Chapter 1
Interactive Visual Analytics of Big Data

Carson K.-S. Leung
University of Manitoba, Canada

Christopher L. Carmichael
University of Manitoba, Canada

Patrick Johnstone
University of Manitoba, Canada

Roy Ruokun Xing
University of Manitoba, Canada

David Sonny Hung-Cheung Yuen
University of Manitoba, Canada

ABSTRACT

High volumes of a wide variety of data can be easily generated at a high velocity in many real-life applications. Implicitly embedded in these big data is previously unknown and potentially useful knowledge such as frequently occurring sets of items, merchandise, or events. Different algorithms have been proposed for either retrieving information about the data or mining the data to find frequent sets, which are usually presented in a lengthy textual list. As “a picture is worth a thousand words”, the use of visual representations can enhance user understanding of the inherent relationships among the mined frequent sets. However, many of the existing visualizers were not designed to visualize these mined frequent sets. This book chapter presents an interactive next-generation visual analytic system. The system enables the management, visualization, and advanced analysis of the original big data and the frequent sets mined from the data.

INTRODUCTION

With advances in technology, high volumes of a wide variety of data can be generated easily. These include:

1. Structured data in relational or transactional databases,
2. Semi-structured data in text documents or the World Wide Web, and
3. Unstructured data in social media or networks.

DOI: 10.4018/978-1-5225-2058-0.ch001
Interactive Visual Analytics of Big Data

This leads us into the new era of big data (Keim et al., 2013; Zhang et al., 2013; Leung, 2014; Leung & Jiang, 2015). Intuitively, big data are interesting high-velocity, high-value, and/or high-variety data with volumes beyond the ability of commonly-used software to capture, manage, and process within a tolerable elapsed time. Hence, new forms of processing data are needed to enable enhanced decision making, insight, knowledge discovery, and process optimization. Moreover, embedded within these data is potentially useful knowledge that professionals, researchers, students, and practitioners want to discover. This calls for both

1. **Information retrieval** (Meng & Lu, 2013; Cuzzocrea et al., 2015), which returns explicit and highly relevant information or resources about data, and
2. **Data mining** (Frawley et al., 1991), which discovers implicit, previously unknown and potentially useful knowledge from data.

A common data mining task is **frequent set mining** (Agrawal et al., 1993), which analyzes the data to find frequently occurring sets of items (e.g., frequently collocated events, frequently purchased bundles of merchandise products). These frequent sets serve as building blocks for many other data mining tasks such as the mining of association rules, correlation, sequences, episodes, emerging patterns, web access patterns, maximal patterns, closed frequent sets, constrained patterns, weighted patterns, and social patterns (Pasquier et al., 1999; Pei et al., 2000; Lakshmanan et al., 2003; Leung et al., 2007; Kumar et al., 2012; Leung et al., 2012; Fariha et al., 2015; Leung et al., 2016). Moreover, these frequently occurring sets of items can be used in mining tasks like classification (Al-Rajab & Lu, 2016) such as **associative classification** (Liu, 2009). Frequent sets can also answer many questions that help users make important decisions for real-life applications in different domains such as health care, bioinformatics, social science, as well as business. For example, knowing the sets of frequently purchased merchandise helps store managers make intelligent business decisions like item shelving, finding the sets of popular elective courses helps students select the combination of courses they wish to take, and discovering the sets of frequently occurring patterns in genes helps professionals and researchers get a better understanding of certain biomedical or social behaviours of human beings.

Frequent set mining has drawn the attention of many researchers as it has played important roles in many data mining tasks and has contributed to various real-life applications. Since the introduction of the frequent set mining problem (Agrawal et al., 1993), numerous algorithms (Han et al., 2007; Cheng & Han, 2009; Leung et al., 2014; Jiang & Leung, 2015) have been proposed to mine frequent sets from databases. Most of these algorithms return the mining results in textual forms such as a very long unsorted list of frequent sets of items. Presenting a large number of frequent sets in such a conventional lengthy list does not lead to ease of understanding. Consequently, users may not easily discover the useful knowledge that is embedded in the databases.

As “a picture is worth a thousand words”, a visual representation matches the power of the human visual and cognitive system. Hence, having a visual representation of the frequent sets makes it easier for users (e.g., professionals, researchers, students, practitioners) to view and analyze the mining results when compared to presenting a lengthy textual list of frequent sets of items. This leads to **visual analytics**, which is the science of analytical reasoning supported by interactive visual interfaces (Thomas & Cook, 2005; Keim et al., 2008; Keim et al., 2009a; Keim et al., 2009b; Heimerl et al., 2016). Since numerous frequent set mining algorithms (which analyze large volumes of data to find frequent sets of items) have