A Comprehensive Study of Healthcare Fraud Detection based on Machine Learning

Shivani S. Waghade

Department of Computer Science and Engineering,
Shri Ramdeobaba College of Engineering and Management, Nagpur - 440 013 (M.S.), India.

Prof. Aarti M. Karandikar

Assistant Professor, Department of Computer Science and Engineering,
Shri Ramdeobaba College of Engineering and Management, Nagpur - 440 013 (M.S.), India.

Abstract

Healthcare is an essential in people's lives and it must be affordable. The healthcare industry is an intricate system with numerous moving components. It is expanding at an expeditious pace. At the same time, fraud in this industry is turning into a critical problem. One of the issues is the misuse of the medical insurance systems. Manual detection of frauds in the healthcare industry is a strenuous work. Recently, machine learning and data mining techniques are used for automatically detecting the healthcare frauds. In this paper, we attempt to give a review on frauds in healthcare industry and the techniques for detecting such frauds. With an emphasis on the techniques used, determining the significant sources and the features of the healthcare data, various available researches were studied in the literature work. From this review it can be concluded that the advanced machine learning techniques and incipiently acquired sources of the healthcare data would be forthcoming subjects of interest in order to make the healthcare affordable, to improve the effectiveness of healthcare fraud detection and to bestow top quality on healthcare systems. Many recent researches, as reviewed in this paper, use machine learning and data mining to detect fraud in healthcare industry. There is a need additional research work to determine different unusual patterns of misuse of health insurance systems and more sophisticated machine learning techniques can be used to improve results.

Keywords: Machine Learning, Fraud Detection, Healthcare, Review, Anomaly Detection

INTRODUCTION

Healthcare has and perpetuates to be an integral component in people's lives. The human body is a compound structure. Hence, it is essential to have specialist physicians qualified to diagnose and treat diseases in different parts of the body. This induces several types of treatment procedures that physicians carry out for patients in different specialties. The aim of the health industry is to successfully serve as many patients as possible. But with every treatment there is a price associated with every service provided. Physicians, drug dealers and medical staff have to be paid for their time and prowess including various medical amenities. Oftentimes these prices are not affordable to the patients. Therefore, insurance schemes are used to dispense costs over all patients in the healthcare system and pay for the requisite people and equipment. As with any insurance system, there is a possibility for misuse or fraud activities.

Healthcare fraud is increasingly apperceived as one of serious social concerns. Clearly, healthcare fraud is a problem for the government and there is a need for more effective detection methods. To detect healthcare fraud, it requires great amount of efforts with extensive medical knowledge.

Traditionally, healthcare fraud detection greatly depends on the experience of domain experts, which is erroneous enough, expensive and time consuming. Manual detection of healthcare fraud involves a few auditors who manually review and identify the suspicious medical insurance claims which requires much effort. But the modern advances of machine learning and data mining techniques led to more efficient and automated detection of healthcare frauds. There has been a growing interest in mining healthcare data for fraud detection in the recent years. This paper reviews the various approaches used for detecting the fraudulent activities in Health insurance claim data.

RELATED WORK

Healthcare fraud has considerably inflated loss for individuals, entities and governments. Combating healthcare fraud has turn out to be a vital concern. Hence, several researchers have developed healthcare fraud detection systems. The job of fraud detection systems is to find, detect and report frauds as they appeared in the system [1], [2].

Commonly, there are two modes in which Fraud detection is rendered. Earlier, in order to detect the fraud, manual fraud audit regulations were implemented [3]. The process of auditing needs abundant knowledge of that field and prowess. These procedures emanate from intricate transactions and takes a lot of time. It involves monotonous and time-consuming manual work. Thus, automatic systems were created to detect frauds efficiently. These complex systems are computer based and integrate a large number of methods and approaches

involved in data mining [1], [3]. Hence, types of fraud, healthcare data and methods for detecting frauds has to be taken for contemplation.

