Chapter 15
Leveraging Applications of Data Mining in Healthcare Using Big Data Analytics: An Overview

Mohammad Hossein Tekieh
University of Ottawa, Canada

Bijan Raahemi
University of Ottawa, Canada

Eric I. Benchimol
University of Ottawa, Canada

ABSTRACT

Big data analytics has been introduced as a set of scalable, distributed algorithms optimized for analysis of massive data in parallel. There are many prospective applications of data mining in healthcare. In this chapter, the authors investigate whether health data exhibits characteristics of big data, and accordingly, whether big data analytics can leverage the data mining applications in healthcare. To answer this interesting question, potential applications are divided into four categories, and each category into sub-categories in a tree structure. The available types of health data are specified, with a discussion of the applicable dimensions of big data for each sub-category. The authors conclude that big data analytics can provide more advantages for the quality of analysis in particular categories of applications of data mining in healthcare, while having less efficacy for other categories.

INTRODUCTION

While collecting, storing, and managing large amounts of digitized data are now technically feasible and affordable, only some useful information is still extracted from a small portion of the gathered data. To discover more information, strong analytical tools are needed for processing and analyzing the collected data, currently on the order of petabytes (Han, Kamber, & Pei, 2011). Data analysis algorithms have

DOI: 10.4018/978-1-5225-2515-8.ch015

also been developed to be able to handle big data collections. In addition, scalable and flexible software technologies have been introduced and are being improved to provide a suitable ecosystem to implement big data algorithms. The package comprising all these new components such as the technologies, algorithms, and methods is known as "big data analytics".

Data mining, as a strong analytical tool, has been applied to large amounts of digitized data collected in various fields – including healthcare – over the past decades. With the introduction of big data analytics, researchers are working to enhance data mining techniques to make the algorithms more scalable and faster. However, whether this enhancement resolves the existing limitations of data analysis studies in the field of healthcare remains unknown. It is necessary to first determine if all "health data" fit into the definition of "big data", before claiming big data analytics as the solution to overcome the limitations of health data analysis.

In this chapter, the authors investigate whether applications of data mining in healthcare can be leveraged by big data analytics by answering the following questions:

1. What are the applications of data mining in healthcare?
2. What are the different types of health data?
3. What are the characteristics of "big data"?
4. Is health data a form of "big data"?
5. Are all types of health data relevant in each application of data mining in healthcare?
6. To what extent do big data analytics enhance the quality of research in each application of data mining in healthcare?

In the introductory section, the application of data mining in healthcare is summarized, and the different types of health data and dimensions of big data are reviewed. Next, the methodology of achieving the above objective is presented and discussed in detail. Finally, the chapter will be concluded by summarizing the answers to the research questions.

BACKGROUND

Whether healthcare data can be considered "big data" is controversial. The phrase "health data" does not refer to a specific type or source of data. Some health data is gathered for specific research studies, but the majority is collected routinely without having pre-defined research questions in mind (Benchimol et al., 2015). There are many types of health data being collected routinely using various approaches, which will be presented later in this section. Often, the only characteristic they share is being related to the healthcare of patients. Each data type has its own characteristics and is collected for a specific reason, such as administration of a healthcare system. Since the majority of health data is not originally collected for research studies, they cannot necessarily be applicable for all types of data analysis studies. However, these health data instances can be valuable sources of information, and to which descriptive and predictive analytical tools such as data mining techniques can be applied to conduct novel analyses.

Data mining techniques have been applied to a large variety of data in many sectors and industries, such as healthcare, finance, retail, and telecommunication (Han et al., 2011). Although various applications of data mining in healthcare have been discussed in the literature, not all types of health data are suited to the requirements of each application. Due to the large and increasing volume of health data