ELSEVIER

Expert Systems with Applications 31 (2006) 56–68

Expert Systems with Applications

www.elsevier.com/locate/eswa

A process-mining framework for the detection of healthcare fraud and abuse

Wan-Shiou Yang [a,*], San-Yih Hwang [b]

[a] Department of Information Management, National Changhua University of Education, No. 1, Jin-De Road, Changhua 500, Taiwan, ROC
[b] Department of Information Management, National Sun Yat-sen University, No. 70, Lien-Hai Road, Kaohsiung 80424, Taiwan, ROC

Abstract

People rely on government-managed health insurance systems, private health insurance systems, or both to share the expensive healthcare costs. With such an intensive need for health insurances, however, health care service providers' fraudulent and abusive behavior has become a serious problem. In this research, we propose a data-mining framework that utilizes the concept of clinical pathways to facilitate automatic and systematic construction of an adaptable and extensible detection model. The proposed approaches have been evaluated objectively by a real-world data set gathered from the National Health Insurance (NHI) program in Taiwan. The empirical experiments show that our detection model is efficient and capable of identifying some fraudulent and abusive cases that are not detected by a manually constructed detection model.

Keywords: Healthcare fraud; Healthcare abuse; Clinical pathways; Classification model; Data mining

1. Introduction

Healthcare has become a major focus of concern and even a political, social, and economics issue in modern society. The medical expenditure required to meet public demand for high-quality and high-technology services is substantial. This phenomenon is likely to become more widespread and more intense due to the increasing average lifespan and decreasing birth rates of humans in many societies. People rely on health insurance systems, which are either sponsored by governments or managed by the private sector, to share the high healthcare costs.

Such an intensive need for health insurance has resulted in fraudulent and abusive behavior becoming a serious problem. According to a report (Health Insurance, 1992) published by the General Accounting Office in the US, healthcare fraud and abuse costs the US as much as 10% of its annual spending on healthcare, representing US$ 100 billion per year. Similar problems have been reported for the health insurance programs of other developed countries (Lassey, Lassey, & Jinks, 1997). The above figures indicate that detecting healthcare fraud and abuse is imperative.

Detecting healthcare fraud and abuse, however, needs intensive medical knowledge. Many health insurance systems rely on human experts to manually review insurance claims and identify suspicious ones. Most of the computer systems that are intended to help detect undesirable behavior require human experts to identify a set of features so as to develop the core of detection models. This results in both system development and claim reviewing being time-consuming, especially for the large government-sponsored national insurance programs in countries such as France, Australia, and Taiwan.

In this research, we propose a process-mining framework that utilizes the concept of clinical pathways to facilitate the automatic and systematic construction of an adaptable and extensible detection model. We take a data-centric point of view and consider healthcare fraud and abuse detection as a data analysis process. The theme of our approach is to apply process-mining techniques to gathered clinical-instance data to construct a model that distinguishes fraudulent behaviors from normal activities. This automatic approach eliminates the need to manually analyze and encode behavior patterns, as well as the guesswork in selecting statistics measures. The proposed framework is evaluated via real-world data to demonstrate its efficiency and accuracy.

This paper is organized as follows. Section 2 examines in more detail the problem of healthcare fraud and abuse. Related research efforts are also reviewed. Section 3 presents the process-mining framework of our research. Sections 4 and 5 describe in detail the methods for building the detection model. Section 6 presents the results of an evaluation of the detection power of the model on a real-world data set gathered from the

* Corresponding author. Tel.: +886 4 723 2105x7611; fax: +886 4 721 1162.
E-mail address: wsyang@cc.ncue.edu.tw (W.-S. Yang).

0957-4174/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2005.09.003

National Health Insurance (NHI) program in Taiwan. Section 7 concludes the current work and discusses the directions of future research.

2. The problem and the related work

The processing of health insurance claims involves three parties: service providers, insurance subscribers, and insurance carriers. The National Health Care Anti-Fraud Association defined healthcare fraud as 'an intentional deception or misrepresentation made by a person, or an entity, with the knowledge that the deception could result in some unauthorized benefit to him or some other entities' and healthcare abuse as 'the provider practices that are inconsistent with sound fiscal, business, or medical practices, and result in an unnecessary cost, or in reimbursement of services that are not medically necessary or that fail to meet professionally recognized standards for health care' (Guidelines to health care fraud, 1991).

