Chapter 96
Knowledge Discovery and Data Mining Applications in the Healthcare Industry: A Comprehensive Study

Iman Barazandeh
Iran University of Science and Technology, Iran & Islamic Azad University, Mahshahr Branch, Iran

Mohammad Reza Gholamian
Iran University of Science and Technology, Iran

ABSTRACT

The healthcare industry is one of the most attractive domains to realize the actionable knowledge discovery objectives. This chapter studies recent researches on knowledge discovery and data mining applications in the healthcare industry and proposes a new classification of these applications. Studies show that knowledge discovery and data mining applications in the healthcare industry can be classified to three major classes, namely patient view, market view, and system view. Patient view includes papers that performed pure data mining on healthcare industry data. Market view includes papers that saw the patients as customers. System view includes papers that developed a decision support system. The goal of this classification is identifying research opportunities and gaps for researchers interested in this context.

INTRODUCTION

Since human learned to inscribe his thinks in the world out of his/her mind, Data has been created and started to growing and its growing accelerates through continuous advances in storing technology during the years and recent years are explosion age of data. Large and valuable volume of data is accumulated in databases and data warehouses in all domains. Online stores store sale details and customer information and interests in their databases. In banking industry account information and transactions are stored. In healthcare industry general patient information and his/her point of care information are stored in

DOI: 10.4018/978-1-5225-5643-5.ch096

databases. These days information is stored either digital or manual because it is proved that information and knowledge are the main success driver in every domain and industry.

However, what we can do with this large volume of data and how we can extract high level knowledge from low level and raw data. It is obvious that we can mine the data to find new and valuable relations and patterns. Pattern is an expression in some language describing a subset of the data or a model applicable to the subset and we can consider a pattern to be knowledge if it exceeds some interestingness threshold that is depends on domain and user definition (Fayyad, Piatetsky-Shapiro & Smyth, 1996). Extracted knowledge can be used to make more effective decisions.

For long years, statisticians used classical statistic methods for pattern identification. Statistics, especially as taught in most statistics texts, might be described as being characterized by data sets which are small and clean, which permit straightforward answers via intensive analysis of single data sets, which are static, which were sampled in an iid manner, which were often collected to answer the particular problem being addressed, and which are solely numeric. None of these apply in the data mining context (Hand, 1998). Data mining technology is presented to pass the constraints of statistic methods. Data mining is a technology that blends traditional data analysis methods with sophisticated algorithms for processing large volumes of data. It has also opened up exciting opportunities for exploring and analyzing new types of data (Tan, Steinbach & Kumar, 2005). Brossette and Hymel (2008) believe that the main tenet of data mining is that the models and patterns contain insights that were previously unsuspected. For that reason alone, data mining is not an exercise in hypothesis-driven exploratory statistics, or hypothesis-driven statistical model building, because "hypothesis-driven" implies previously suspected. Data mining is a new discipline lying at the interface of statistics, database technology, pattern recognition, machine learning, and other areas (Hand, 1998).

There are several definitions for data mining, but all of these definitions have a same understating of underlying concept and there are keywords that are common in all of them. Tan et al. (2005) define data mining as the process of automatically discovering useful information in large data repositories. From Fayyad et al. (1996) point of view knowledge discovery in databases (KDD) is the overall process of discovering useful knowledge from data, and data mining refers to a particular step in this process that is the application of specific algorithms for extracting patterns from data. Usefulness is depends to domain of problem and user definition. Data mining is always associated with analysis. Everywhere that analysis of a small or large data set is needed, data mining can be useful.

Healthcare industry is one of the most interesting areas in which data mining may have an important practical impact. In healthcare, data mining is becoming increasingly popular for several reasons: the extremely large amounts of data; the need for organizations to make decisions based on the analysis of clinical and financial data; and the power to generate information that is fundamentally useful to all parties involved in the healthcare industry (Santos, Malheiros, Cavalheiro & Parente de Oliveira, 2013). Databases are growing in hospitals, clinics, medical research centers, pharmaceutical companies and other related businesses. Researchers and practitioners of this industry seek for solutions to enable them using hidden patterns of data, to extract valid knowledge for more accurate and timely diagnosis, effective genetic data analysis, more effective care, drug discovery, drug repositioning, more effective monitoring and evaluating system, outlier detection, reducing errors, improving decision making for physicians and personnel performance, improving customer relationship management in hospitals and clinics, healthcare tourism and also better understanding of characteristics of care processes. To the best of our knowledge there is not a survey that investigates general applications of knowledge discovery and data mining in healthcare industry. Thus, this chapter proposes a new classification of applications of knowledge dis-