Chapter 14

Exploration of Healthcare Using Data Mining Techniques

Anindita Desarkar
Birla Institute of Technology Mesra – Kolkata, India

Ajanta Das
Birla Institute of Technology Mesra – Lalpur, India

ABSTRACT

Huge amount of data is generated from Healthcare transactions where data are complex, voluminous and heterogeneous in nature. This large dataset can be used as an ideal store which can be analyzed for knowledge discovery as well as various future predictions. So, Data mining is becoming increasingly popular as it offers set of innovative tools and techniques to handle this kind of data set whereas traditional methods have limitations for that. In summary, providing the better patient care and reduction in healthcare cost are two major goals of application of data mining in healthcare. Initially, this chapter explores on the various types of eHealth data and its characteristics. Subsequently it explores various domains in healthcare sector and shows how data mining plays a major role in those domains. Finally, it describes few common data mining techniques and their applications in eHealth domain.

INTRODUCTION

eHealth is an emerging field which uses the combination of medical informatics, public health and business to satisfy the needs of common people including patients, healthcare professionals, healthcare service providers and policy makers as well (Sartipi et al., 2008). Nowadays huge amount of complex and heterogeneous data is generated from healthcare industry rapidly. It includes data from various medical devices, electronic patient record, hospital resource details, diagnosis results etc. This large amount of dataset is a major source of resources which can be processed and analyzed for knowledge discovery. It can be used for future prediction, appropriate decision making and cost savings as well. But processing and analyzing become a challenge here as data volume is large and it’s diverse in nature. These challenges include technical difficulty of matching patients’ records, data exchange among various EHR systems and geographies and properly extract relevant information from the huge dataset so that it can be used effectively and efficiently by healthcare providers.

DOI: 10.4018/978-1-5225-5222-2.ch014
Data mining is a methodology of finding relevant, innovative and useful patterns in data. This includes various numbers of techniques like clustering, classification, data summarization, change analysis etc. It comes into the picture here as it offers a set of innovative tools and techniques which can be used to process and analyze this large amount of data to discover the hidden patterns. These tools have the ability to work with such complex dataset though the files are huge in volume and heterogeneously structured. Healthcare professionals are benefited to a large extend in the decision making by using this knowledge source. For instance, data mining can be the solution to find the most effective treatment for a specific condition, patterns can be identified which are due to medicine side effect or readmission in the hospital. It also helps the patient to identify the most cost effective treatment for a specific disease (Koh & Tan, 2011). There are good number of data mining tools already available in the market which includes ODND, NCC2 which is used in the area of heart disease, WEKA – used for the betterment of cancer patients, K-means clustering for brain cancer, WEKA 3.6 for HIV/AIDS, SNP for Hepatitis C, RST for kidney related problems.

The organization of the chapter is as follows. Background section talks about the related work. The next section describes various features and sources of big data and analyses the eHealth data in the light of big data. Section 2 explores eHealth data with the help of various well known data mining techniques. It includes brief overview of data mining, motivations behind exploration of e-Health data, major dimensions of healthcare industry where data mining applications are used, few well known data mining techniques along with their detailed application in the healthcare domain. Next section concludes the chapter.

**BACKGROUND**

Data mining which is a newer technology, came into picture predominantly only in 1994. After that it was used in almost every domains including healthcare management. Various data mining applications can be created to measure the efficiency of current medical treatments. It can generate a report regarding the most appropriate and effective action on that condition by analyzing the causes, symptoms and course of treatments. As example, the patient’s condition can be compared after applying different drugs for a specific disease to find the most appropriate and cost effective treatment. United Healthcare has implemented it by mining its treatment record data to provide better treatment to its patients in reduced cost (Durairaj & Ranjani, 2013).

Jayanthi Ranjan already presented the discovery and extraction of useful data patterns to find observable patterns from huge volume. It proves the advantages of data mining in pharma industry in decision making process along with other issues like adverse reactions to the drugs (Ranjan, 2007).

K. Srinivas, B. Kavitha Rani and Dr. A. Goverdhan had used classification based data mining techniques on huge data set to predict the likelihood of having heart diseases for a patient. It takes age, sex, blood pressure, blood pressure as inputs from the patient profile (Srinivas, 2010).

Shweta Kharya had discussed the application of various data mining techniques in breast cancer diagnosis and prognosis decision tree was found to be the best predictor having 93.26% accuracy on benchmark data as well as SEER data set. Elias Lemuye discussed the application of Apriori algorithm and WEKA 3.6 tool in predicting the HIV virus (Kharya, 2012).
Related Content

Content-Aware Caching for Cooperative Transcoding Proxies
Kyungbaek Kim, Byungjip Kim and Daeyeon Park (2008). Encyclopedia of Internet Technologies and Applications (pp. 112-118).
www.igi-global.com/chapter/content-aware-caching-cooperative-transcoding/16842?camid=4v1a

ALBA Cooperative Environment for Scientific Experiments
Andrea Bosin, Nicoletta Dessi, Maria Grazia Fugini, Diego Liberati and Barbara Pes (2008). Encyclopedia of Internet Technologies and Applications (pp. 52-58).
www.igi-global.com/chapter/alba-cooperative-environment-scientific-experiments/16833?camid=4v1a

Joint Angular and Time Diversity of Multi-Antenna CDMA Systems in Wireless Fading Channels
Feng She, Hsiao Hwa Chen and Hongyang Li (2013). Security, Design, and Architecture for Broadband and Wireless Network Technologies (pp. 1-14).
www.igi-global.com/chapter/joint-angular-time-diversity-multi/77406?camid=4v1a

On the Performance of Content Delivery Clouds
Mukaddim Pathan, James Broberg and Rajkumar Buyya (2012). Next Generation Content Delivery Infrastructures: Emerging Paradigms and Technologies (pp. 29-54).
www.igi-global.com/chapter/performance-content-delivery-clouds/66992?camid=4v1a