

Research on Anomaly Analysis and Judgment Method of Healthcare Based on Data-drive

Linlin Su

The University of Hong Kong, Hong Kong, China
leensu00@gmail.com

Abstract: With the rapid development of social healthcare, the number of insured individuals is gradually increasing, and information technology is gradually developing. Some unreasonable medical reporting behaviors that deceive large amounts of medical insurance funds from national medical insurance institutions, hospitals, and other medical industry institutions by fabricating physical conditions and concealing real situations are seriously endangering the construction of the social health development system. At the same time, the scale of the Cyber Physical System (CPS) for medical information is constantly increasing, and the inconsistency and inefficient utilization of existing medical information standards have led to a rapid increase in the overall complexity of the system. Ordinary real-time processing systems are unable to meet the computational needs of geometrically increasing data volumes. After years of medical informatization and the application of various medical information systems, it is difficult to achieve data sharing, mapping, and fusion based on medical insurance analysis, which includes a large amount of medical visit records and electronic data covering medical treatment time, treatment costs, and other contents. This study explores the new technological innovations sought by the medical field and proposes to conduct research and analysis on healthcare service datasets in the context of big data, focusing on the relevant methods of medical insurance anomaly analysis and big data mining techniques. Methods such as medical information standardization, relationship extraction, machine learning, and big data analysis are introduced to study the structuring of medical insurance data, big data assisted medical insurance case coding, medical insurance abnormal behavior mining, and active medical insurance judgment methods. It has important practical significance and value in promoting the standardization and structuring of medical insurance data, improving the efficiency of medical insurance anomaly analysis and detection, mining medical insurance quality, and assisting medical insurance research and judgment.

Keywords: Medical Insurance, Data-drive, Anomaly Analysis.

1. Research Background

In recent years, the issue of abnormal medical insurance costs has been of great concern. Since 2007, federal law enforcement agencies in the United States have increased their efforts to combat healthcare fraud. So far, nearly 2300 suspects have been prosecuted for fraudulent billing, with a total amount involved exceeding \$7 billion. In addition, a large public healthcare insurance company in Germany bribed doctors to exaggerate patients' conditions, such as describing mild hypertension as severe hypertension in order to defraud the German National Health Fund of more funding. In 2016, it was ultimately fined 7 million euros by the German Federal Insurance Agency. The deviant behavior of medical insurance has to some extent caused an unreasonable increase of medical expenses, seriously harming the legitimate interests of insured individuals, and bringing huge financial losses to the medical industry. Although various countries have saved some financial costs and recovered certain economic losses for medical insurance by carrying out antifraud work in medical insurance, various insurance fraud phenomena are also becoming increasingly serious, and such illegal methods have provoked complexity and diversity for the development of the medical insurance industry [1-3].

The rapid development of data and communication technology has led to a geometric multiple growth of data in daily life, and big data technology has become a current research interest. The medical system is closely related to people's life safety, and the aging society leads to increasingly prominent issues such as healthcare and elderly care. The

question of how to alleviate the resulting medical burden is becoming increasingly urgent. Due to the close relationship between the treatment and research processes of hospitals, clinics, medical stations, and medical imaging, electronic medical records, and other data, traditional relational databases cannot process and analyze such a large amount of data in a timely manner. Therefore, big data technology can effectively reflect its value in the medical field [4]. Regarding the popular cloud computing and big data technologies, Zhang et al. discussed combining this with healthcare service applications and proposed a data management layer framework for distributed storage and parallel computing but did not provide detailed technical details [5]. Alhumud et al. hope to achieve interoperability of health data in medical CPS systems through cloud platforms and propose a conceptual framework that supports medical professionals to analyze, share, and make decisions on heterogeneous data [6]. Hossain et al. studied a monitoring system that combines voice and EEG signals, which is deployed on cloud CPS and can support patient healthcare [7].

