Expert Systems with Applications 39 (2012) 7441–7450

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A scoring model to detect abusive billing patterns in health insurance claims

Hyunjung Shin [a], Hayoung Park [b,*], Junwoo Lee [a], Won Chul Jhee [c]

[a] Department of Industrial and Information Systems Engineering, Ajou University, San 5 Woncheon-dong, Yeongtong-gu, Suwon 443-749, Republic of Korea
[b] Technology Management, Economics, and Policy Graduate Program, Seoul National University, 599 Gwanak-ro, Gwanak-gu, Seoul 151-744, Republic of Korea
[c] Department of Industrial and Information Engineering, Hongik University, 72-1 Sangsu-dong, Mapo-gu Seoul 121-791, Republic of Korea

ARTICLE INFO

Keywords:
Health insurance claims
Medical abuse detection
Fraud detection
Degrees of anomaly
Data mining

ABSTRACT

We propose a scoring model that detects outpatient clinics with abusive utilization patterns based on profiling information extracted from electronic insurance claims. The model consists of (1) scoring to quantify the degree of abusiveness and (2) segmentation to categorize the problematic providers with similar utilization patterns. We performed the modeling for 3705 Korean internal medicine clinics. We applied data from practitioner claims submitted to the National Health Insurance Corporation for outpatient care during the 3rd quarter of 2007 and used 4th quarter data to validate the model. We considered the Health Insurance Review and Assessment Services decisions on interventions to be accurate for model validation. We compared the conditional probability distributions of the composite degree of anomaly (CDA) score formulated for intervention and non-intervention groups. To assess the validity of the model, we examined confusion matrices by intervention history and group as defined by the CDA score. The CDA aggregated 38 indicators of abusiveness for individual clinics, which were grouped based on the CDAs, and we used the decision tree to further segment them into homogeneous clusters based on their utilization patterns. The validation indicated that the proposed model was largely consistent with the manual detection techniques currently used to identify potential abusers. The proposed model, which can be used to automate abuse detection, is flexible and easy to update. It may present an opportunity to fight escalating healthcare costs in the era of increasing availability of electronic healthcare information.

1. Introduction

Increasing healthcare costs have burdened the economies of almost every developed and developing country, and the problem is worsening with an aging population and advancing health technology (Organisation for Economic Co-operation, 2008; 2009). Ongoing efforts against medical abuse and fraud include steps to reduce inappropriate use of healthcare funded by third-party payers, but the process is costly (Center for Medicare, 2008a; Feldman, 2001; Pontell, Jesilow, & Geis, 1982; Rai, 2001; Shane, 2000). Various estimates suggest that the magnitude of the problem, measured as a percentage of the healthcare budget, would range from 3 to 10% in the United States; however, key statistics for South Korea are unavailable. The Improper Medicare FFS Report indicated that 3.7% of US Medicare payments were inappropriate, which amounted to 10.2 billion USD in FY 2007, and in its 2007 crimes report, the US Federal Bureau of Investigation (FBI) estimated the figure to be as high as 10% (Center for Medicare, 2008b; National Health Care Anti-Fraud Association, 2009). The FBI identified healthcare fraud schemes such as billing for unprovided services, upcoding services and items for higher payments, submitting duplicate claims, unbundling services that should be billed as a single item, providing medically excessive and unnecessary services, and kickbacks (US Federal Bureau of Investigation, 2009). Medical abuse and fraud compromise both healthcare costs and quality. They also harm honest and ethical healthcare providers.

The detection of abusive and fraudulent practice in healthcare is difficult because uncertainties inherent in medical practices result in variable care processes (Eisenberg, 2002; Henderson, 2009). Therefore, medical experts must review each case, which can be time consuming and expensive. Advances in information technology and digitization of healthcare information, such as electronic medical records, bills, and claims, opened a new venue for efficient and effective medical abuse and fraud detection. Data mining and machine learning technologies have been widely used for fraud detection and auditing in the auto and life insurance, banking, credit card, and mortgage industries, and since the late 1990s, similar efforts have been made in healthcare (Hager et al., 2006; Li, Huang, Jin, & Shi, 2008).

Interest in fraud detection research has been gaining strength mainly in developed countries, and the scope of the research is expanding (Phua, Lee, Smith, & Gayler, 2005). Fraud detection models are most often cited for the national security, industrial

* Corresponding author. Tel.: +82 2 880 2575; fax: +82 2 886 8220.
E-mail addresses: shin@ajou.ac.kr (H. Shin), hayoungpark@snu.ac.kr (H. Park), neo100in@ajou.ac.kr (J. Lee), jhee@hongik.ac.kr (W.C. Jhee).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2012.01.105

information security, credit card, e-commerce, insurance, and telecommunication industries. Traditional statistical methods, data mining algorithms, and new machine-learning methods are used in the detection models. Although various algorithms are applied depending on the nature of the problem aimed at in the healthcare domain, neural network algorithms which result in superior performance and decision tree algorithms which render easy to understand results are two of the most popular methodologies (Bonchi, Giannotti, Mainetto, & Pedreschi, 1999; He, Hawkins, Graco, & Yao, 2000; Major & Riedinger, 2002).

Newer technologies are introduced to detect fraud and recently researchers have combined multiple methodologies. For example, some researchers used fuzzy logic in medical claims assessment, a combination of heuristic searches and the activity rule group for fast data review, and neural network algorithms to automatically classify claims information (Brockett, Xia, & Derrig, 1998; Cox, 1995). Also a combination of a neural network and a genetic algorithm along with application of Naive Bayes was attempted in the assessment of fraud in claims (Viaene, Richard, & Dedene, 2005). Koh and Tan (2005) introduced cases of data mining application, including for fraud and abuse detection, in the broad spectrum of healthcare management.