The above definitions indicate that undesirable behavior can be performed by any of the three parties. However, further studies (He, Wang, Graco, & Hawkins, 1997; Pflaum & Rivers, 1991; Health care fraud, 2002) suggest that service providers account for the greatest proportion of fraud and abuse. The perpetrators of some types of fraud schemes (e.g. surgeries, invasive testing, and certain drug therapies) even deliberately and callously place their trusting patients at significant physical risk. Therefore, in this research, we focus on the detection of fraudulent and abusive behavior by service providers.

Currently, detecting such fraud and abuse relies heavily on medical knowledge. The carriers of nearly every insurance program around the world employ experts, who are pre-eminent in their specialty, to detect suspicious claims in their programs (He et al., 1997). These experts review medical claims, and verify the necessity of services according to the conditions of the patients. It is clear that this task is both effort and time-consuming, especially in the case of large-scale insurance programs.

The huge human effort raises the need of using information techniques to detect suspicious cases (Frieden, 1992). EFD, an expert system developed by Travel-Insurance Companies (Major & Riedinger, 1995), utilizes micro-knowledge (behavioral heuristics) coupled with information theory to select rules for performing the task of identifying outliers (fraud). Herb and Tom (Herb & Tom, 1995) derived fraud indicators and rules from the knowledge and experience of human experts to develop a computer-based expert to facilitate the work of insurance carriers.

Another line of research focuses on the use of more recent machine learning technologies. In this approach, features are often identified by expert consultants and used in the subsequent development of induction schemes. For example, the research by Sokol et al. (Sokol, Garcia, Rodriguez, West, & Johnson, 2001; Sokol, Gaarcia, West, Rodriguez, & Johnson, 2001), funded by the Health Care Financing Administration and the Office of the Inspector General in US, built a model that aimed to discriminate between normal and suspicious claims. For each care service, such as chiropractic services, laboratory/radiology procedures, and preventive medical services, a set of features is identified and an inductive model is accordingly developed to detect suspicious claims in a particular care service.

The work reported in Hall (1996); He et al. (1997)), funded by the Health Insurance Commission of Australia, aims to detect service providers who are practicing inappropriately. In this work, after discriminating features are determined (typically 25–30 features identified by specialists), a fuzzy–logic, neural-network-based induction algorithm is used to generate the detection model. This model is then used to tag suspicious service providers. Similarly, the work reported by Blue Cross and Blue Shield organizations (Cox, 1995) uses a fuzzy-based system to manage the claim profiles of service providers.

While the above-mentioned approaches reduce the workload of human experts, the enormous knowledge engineering task of identifying either discriminating rules or discriminating features is still required. Moreover, due to the manual and ad hoc nature of the development process, the resultant prototypes have limited extensibility and adaptability.

3. Research framework

In this section we introduce the concept of clinical pathways and our proposed process-mining framework that distinguishes fraudulent and abusive cases from normal ones.

The concept of *clinical pathways* (or *integrated care pathways*) was initiated in the early 1990s, and defined as 'multidisciplinary care plans, in which diagnosis and therapeutic intervention are performed by physicians, nurses, and other staff for a particular diagnosis or procedure' (Healy et al., 1998; Ireson, 1997). Clinical pathways are typically driven by physician orders and clinical industry and local standards of care. Once the pathways are created, they are viewed as algorithms of the decisions to be made and the care to be provided to a given patient or patient group. For example, the pathway of cholecystectomy (Ireson, 1997) begins with the preadmission process, which mainly involves preadmission testing and anesthesia consultation, goes though several assessments, surgery, and physician orders, and ends with a follow-up visit at the surgeon's office.

The application of clinical pathways is an efficient approach to analyzing and controlling clinical care processes. It aims to have medical staff performing the care services in the *right order*, enabling best practice—without rework and resource waste—to be implemented. Consider the cholecystectomy pathway. Care activities are sequenced on a timeline so that physicians can make suitable orders in accordance with the test results in the preadmission step, and anesthetic can be executed during the performance of surgery on the basis of anesthesia consult. In today's competitive healthcare environment, the competition advantage of a healthcare institution relies not only on outstanding professional quality but also on the agility of clinical care processes, and so the concept of clinical