectively on any team projects set by the employer. As a result, team working and collaborative learning are increasingly being incorporated into student assessments. From an educational point of view, collaborative learning is also a desired educational goal, enabling students to generate a richer pool of ideas than if they were to work on their own, and to learn from each other through social interaction, promoting deeper understanding and constructivist learning (Jonassen, 1999). However it is often difficult to ensure a group operates effectively and productively, with all members experiencing high levels of satisfaction in the process, and this applies equally to teams in the workplace as well as to teams of students. This has prompted research into ways that data from group work processes and outcomes can be captured and analysed to predict a team’s performance and ultimately allow timely interventions to be made to help dysfunctional groups. Given that a lot of group interaction is now conducted electronically, this raises possibilities for mining Big Data generated from these interactions. In the field of education, research into mining learning data to predict an individual student’s success is now in its maturity. However, much less appears to exist on mining data to predict group work success, whether in the field of education or in the workplace.

BIG DATA

The term ‘Big Data’ is often described in terms of volume, variety and velocity: Volume: Big Data implies enormous volumes of data being generated primarily by websites, sensors, social media, and so on; Variety refers to the many sources and types of data, both structured and unstructured. Structured data is well defined and can easily be represented as numbers or categories, whereas unstructured can encompass textual information, videos, photos, which creates problems for storing and analysing such data. Velocity refers to the speed with which Big Data is generated - most is produced by machines, rather than by humans and the flow of data is often continuous and massive. Accelerating digitization taking place in all areas of industry and society has meant that Big Data is appearing at an increasing rate in every domain and is available for analysis in ever more creative ways.

This is also true in the context of collaborative learning, where large data sets are potentially available for analysis from students’ interactions with online learning and learning support systems. Learning Management Systems (LMS) such as Moodle and Blackboard record every action a student makes while using the platform, generating a digital footprint of their activity. As well as providing scope for a range of quantitative measures of student activity, these data logs can reveal associations between learners and the structure of networks to which they belong.

In addition, large volumes of textual data can be collected from emails, discussion boards, wikis, phone transcripts, and so forth. As this is primarily unstructured data it needs to be pre-processed and coded in some way prior to analysis.