Attempts to apply data mining methodologies at the national and state level can be found in the previous studies as well. The Health Insurance Commission of Australia, which administrates the Medicare program for the Australian federal government, used online unsupervised learning algorithm based on finite mixture model to detect outliers in the utilization of pathology services (Yamanishi, Takeuchi, Williams, & Milne, 2004) and a combination of two neural network algorithms, the multi-layered perceptron (MLP) and Self Organizing Map, to identify abnormal patterns from the practice profiles of general practitioners (He, Wang, Grac, & Hawkins, 1997). The National Health Insurance (NHI) program of Taiwan developed disease-specific clinical pathways to identify fraudulent claims. The detection model, based on a process mining framework, automatically and systematically identified practices that deviated from the pathways, which could indicate abuse and fraud (Yang & Hwang, 2006). Also the NHI attempted to apply a model that combined fuzzy sets theory and a Bayesian classifier to a claims audit (Chan & Lan, 2001). The application of MLP neural networks in medical abuse and fraud detection enabled a Chilean private health-insurance company to install a real time-based detection process that brought considerable savings to the company (Ortega, Figueroa, & Ruz, 2006).

Since 1989, the entire population of South Korea (49.5 million) has been covered by a uniform insurance policy administered by the National Health Insurance Corporation (NHIC), except for approximately 3.6%, who are covered through the medical aid program funded by the general tax. Physicians and hospitals are reimbursed based on a fee-for-service mechanism, based on a fee schedule predetermined annually by the government. Although fees are strictly regulated by the government, the system is vulnerable to providers' abusive utilization and billing behavior, which causes unnecessary increases in healthcare costs. The insurer instituted a prepayment claims review and audit process to prevent improper utilization from reimbursement, and the Health Insurance Review and Assessment Services (HIRA) is dedicated to claims review and audit. Despite all the concerted efforts by the NHIC and HIRA, between 1990 and 2007 the health insurance budget expanded at an average annual rate of 16% (Organisation for Economic Co-operation & Development, 2009; Health Insurance Review & National Health Insurance Corporation, 2008).

The magnitude of information reviewed and audited by the 1700 employees of the HIRA is enormous and growing fast. In 2007, a total of 968 million claims were submitted to the HIRA for reimbursement, and nearly 37% of them were outpatient claims submitted by clinics (or physician practices). The average annual growth rate of claims between 2000 and 2007 was 13% and the average size of an outpatient claim was 1.9 KB (8.7 KB for inpatient claims). At first, every claim was manually reviewed to determine the amount of reimbursement, but organizational expansion proved politically infeasible, HIRA staff quickly realized that the practice was unsustainable. HIRA management saw opportunities in information technology and focused a strategy to simultaneously enhance the effectiveness and efficiency of the organization.

Claims were digitized starting in 1994 and the Electronic Data Interface (EDI) based billing system was introduced in 1996. As of 2008, 97% of 78,410 clinics, hospitals, and pharmacies submitted electronic claims. A data warehouse (DW) was built in 2003 so reviewers were better equipped with knowledge extracted from claims information. The size of the DW, which keeps 5-year claims information, was 142 TB in 2008. In another initiative, launched in 2002 to capitalize on the majority of claims being in electronic forms, reviewers focused on potential abusers to prevent waste, the Comprehensive Intervention Program, instead of post-utilization reviews. Under this program, machines do most of the post-utilization reviews on outpatient records while the reviewers undertake manual review to detect and educate and communicate with the small percentage of providers with abusive utilization behavior. Twenty-six thousand clinics submitted 67% of the outpatient claims, but the contents tended to be simple compared to the inpatient and outpatient claims submitted by hospitals.

Reviewers manually selected clinics based on approximately 180 indicators routinely computed in the DW using individual providers' claims data. Some examples of the indicators that comprise the case-mix adjusted costliness indices (CIs) for total charges and charges for categories of services such as IV, procedures, antibiotics, expensive medications, and lab work. The case-mix adjusted indicators of intensities of utilization also include data on the numbers of prescription medications and the days medications are prescribed as well. All the information characterized by about 180 indicators is difficult to amalgamate manually, therefore providers were selected for further investigations based on rankings of individual indicators regardless of the significance of the problem found. For example, a provider ranked in the top 3‰ for one indicator but below 50‰ in all other indicators could be selected, but one that ranked in the top 10‰ for all indicators may not be selected. The selection process based on these rankings shows obvious flaws. Furthermore, the manual selection of clinics with abusive billing patterns grew increasingly complicated because new treatments and medicines increased the information considered in the selection. The process has been criticized for lacking rationality, consistency, and interpretability.

We formulated a model that detects healthcare providers who show a pattern of abusive behavior in the provision of outpatient care. The proposed model was designed to automatically process large amounts of information contained in healthcare insurance claims and to generate an index that can be used to decide whether further investigation of the practitioner for subsequent intervention is warranted. We also applied the decision tree method to create clear explanations about the characteristics that make a provider a suspect of abuse.

2. Material and methods

2.1. Modeling

The proposed model is designed through two phases of modeling: scoring and segmentation. The scoring model quantifies the degree of abusiveness in a provider's billing pattern and the segmentation model groups providers based on the resulting