

Applications of Data Mining in the Healthcare Industry

John Wang

Montclair State University, USA

Xiaohua Hu

Drexel University, USA

Dan Zhu

Iowa State University, USA

INTRODUCTION

The workflow of health care organizations involves the generation and collection of various kinds of data relating to clinical practices, clinical trials, patient information, resource administration, policies, and research. Traditionally, statistical techniques are used to derive some operational information from the data. Data mining, a new method, provides the opportunity to derive, in an exploratory and interactive manner, valuable health care knowledge in terms of associations, sequential patterns, classifications, predictions, and symbolic rules. Such inductively-derived health care knowledge can provide strategic insights into the practice delivery of health care.

As the medical field expands, it is the duty of each physician to evaluate and protect each patient from diseases, side effects, and medical mishaps. Armed with a scalpel, stethoscope, and other accretions, physicians are now armed with data mining as a tool for expanding their knowledge base. Data mining is available to every aspect within the health care industry. It is multifaceted and used in areas like insurance to detect fraud, the pharmaceutical industry to evaluate side effects of drugs, and even detection of certain diseases based on genetics.

BACKGROUND

The progress of data mining has been around for decades. Although data mining is an evolution of a field that can be traced back many years, the terminology

itself has been only been introduced in the 1990s. By using pattern recognition technologies and statistical and mathematical techniques to sift through warehoused information, data mining helps analysts recognize significant facts, relationships, trend, patterns, exceptions, and anomalies. The use of data mining can advance a company's position by creating a sustainable competitive advantage. Data mining is the science of managing and analyzing large datasets and discovering novel patterns (Davenport & Harris, 2007; Olafsson, 2006; Wang, 2006).

Data mining is becoming more prevalent in the health care industry because of the vast quantities of data stored in a multitude of medical systems, more specifically systems of health care providers, hospitals, and other medical institutions (Veletos, 2003). Recognizing patterns of data in order to discover valuable information, new facts, and relationships among variables are important in making business decisions that would best minimize costs, maximize returns, and create operating efficiency without compromising the quality of patient care.

The growth in data mining and its advantages over the traditional methods of collecting data in the industry, both for reducing costs and providing patient care, will be discussed. The bulk of the material focuses on various applications of data mining in the health care industry under three major areas. These areas include: data mining in health insurance plans; data mining in clinical care, and data mining in health care administration. Also, emerging future trends in data mining in the health care industry will also be explored.

MAIN FOCUS

Data Mining in Health Insurance Plans

Financially oriented applications, such as the actuarial cost-risk models utilized by health insurance companies, as well as models used to track and identify potential claim fraud, have been the primary focus of data mining in the health care industry. Along with these applications, predictive models have also been applied to issues that impact cost, such as predictions of length of stay, forecasting treatment costs, total charges, and even mortality. These applications contribute significantly to the driving force that has helped this industry evolve to the gigantic sector that it is in the market: how can health care organizations minimize cost, raise its quality of service, and maintain its competitive edge at the same time.

Fraud Detection

Research indicates that health care fraud and abuse costs the U.S. as much as \$100 billion per year. That equates to about 10% of annual spending on health care. Numbers like that represent a huge red flag for action in this industry. Fraudulent and abusive behavior has become a serious problem. It is an area that requires significant attention and insight in order to prevent the unnecessary expenses associated with fraudulent transactions. Fraud detection takes on many forms including: claim fraud, premium avoidance fraud, and indemnity fraud (as in the case of disability or workers compensation insurance).

Claim fraud is currently the most predominant form of fraudulent activity and is a major concern of health care organizations. Medical claim fraud takes place when medical providers deliver unnecessary treatments and services or file claims to the insurance companies for services that have not been rendered in actuality. Due to the bulk of medical claims being processed side by side with a great multitude of insured patients and medical providers, detection of fraudulent claims poses a great challenge to health insurance companies. There is a considerable degree of difficulty involved in discerning whether a particular transaction is valid or illegitimate due to the enormous quantity of claims being processed.

Another type of medical fraud mentioned above is *premium avoidance fraud* which can occur under

a variety of circumstances. Some examples include: (1) A policy holder attempting to reduce premiums by providing misinformation to the insurance carrier, such as fewer employees or lower payroll; (2) A policy holder submits incorrect classification to qualify for lower premium rates; (3) A policy holder changes insurance to cover injured employee and falsifies accident date; or (4) An employee changes a date-of-birth for a dependent that would otherwise not qualify for coverage.

Indemnity fraud, on the other hand, occurs when an insured customer receives workers compensation indemnity benefits (exemption from paying incurred liabilities) when in reality, they are no longer qualified to do so due to reasons such as capability of returning to work or discontinuation of medical treatments.

All of the forms of fraudulent activities mentioned above represent a consequential opportunity for high-value data mining applications in the insurance industry.

Yang and Hwang (2006) proposed a process-mining framework that uses clinical pathways to distinguish fraudulent and abusive cases from normal ones. The framework for their research consisted of first breaking the health care industry down by service groups. Then, for each care service, a set of features is identified and an inductive model is developed to detect suspicious claims for each particular service group. All of these methods reduce the workload of human experts.

Policy Holder Retention

In order to sustain desired revenues, programs desired to retain current members are vital to a medical insurance organizations. A significant threat to the insurance industry is a reduction in the number of policy holders. It has been recognized that retention of existing customers is about 5-10 times more cost effective than acquiring new ones (Haux, Elske, Werner, & Knaup, 2002). Therefore, it is considered wise for insurance companies to distinguish members who are at risk of switching their membership to another provider.

Data mining can be used to identify the potential loss of clients switching to other companies. Historical data contains usage patterns and other important customer characteristics that when discovered can be used to identify satisfied or unhappy policyholders. These correlations between historical data values are very complex. Predicting policyholders who are at risk of changing their service to another provider with a

4 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/applications-data-mining-healthcare-industry/12924

Related Content

GIS Based Health Information Management through LETL, Multi Criteria Query, Analysis, Visualization

Sunitha Abburu (2019). *International Journal of E-Health and Medical Communications* (pp. 38-54).
www.irma-international.org/article/gis-based-health-information-management-through-letl-multi-criteria-query-analysis-visualization/215342

On the Integration of Clinical Archetypes with Ontologies and Rules

Leonardo Lezcano (2013). *Interoperability in Healthcare Information Systems: Standards, Management, and Technology* (pp. 82-147).
www.irma-international.org/chapter/on-the-integration-of-clinical-archetypes-with-ontologies-and-rules/106576

Adoption of Electronic Health Care Records: Physician Heuristics and Hesitancy

Jerald D. Hatton, Thomas M. Schmidt and Jonatan Jelen (2013). *Information Systems and Technologies for Enhancing Health and Social Care* (pp. 148-165).
www.irma-international.org/chapter/adoption-electronic-health-care-records/75626

Robotics and Tele-Rehabilitation: Recent Advancements, Future Trends

Amir Mobini, Saeed Behzadipour and Mahmoud Saadat Foumani (2013). *International Journal of Reliable and Quality E-Healthcare* (pp. 1-13).
www.irma-international.org/article/robotics-and-tele-rehabilitation/103828

A Content-Based Approach to Medical Images Retrieval

Mana Tarjoman, Emad Fatemizadeh and Kambiz Badie (2013). *International Journal of Healthcare Information Systems and Informatics* (pp. 15-27).
www.irma-international.org/article/a-content-based-approach-to-medical-images-retrieval/78928