A. Types of Frauds in Healthcare

Healthcare fraud has different fraudulent behaviors change to the occasion. It is a specific topic for every country. There are different types of fraud that occur in the healthcare industry. The types of frauds can be classified on the basis of which group or individuals are engaged in the fraud [4], [5]:

- Fraud by Service Providers

    – Service providers' may bill for the medical services that are not actually performed;

    – Service providers' may bill for each stage of a medical procedure as if it were a separate treatment; also called as Unbundling

    – Service provider may bill for expensive medical services than the one actually performed;

    – Just to generate insurance payments, service providers may perform unnecessary medical services;

    – Just to obtain insurance the service providers may misrepresent non-covered treatments as medically necessary covered treatments;

    – To validate the medical procedures that are not actually needed, service providers may falsify patients' diagnosis and/or treatment histories;

- Fraud by Insurance subscribers:

    – For obtaining a lower premium rate, records of employment / eligibility can be falsified;

    – Subscribers may file claims for medical services which are not actually received;

    – To illegally claim the insurance benefits, subscribers may use other persons' coverage or insurance card.

- Frauds by Insurance carriers:

    – Fake reimbursements;

    – Misrepresenting benefit / service statements.

- Conspiracy frauds:

    – In such frauds more than one party are involved; for example, fraudulent activity may include a patient and a doctor or insurance company.

B. Data for Healthcare Fraud

The raw data for healthcare fraud detection are generally insurance claims which comes from many different sources.

The other kinds of data used in the healthcare fraud detection, other than the insurance claims data, are data of physicians, data of prescriptions given by the physicians, data of the medication or drugs prescribed and data of bills and transactions [6]. Each country has unique characteristics for its healthcare system data. Thus, the work done in fraud detection is evaluated by taking into consideration the knowledge of the governmental health data. U.S. Health Care Financing Administration (HCFA) is a major governmental health department. Medicare and Medicaid are two health care programs in the U.S. Mostly researchers, to detect frauds and abuse in the healthcare systems, uses Medicare or Medicaid data which involves data of medication and drugs, bills and transactions and medical providers. [6], [7], [8], [9], [10], [11], [12], [13], [14], [19].

C. Methods for Healthcare Fraud Detection

There are various intricate and complicated patterns with many small trivial specifications comprised in Fraud, whose data is gathered over a prolonged time span. It is extremely difficult to detect these patterns in the present times, where we have a huge accumulation of the data and very few means for evaluating them[15]. Traditionally, a few auditors used to handle thousands of health care claims. Thus, usually only experienced investigators would manage fraud detection. But due to large collection of data this method becomes time-consuming and inefficient. Improvements in data mining and machine learning tools bring attention to automated systems for fraud detection [16]. For detection of anomaly and detection of fraud behavioral profiling methods based on machine learning techniques are used and for this purpose, behavior pattern of each person, involved in the healthcare system, is configured to observe and check it for any deduction from the standards [17]. Data mining methods are classified into two categories as supervised and unsupervised learning by most of the researchers [18], [16]. But, in some cases, along with these two approaches, semi-supervised learning is also involved in the classification [6], [1].

LITERATURE REVIEW

The Centers for Medicare and Medicaid Services (CMS) releases healthcare data which is used by most of the researchers for healthcare fraud detection.

Srinivasan et al. [19] proposed an anomaly detection method by applying Rule-based Data Mining, an unsupervised technique, on the insurance claims data acquired from Medicare data. Applications for analyzing health insurance claims leverage big data to detect fraud, abuse, waste, and errors were devised. Medical insurance claim anomalies were detected using these applications that avail private health insurers identify hidden cost overruns that transaction processing systems can't detect.

Branting et al. used Healthcare data sourced from Medicare and Medicaid and applied supervised techniques along with graph analytics and decision tree [9]. They proposed an approach to estimating healthcare fraud risk that applies network algorithms to graphs derived from open source datasets.

