Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 2, No. 3, 2022

Trust Evaluation of Doctor Behavior Based on GCN Network in Medical Big-Data Access Control

Xun Tian

Yunnan University of Finance and Economics, School of Information, Kunming 650221g, China
gk700.gjgcs@foxmail.com

Abstract: The establishment of health care big data has brought great convenience to population health and medical research, but at the same time a series of privacy protection issues must be considered as a result. In this paper, we propose a graphical convolutional neural network to detect the access behavior of doctors in medical big data. In this paper, we propose a graphical convolutional neural network to model the access behavior of doctors in medical big data and perform trust evaluation, so as restrict such doctors or behaviors. In this paper, by taking the doctor behavior features and the doctor-doctor relationship network as input, the GCN network is used to supervise the learning of the department to which the doctor belongs, and the last layer is used as the characterization learning result. Finally, the similarity between doctor and department is used as the doctor behavior trust evaluation index. The experimental results show that the proposed model in this paper can well identify doctors' behaviors with malicious intent and assign a low trust value, laying the foundation for further research.

Keywords: Trust evaluation; Medical big-data; GCN.

1. Introduction

With the rapid development of Internet information technology, all walks of life have stepped into the era of big data. Especially in the medical field, the era of big data has created an unprecedented and great convergence of digital with health and medical care. This convergence has enabled new medical research and new health services to manifest great value and potential. In the healthcare big data environment, based on data standards and data integration, doctors can call up patients' historical information and even genetic information during the treatment of patients. In this way, doctors can provide better and more targeted specific treatment plans to specific patients at specific points in time. And big data in healthcare can bring many benefits to medical institutions, government management departments and related enterprises. However, each stage of collecting, mining, analyzing and utilizing big data in healthcare requires a good health big data industry ecosystem. One of the basic and important features of this industrial ecosystem is "openness". However, the "openness" of health care big data will increase the risk of data leakage to a certain extent. For example, in September 2017, the service information system of a hospital in China was hacked, resulting in the leakage and trafficking of a large amount of citizens' information; in January 2018, hundreds of thousands of information about newborn babies and pregnant women due to unauthorized access by staff of a community health service center in China was leaked. Therefore, preventing the leakage of healthcare big data has become an urgent issue now. Currently, the main access control model uses a static authorization approach. It uses fixed policies and does not consider uncertainty and big data environments. It is difficult to apply in scenarios where authorization changes frequently in big data environments. The main direction of scholars in recent years is dynamic access control methods. Among them, risk-based and trust-based are two more mainstream approaches. And this paper finds the doctors with malicious access behavior by modeling the trust value of doctors' access behavior.

2. Related works

In the context of big data in health care, medical information construction is bound to break the traditional data silos and move toward sharing and openness. However, in the process of data "flow", there are many hidden problems. Among them, the privacy leakage of data has become the core concern of big data in health care. At present, many scholars have conducted research on the privacy leakage of data and proposed many solutions in terms of technology. Examples include pseudonymization of patient identity, encryption of patient data and information, creation of public and private clouds to handle sensitive data, privacy-protected data distribution, privacy-centric access control and data outsourcing, and dynamic reconstruction of data [1-4]. In the healthcare domain, Soceanu propose an advanced encryption scheme and attribute-based access control authorization framework for clinical data privacy and security issues. Allowing partial visibility and effective protection of authorized parts. Suneetha . proposed a new framework spark using K-anonymization and L-diversity to mask patient sensitive information. It ensures that the shared data does not reveal the original medical data. And isolate sensitive data before transferring to the system. Literature [7,8] used blockchain technology for effective management of big data set access control to solve the data privacy security problem. Lee designed a medical big data privacy protection system based on Diffie-Hellman protocol to protect patients' private information and avoid the leakage of medical data. Wu built a medical big data privacy protection platform based on IoT . For medical big data transmission sharing security. Medical data security is provided. Y Yang proposed a privacy-preserving electronic health record (EHR). and developed a policy update mechanism based on keyword matching. To achieve flexible access updates without compromising privacy. Mariagrazia proposed an ABAC model-based subject and object access control principle based on risk scenarios. It achieves dynamic modification of authorized adaptive access

control based on risk values. L Chen proposed a risk-aware RBAC model. The model consists of three components: user's trust, user's ability to assume roles, and appropriateness between roles and permissions. Riaz added trust and risk relative to traditional access control considering different sensitivity requirements of different applications. Two dynamic risk-based access control system decision methods are proposed.

