Chapter 3
Big Data Mining

N. G. Bhuvaneswari Amma
Indian Institute of Information Technology Srirangam, India

ABSTRACT

Big data is a term used to describe very large amount of structured, semi-structured and unstructured data that is difficult to process using the traditional processing techniques. It is now expanding in all science and engineering domains. The key attributes of big data are volume, velocity, variety, validity, veracity, value, and visibility. In today's world, everyone is using social networking applications like Facebook, Twitter, YouTube, etc. These applications allow the users to create the contents for free of cost and it becomes huge volume of web data. These data are important in the competitive business world for making decisions. In this context, big data mining plays a major role which is different from the traditional data mining. The process of extracting useful information from large datasets or streams of data, due to its volume, velocity, variety, validity, veracity, value and visibility is termed as Big Data Mining.

1. INTRODUCTION

Big Data is originated due to the fact that huge amount of data is created every day like Google has more than 1 billion queries per day, Twitter has more than 250 million tweets per day, Facebook has more than 800 million updates per day, and YouTube has more than 4 billion views per day etc. These data are produced in the order of zetabytes, and it is growing around 40% every year (Wei & Albert, 2012). The need for Big Data Mining (BDM) is to extract useful information from large datasets because companies like Google, Apple, Facebook, Yahoo, Twitter are started to look carefully to these data to find useful patterns to improve their user experience. With the help of BDM, one can find useful pattern from mobile data too such as what the users do with the mobile.

1.1 Dimensions of Big Data

Big data differs from other data in seven dimensions such as Volume, Velocity, Variety, Validity, Veracity, Value and Variability. Table 1 shows the seven dimensions of big data with the characteristic of all the seven Vs. The data used in BDM must be based on the seven Vs.

DOI: 10.4018/978-1-5225-0182-4.ch003
1.2 Need for BDM

The main objective of BDM is knowledge extraction from big data which is used in the process of Business Intelligence (BI). Figure 1 shows the steps in BI. It consists of four steps namely Data Acquisition, Information Conversion, Knowledge Extraction, and Actionable Plans.

1.2.1 Data to Information

The process of determining what data is to be collected and managed and in what context.

1.2.2 Information to Knowledge

The process which involves the analytical components such as data warehousing, online analytical processing, data profiling, business rule analysis, and data mining.

Figure 1. Steps in business intelligence