Chapter 6
For Better Healthcare Mining Health Data

Güney Gürsel
Command post of Gendarme Logistics, School ol of Technical and Auxiliary Forces, Turkey

ABSTRACT

Data mining has great contributions to the healthcare such as support for effective treatment, healthcare management, customer relation management, fraud and abuse detection and decision making. The common data mining methods used in healthcare are Artificial Neural Network, Decision trees, Genetic Algorithms, Nearest neighbor method, Logistic regression, Fuzzy logic, Fuzzy based Neural Networks, Bayesian Networks and Support Vector Machines. The most used task is classification. Because of the complexity and toughness of medical domain, data mining is not an easy task to accomplish. In addition, privacy and security of patient data is a big issue to deal with because of the sensitivity of healthcare data. There exist additional serious challenges. This chapter is a descriptive study aimed to provide an acquaintance to data mining and its usage and applications in healthcare domain. The use of Data mining in healthcare informatics and challenges will be examined.

INTRODUCTION

From primary care institutions to big healthcare centers, every healthcare organization uses an information system. These healthcare information systems (HCIS) store, process and retrieve healthcare data. Healthcare data is very valuable in today's world. By the help of rapidly developing healthcare informatics, there are efforts to use the valuable data stored electronically in HCIS databases to improve healthcare. Healthcare staff expects more than e-recording the data from HCISs. Besides using healthcare data for care giving, healthcare centers and academic centers use these data for education and research. Research in medical area is not limited to healthcare development such as developing new healing techniques and drugs. There are healthcare informatics fields such as structured data entry, constructing longitudinal patient data, image processing, etc. Any research area dealing with huge and valuable data such as medical domain requires creative techniques supported with computers and computer systems to utilize it. Data mining techniques are good examples for these required creative techniques.

Life expectancy has risen. We live longer when we compared to past. This improvement is tremendous for human being. But it has a cost. We spend more to healthcare when compared to past. The healthcare expenses are becoming unmanageable. Payers put pressure on healthcare organizations to decrease the cost. Then is it possible to decrease the cost while increasing the quality of healthcare? The answer is "it depends". It depends on the extent that we utilize the data and information we have.

It is stated that every healthcare institution uses an information system. During the daily healthcare service, every piece of patient data is recorded by means of HCISs. In short, we have huge amounts of patient data and it is growing every day. These huge amounts of data are called as "Big

Data". These "Big Data" should be used to get information and knowledge. The most fundamental challenge about Big Data is to explore the large volumes of data and extract useful information or knowledge for future use (Wu, Zhu, Wu, & Ding, 2014).

Data mining is one of the techniques to utilize this valuable healthcare data. There are many different definitions available in the literature. Drawing from the literature, we can define data mining as the process of analyzing data and discovering knowledge, patterns, associations, rules, anomalies, sequences that are non-trivial, implicit, previously unknown and potentially useful from databases.

Healthcare institutions are geographically distributed to ease healthcare access. Data about a patient can be scattered into many HCIS databases. There should be a technique to handle these scattered data to analyze and mine. Distributed data mining deals with this scattered data problem and will be covered through the chapter.

Patient health data is very sensitive. Researches about patient health data should comply with some ethical and legal issues to protect patient privacy as well as data mining studies. Privacy is term used for notion of the confidentiality and access restrictions of patients' protected health information (PHI) which contains sensitive and personal information (Sun, Zhu, Zhang, & Fang, 2012). Patient privacy refers to any medical data (medical condition, test result, payment information, etc.) in any form (paper, electronic, etc.) that belongs to a patient, the meaning is about what is protected and who will be permitted to use this information (Upstate Medical University, 2011). Protecting patient privacy in data mining will be a part and it is examined in the chapter.

The complexity and toughness of the medical environment is common information in the literature. The complexity and toughness of the medical domain is a challenge to every technology and application as well as data mining. In this chapter, the challenges specific to data mining will also be examined.

This chapter is a descriptive study that examines the concepts, issues related and techniques used in data mining in healthcare big data. The purpose of the study is to give an idea about the data mining tasks and techniques used in healthcare.

The chapter is organized as follows: In the background section, the universal basics, not related to healthcare only, of data mining will be covered. The main part is in the heading of "MEDICAL DATA MINING" comprising the contribution of the data mining to the healthcare, data mining techniques used in healthcare, temporal data mining in healthcare, distributed data mining in healthcare, challenges, privacy and security issues, data mining for healthcare management. In the end, are the future research directions and conclusion parts interpreting the chapter.

22 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/for-better-healthcare-mining-health-data/172554

Related Content

Fuzzy Logic-Based Predictive Model for the Risk of Sexually Transmitted Diseases (STD) in Nigeria
Jeremiah A. Balogun, Florence Alaba Oladeji, Olajide Blessing Olajide, Adanze O. Asinobi, Olayinka Olufunmilayo Olusanyaand Peter Adebayo Idowu (2020). *International Journal of Big Data and Analytics in Healthcare (pp. 38-57).*
www.irma-international.org/article/fuzzy-logic-based-predictive-model-for-the-risk-of-sexually-transmitted-diseases-std-in-nigeria/259987

Social Media Big Data Analytics for Demand Forecasting: Development and Case Implementation of an Innovative Framework
Rehan Iftikharand Mohammad Saud Khan (2022). *Research Anthology on Big Data Analytics, Architectures, and Applications (pp. 902-920).*
www.irma-international.org/chapter/social-media-big-data-analytics-for-demand-forecasting/291017

E-LearningFacultyModules.org
Roger McHaney, Lynda Spireand Rosemary Boggs (2014). *Packaging Digital Information for Enhanced Learning and Analysis: Data Visualization, Spatialization, and Multidimensionality (pp. 103-119).*
www.irma-international.org/chapter/e-learningfacultymodulesorg/80214

A Predictive Analytics Framework for Blood Donor Classification
Kavita Pabrejaand Akanksha Bhasin (2021). *International Journal of Big Data and Analytics in Healthcare (pp. 1-14).*
www.irma-international.org/article/a-predictive-analytics-framework-for-blood-donor-classification/277644

Blockchain and Machine Learning Framework for Financial Performance in Pharmaceutical Supply Chains
Russell Sadeghi, Ava Hajianand Meysam Rabiee (2023). *Advancement in Business Analytics Tools for Higher Financial Performance (pp. 112-128).*
www.irma-international.org/chapter/blockchain-and-machine-learning-framework-for-financial-performance-in-pharmaceutical-supply-chains/328300