To address the above problems, this paper transforms doctors' high-dimensional access features into low-dimensional section relevance features based on their historical access records using graph convolutional neural networks, and then uses the Euclidean distance similarity between each section doctor and the section center feature as the reputation evaluation index based on the reduced-dimensional features. Therefore, the research in this paper provides a new idea for the access control research of healthcare big data.

3. Trust evaluation method

In this paper, we focus on supervised learning by using the graph of relationships between doctor access object features and doctors as the input to the GCN network.

3.1. Doctor Characteristics and Relationship Chart

For an electronic case Ec, let there be m kinds of data related to diagnosis and treatment in the case, such as blood and urine laboratory data, ECG, chest X-ray, clinical symptoms, etc.. Use the one-hot code to indicate whether there is a record of the corresponding data in the case Ec table, noted as $Ec = [f_1, f_2, \dots, f_m]$, where $f_i = 1$ means that item i has a record.

A doctor's access record Dr should contain the case records of all non-primary patients accessed within a certain work duration, denoted as $Rd = [Dr_1, Dr_2, \dots, Dr_n]$. Since any Rp_i is a vector of fixed dimensions, Rd can also be represented as a two-dimensional matrix as follows:

$$Dr = \begin{bmatrix} f_{11} & \cdots & f_{1m} \\ \vdots & \ddots & \vdots \\ f_{n1} & \cdots & f_{nm} \end{bmatrix}$$

The data features required for a doctor access, Rd, can be characterized based on all historical data features from the doctor's access to the non-primary patient's medical record as follows:

$$Dr = \left[\sum_i^n f_{i1}, \sum_i^n f_{i2}, \dots, \sum_i^n f_{im} \right]$$

A doctor access to a patient case creates a relationship between the doctor and the case, and the element level of the weighted adjacency matrix A_{d2c} of the doctor-case relationship graph is represented as:

$$A_{d2c}(i,j) = k\ (Doctor\ i\ access\ medical\ record\ j\ k\ times)$$

Doctors who access the same case in the diagnostic process by correlation, and therefore can be transformed from a doctor-case relationship graph to a doctor-doctor relationship graph, whose weighted adjacency matrix A_{d2d} can be calculated as following:

$$A_{d2d} = \sqrt{A_{d-c} \cdot A_{d-c}^T}$$

The Laplace matrix is the central object used to study the structural properties of the graph. The Laplace matrix is defined as follows: $L = D - A$, D is the diagonal matrix of degrees of A. The Laplacian matrix also has a regularized representation:

$$L_{sym} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$$

In this paper we use a two-layer GCN network with fixed graph filters:

$$\tilde{L}_{sym} = L_{sym} + I$$
$$H = \sigma\left(\tilde{L}_{sym} X W_1 + B_1\right)$$
$$Y = softmax\left(\tilde{L}_{sym} H W_2 + B_2\right)$$

Similarly, the central feature points of all doctors in each department are found by deep learning, and the cosine similarity between each doctor and the central feature point is calculated separately as the trust evaluation index:

$$Trust_i = sim(y_i, \hat{y}) = \frac{y_i \hat{y}}{\|y_i\| \|\hat{y}\|}$$

3.2. Simulation experiment

This experiment was conducted by computer simulation to generate 10 departments with different medical records of 50 doctors in each department for access, totaling 200,000 access of data. The sequence of 40 doctors with normal access and those with malicious access in each department.