In terms of the current level of informatization in the medical industry, medical data exhibits four basic characteristics of big data, namely, large data volume, multiple types of data, high inherent value, and fast generation [8]. After years of medical informatization and the application of various medical information systems, hospitals and medical insurance management departments have accumulated a large amount of medical treatment records and electronic data, covering such content as treatment time and treatment costs. However, the lack of uniformity and

inefficient utilization of existing medical information standards makes it difficult to promote data sharing, data mapping, and data fusion based on medical insurance analysis [9]. Meanwhile, traditional medical insurance anomaly analysis often relies on medical practitioners to manually formulate rules and conduct simple statistical analysis based on experience, making it difficult to accurately extract complete abnormal behavior information from the vast and complex medical insurance data [10]. Therefore, effectively conducting research and analysis on medical insurance datasets in the context of big data, mining hidden abnormal behaviors, is of great significance for assisting medical insurance research and judgment, controlling unreasonable cost growth in China's medical industry, and promoting the healthy development of the social medical insurance industry [11-14].

2. Research Idea and Purpose

2.1. Data Driven Coding Strategy for Medical Insurance Medical Records

To address the problem of structured medical insurance data and big data assisting in medical insurance case coding, we aim to improve the efficiency and accuracy of medical insurance case coding conversion and upgrading, ensure the accuracy of medical insurance reimbursement, and address key issues during the implementation of ICD-11. Propose the establishment of an ICD-11 implementation evaluation table, design from the perspective of big data analysis technology, build a cross regional ICD-11 information service architecture, implement medical insurance medical record coding conversion and mapping based on the MapReduce model, and automatically encode medical insurance medical records for big data analysis.

2.2. Effective Identification Strategy for Medical Insurance Aggregation Behavior Based on Distributed Data-Driven Approach

Based on new medical insurance medical record coding system, an improved distributed medical insurance aggregation behavior mining method is proposed, which includes establishing a structured model of medical behavior matrix and medical aggregation behavior matrix, using MapReduce transpose, establishing MapReduce frequent item sets, and MapReduce links. This method continuously generates high-order matrices through cross operations and analyzes and verifies real outpatient reimbursement data. The purpose is to improve the efficiency and accuracy of high-dimensional frequent itemset data detection and achieve effective identification of abnormal medical clustering behavior.

2.3. Patient Analysis and Fraudulent Patient Labeling Strategy Based on Patient Information Weighted Tree Structure

Based on the research on abnormal behavior in medical insurance aggregation, this study focuses on similar medical behavior information of different types of patients and therefore proposes a patient analysis and fraudulent patient labeling method, that is based on patient information weighted tree structure. By providing effective fraudulent labels to reduce the number of false positives, it overcomes

the problem of patient information homogeneity faced by medical fraud detection and improves the efficiency of fraud detection.

2.4. Research Strategy for Active Research and Judgment of Medical Insurance Based on Intelligent Big Data

Based on the existing big data theory, a new big data analysis model is proposed, which includes intelligent big data analysis, construction of a medical and health big data analysis platform, and a big data-driven research framework for medical insurance abnormal behavior analysis, active judgment, and innovative management of medical care. It involves the exploration of the transformation of existing static big data into intelligent and proactive big data models at the management decision-making level.

3. Research Methodology

The research methodologies focus on the following three key orientations, which are usually not separated due to the logical connection between them. The specific research ideas are as follows.

3.1. A Hadoop Based Implementation Method for Medical Insurance Medical Record Coding

Firstly, establish an ICD-11 implementation evaluation form, and propose specific steps for different stages of ICD-11 implementation through three stages: pre-implementation, during implementation, and post-implementation. Based on the three key stages of ensuring the reporting of medical record data, the migration and mapping of old coding standards, and the implementation of ICD-11 automatic coding, this paper proposes a cross regional ICD-11 information service architecture, a MapReduce model based on medical insurance case coding conversion and mapping method, and a medical insurance case automatic coding method for big data analysis from the perspective of big data analysis technology.

3.2. A Method for Mining Abnormal Behavior in Medical Insurance Aggregation Based on Frequent Patterns

Relying on the new medical insurance medical record coding system, this paper analyzes the challenges faced by the detection of medical aggregation abnormal behavior. In response to the complexity of medical aggregation abnormal behavior, a structured model and medical behavior matrix of medical aggregation behavior are constructed, and association rule mining methods are introduced to transform medical aggregation behavior mining into a frequent pattern mining problem. Using big data analysis technology, combined with the MapReduce distributed computing model and the actual scenario of medical insurance reimbursement, an improved distributed based medical aggregation behavior mining method is proposed to overcome the shortcomings of association rule mining methods in the medical field.