A research, which uses the CMS 2012 data, determined a way a physician practices by analyzing the physicians past schooling [20]. By presenting a geographical analysis with the national distribution of school procedure payments and charges, they compared medical school charges, procedures, and payments as well as find possible anomalies in the data. The authors attempt to identify the physicians who are misusing or inefficiently using medical insurance systems by finding correlations between educational backgrounds and the practices and procedures physicians perform.

Ko et al. specifically considered only one field, Urology, while using 2012 CMS data [21]. The authors attempt to determine an estimated savings from a standardized service utilization by analyzing variability among Urologists within the field's service utilization and payment.

A study, which uses 2013 CMS dataset, built a machine learning model to detect when physicians exhibit anomalous behavior in their medical insurance claims [8]. It attempts to determine if, and when, physicians are acting outside the norm of their respective specialty, which could indicate misuse, fraud, or lack of knowledge around billing procedures.

The model is evaluated by calculating precision, recall, and Fscore with 5-fold cross-validation. It uses the multinomial Naïve Bayes algorithm. The model predicts several classes of physicians with an F-score over 0.90 and these results show that it is possible to effectively use machine learning in a novel way to classify physicians into their respective fields solely utilizing the procedures they bill for. It identifies physicians who are potentially misusing healthcare insurance systems for further investigation.

CONCLUSION

In this paper, healthcare fraud, types of healthcare frauds, types and sources of healthcare data, and methods for healthcare frauds were studied. Various studies are reviewed in the literature. It is deduced that in the healthcare industry, 'Data' is a paramount issue. The major part of the data comes from governmental resources and private insurance companies. Mainly, machine learning and data mining are used for Healthcare fraud detection. Supervised, unsupervised and semi-supervised learning are the three categories of Machine learning approaches. In most of the cases, semi-supervised learning approaches are used by many researchers. But, to detect frauds in healthcare system more efficiently, new semi-supervised learning approaches can be proposed in few cases. But, to conceal all the instances of the healthcare fraud, there doesn't exist any particular standard approach or patterns. It can be concluded from this review that the advanced machine learning techniques and newly acquired sources of the healthcare data would be forthcoming subjects of interest in order to make the healthcare affordable, to improve the effectiveness of healthcare fraud detection and to bestow top quality on healthcare systems.

FUTURE SCOPE

After reviewing different studies on healthcare fraud detection, it can be concluded that frauds or abuse that occur in health insurance systems can be of different unusual patterns. To detect such suspicious patterns more research work is needed by using advanced data mining and machine learning techniques. There is a need to propose new methods while considering minute details of healthcare data. To achieve this, correlations between different entities of healthcare data can be taken into account.

REFERENCES

[1] Abdallah, A., Maarof, M. A., & Zainal, A. (2016). Fraud detection system: A survey. *Journal of Network and Computer Applications*, *68*, 90-113.

[2] Behdad, Mohammad, et al. "Nature-inspired techniques in the context of fraud detection." *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 42.6 (2012): 1273-1290.

[3] Konasani, Venkatareddy, Mukul Biswas, and Praveen Krishnan Koleth. "Healthcare fraud management using big data analytics." *An Unpublished Report by Trendwise Analytics, Bangalore, India* (2012).

[4] National Health Care Anti-Fraud Association. "Health Care Fraud–A Serious and Costly Reality For All Americans." *April2005* (2007).

[5] Yang, Wan-Shiou. "A Process Pattern Mining Framework for the Detection of Health Care Fraud and Abuse." *National Sun Yat-Sen University, Taiwan* (2003)..

[6] Liu, Qi, and Miklos Vasarhelyi. "Healthcare fraud detection: A survey and a clustering model incorporating Geo-location information." In *29th World Continuous Auditing and Reporting Symposium (29WCARS), Brisbane, Australia*. 2013..