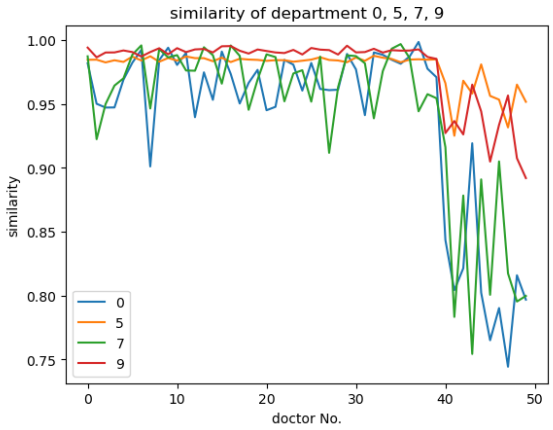

Fig 1. Similarity of the between doctors and the center of the department

Through simulation experiments we get the results in Fig 1. According to Figure 1 we can easily find the low similarity between malicious doctors and section averages. It shows that our method effectively distinguishes doctors with malicious accessing behavior.

4. Conclusion

In this paper, we propose a trust evaluation model for medical environment, which uses graph neural network algorithm to reduce the high-dimensional access data of all doctors into low-dimensional section relevance data based on the historical access records of doctors, and then learns the average features of the section based on the reduced features of doctors in the same section, and cosine similarity to evaluate the trust value of doctors within the section. The trust model proposed in this paper is different from the traditional trust model, this model can not only be based on whether the data of doctors' access to patient's medical records are related to diagnosis and treatment, but also based on the network topology relationship between doctors at the same time, this model is more conducive to give credibility judgment on doctors' lines, which brings great reference value for the

subsequent access control research. The model is proved to be effective in the simulation experiments at the end of the paper.

References

[1] AGGARWAL C C, PHILIP S Y . A general survey of privacy-preserving data mining models and algorithms [M]. Privacy-preserving data mining. Springer. 2008: 11-52.

[2] FERNáNDEZ-ALEMáN J L, SEñOR I C, LOZOYA P Á O, et al. Security and privacy in electronic health records: A systematic literature review [J]. Journal of biomedical informatics, 2013, 46(3): 541-62.

[3] ABBAS A, KHAN S U. A review on the state-of-the-art privacy-preserving approaches in the e-health clouds [J]. IEEE Journal of Biomedical and Health Informatics, 2014, 18(4): 1431-41.

[4] CAMARA C, PERIS-LOPEZ P , TAPIADOR J E. Security and privacy issues in implantable medical devices: A comprehensive survey [J]. Journal of biomedical informatics, 2015, 55(272-89.

[5] SOCEANU A, VASYLENKO M, EGNER A, et al. Managing the privacy and security of ehealth data; proceedings of the 2015 20th International Conference on Control Systems and Computer Science, F, 2015 [C]. IEEE.

[6] SUNEETHA V, SURESH S, JHANANIE V. A Novel Framework using Apache Spark for Privacy Preservation of Healthcare Big Data; proceedings of the 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), F, 2020 [C]. IEEE.

[7] CHEN X, ZHU H, GENG D, et al. Merging RFID and Blockchain Technologies to Accelerate Big Data Medical Research Based on Physiological Signals [J]. Journal of Healthcare Engineering, 2020.

[8] UCHIBEKE U U, SCHNEIDER K A, KASSANI S H, et al. Blockchain access control Ecosystem for Big Data security; proceedings of the 2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), F, 2018 [C]. IEEE.

[9] LEE N-Y, WU B-H. Privacy Protection Technology and Access Control Mechanism for Medical Big Data; proceedings of the 2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI), F, 2017 [C]. IEEE.

[10] WU X, ZHANG Y , WANG A, et al. MNSSp3: Medical big data privacy protection platform based on Internet of things [J]. NEURAL COMPUTING & APPLICATIONS, 2020,

[11] YANG Y, ZHENG X, GUO W , et al. Privacy-preserving fusion of IoT and big data for e-health [J]. Future Generation Computer Systems, 2018, 86(1437-55.

[12] TEIMOURIKIA M F A M. Access Control Privileges Management for Risk Areas [J]. Information Systems for Crisis Response and Management in Mediterranean Countries, 2014,

[13] CHEN L, CRAMPTON J. Risk-aware role-based access control; proceedings of the International Workshop on Security and Trust Management, F, 2011 [C]. Springer.

[14] SHAIKH R A, ADI K, LOGRIPPO L. Dynamic risk-based decision methods for access control systems [J]. computers & security, 2012, 31(4): 447-64.