3.3. A Medical Insurance Fraud Labeling Method Based on Patient Information Weighted Tree

On the basis of the principle of multi tree, a patient

information weighted multi tree is constructed for similar medical claim information of different patient types. The patient information column value is regarded as a node, and each complete path between the root node and leaf node is regarded as a detailed information set for the corresponding patient type. The weight and threshold of the edge are calculated based on the relative frequency of the information set. Mapping the patient's medical insurance reimbursement information to a list of patient information containing genuine fraudulent information is done to detect fraudulent claims. Compared with traditional mapping methods, mapping based on patient information weighted trees reduces the number of false positives and improves the efficiency of fraud detection by providing effective fraud markers.

3.4. A Proactive Medical Insurance Assessment Method Based on Intelligent Big Data

Analyze the limitations and complex issues faced by existing big data in the field of healthcare and propose an intelligent big data conceptual model. This model effectively utilizes the relationships between big data sets and establishes an intelligent big data-driven healthcare active judgment framework by integrating key technologies and methods such as healthcare big data structured processing, dataset semantic analysis, machine learning method set mapping, and exploratory visualization. Supported by big data components such as Hadoop, Hbase, Spark, etc., it is suitable for practical medical insurance research and judgment support scenarios and is set provide reference for the development of medical informatization in the era of medical and health big data.

References

- [1] Lin, Y. (2022). Analysis of the current situation of abnormal research on medical insurance at home and abroad. *Insurance Research*, (12), 115–122.
- [2] Xia, H., Wang, K., & Zhang, S. C. (2019). Abnormal and anti abnormal issues in medical insurance. *Modern Preventive Medicine*, 34(20), 3907–3908.
- [3] Tang, J. S. (2018). Cracking down on medical insurance fraud and maintaining fund security. *China Medical Insurance*, (5), 5–7.
- [4] Tang, M. W., Su, X. N., & Xiao, L. J. (2018). An intelligence analysis framework for big data. *Journal of Intelligence*, 37(5), 21–30.
- [5] Zhang, Y., Qiu, M., Tsai, C. W., et al. (2021). Health-CPS: Healthcare cyber-physical system assisted by cloud and big data. *IEEE Systems Journal*, 11(1), 88–95. <https://doi.org/10.1109/JSYST.2017.2786411>
- [6] Alhumud, M. A., Hossain, M. A., & Masud, M. (2016). Perspective of health data interoperability on cloud-based medical cyber-physical systems. In 2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW) (pp. 1–6). IEEE. <https://doi.org/10.1109/ICMEW.2016.7416234>
- [7] Hossain, M. S. (2015). Cloud-supported cyber-physical localization framework for patients monitoring. *IEEE Systems Journal*, 11(1), 118–127. <https://doi.org/10.1109/JSYST.2015.2417596>
- [8] Zhang, Z., Zhou, Y., Du, S. H., Luo, X. Q., & Mei, T. (2018). Medical big data and its opportunities and challenges. *Journal of Medical Informatics*, (6), 2–8.
- [9] (2020). Tackling user requirements and challenges in robotics. In *Recent advances in intelligent engineering* (pp. 253–277). Springer, Cham.
- [10] Inoubli, W., Aridhi, S., Mezni, H., et al. (2018). An experimental survey on big data frameworks. *Future Generation Computer Systems*, 86, 546–564. <https://doi.org/10.1016/j.future.2018.04.039>
- [11] Chen, H. J., & Han, Y. J. (2018). Application of data mining and information fusion in fraud identification in the insurance industry. *Computer and Modernization*, (9), 110–112.
- [12] Gao, Y. T. (2015). Research on fraud identification of new rural cooperative medical insurance based on outlier detection (Master's thesis). Harbin University of Commerce, Heilongjiang.
- [13] Cao, H. F. (2020). Research on medical insurance anomaly detection based on SOFM neural network and random forest (Master's thesis). Beijing Jiaotong University, Beijing.
- [14] Huang, H. S. (2020). Research on medical insurance anomaly detection based on improved fuzzy clustering algorithm (Master's thesis). Beijing Jiaotong University, Beijing.