[7] Thornton, Dallas, Roland M. Mueller, Paulus Schoutsen, and Jos van Hillegersberg. "Predicting healthcare fraud in medicaid: a multidimensional data model and analysis techniques for fraud detection." *Procedia technology* 9 (2013): 1252-1264.

[8] Bauder, Richard A., Taghi M. Khoshgoftaar, Aaron Richter, and Matthew Herland. "Predicting medical provider specialties to detect anomalous insurance claims." In *Tools with Artificial Intelligence (ICTAI), 2016 IEEE 28th International Conference on*, pp. 784-790. IEEE, 2016.

[9] Branting, L. Karl, Flo Reeder, Jeffrey Gold, and Timothy Champney. "Graph analytics for healthcare fraud risk estimation." In *Advances in Social Networks Analysis and Mining (ASONAM), 2016 IEEE/ACM International Conference on*, pp. 845-851. IEEE, 2016.

[10]   Musal, Rasim Muzaffer. "Two models to investigate Medicare fraud within unsupervised databases." *Expert Systems with Applications* 37, no. 12 (2010): 8628-8633.

[11]   Copeland, Leanndra, Dana Edberg, Anna K. Panorska, and Jeanne Wendel. "Applying business intelligence concepts to Medicaid claim fraud detection." *Journal of Information Systems Applied Research* 5, no. 1 (2012): 51.

[12]   Bauder, Richard A., and Taghi M. Khoshgoftaar. "A probabilistic programming approach for outlier detection in healthcare claims." In *Machine Learning and Applications (ICMLA), 2016 15th IEEE International Conference on*, pp. 347-354. IEEE, 2016.

[13]   Bauder, Richard A., and Taghi M. Khoshgoftaar. "A novel method for fraudulent medicare claims detection from expected payment deviations (application paper)." In *Information Reuse and Integration (IRI), 2016 IEEE 17th International Conference on*, pp. 11-19. IEEE, 2016.

[14]   van Capelleveen, Guido, Mannes Poel, Roland M. Mueller, Dallas Thornton, and Jos van Hillegersberg. "Outlier detection in healthcare fraud: A case study in the Medicaid dental domain." *International journal of accounting information systems* 21 (2016): 18-31.

[15]   Rudman, William J., John S. Eberhardt, William Pierce, and Susan Hart-Hester. "Healthcare fraud and abuse." *Perspectives in Health Information Management/AHIMA, American Health Information Management Association* 6, no. Fall (2009).

[16]   Joudaki, Hossein, Arash Rashidian, Behrouz Minaei-Bidgoli, Mahmood Mahmoodi, Bijan Geraili, Mahdi Nasiri, and Mohammad Arab. "Using data mining to detect health care fraud and abuse: a review of literature." *Global journal of health science* 7, no. 1 (2015): 194.

[17]   Jyothsna, V., VV Rama Prasad, and K. Munivara Prasad. "A review of anomaly based intrusion detection systems." *International Journal of Computer Applications* 28, no. 7 (2011): 26-35.

[18]   Li, Jing, Kuei-Ying Huang, Jionghua Jin, and Jianjun Shi. "A survey on statistical methods for health care fraud detection." *Health care management science* 11, no. 3 (2008): 275-287.

[19]   Srinivasan, Uma, and Bavani Arunasalam. "Leveraging big data analytics to reduce healthcare costs." *IT professional* 15, no. 6 (2013): 21-28.

[20]   Feldman, Keith, and Nitesh V. Chawla. "Does medical school training relate to practice? Evidence from big data." *Big data* 3, no. 2 (2015): 103-113.

[21]   Ko, Joan S., Heather Chalfin, Bruce J. Trock, Zhaoyong Feng, Elizabeth Humphreys, Sung-Woo Park, H. Ballentine Carter, Kevin D. Frick, and Misop Han. "Variability in Medicare utilization and payment among urologists." *Urology* 85, no. 5 (2015): 1045-1051.