Analysis and Evaluation on the Sharing Security of Medical Big Data based on Analytic Hierarchy Process

Meijing Shan

Department of Information Science and Technology, East China University of Political Science and Law, Shanghai , China

shanmeijing@ecupl.edu.cn

Keywords: Medical big data, Analytic hierarchy process, Secure sharing

Abstract: Opening and sharing of medical big data can provide patients, medical technicians and medical researchers the required information, as well as new medical and medical equipment research and development staff. In this study, analytic hierarchy process is used to evaluate the security of medical large data sharing. Medical big data involves the patient's privacy and other sensitive information, so that its security in information sharing constraints more than general large data. According to method Delphi, the medical big data security sharing indexes are selected. Focusing on these indexes, the hierarchy evaluation model of index is established by using AHP method, and the weight of each index and the influence to each other are determined. Experiments showed that the system can evaluate the sharing security of medical big data effectively. This study proposed a evaluation method based on the sharing of medical large data security, which provides a new technology platform and solution for the information management of the hospital.

1. Introduction

During the 13th Five-year Period, the smart city ideas which based on the new generation of information technology bring a new urban life expectation for the goverments and people. In Smart city construction, there are three big data application key areas: transportation, health and peace. China's medical and health services, health statistical survey and various medical and health industry information system has huge amount of data. For example, each CT image contains about 150 MB data, each genome sequence file is about 750 MB, each pathology plan is close to 5 GB, times the population and the average life expectancy in China. Then each community hospital or medium-sized pharmaceutical enterprises will generate and accumulate up to TB or even PB of structured and unstructured data.

The big data resources in medical and health include medical services data of EHRs, the cost and charge data of hospital and health care, medical research academic, social, government data, manufacturers of medicine, medical , clinical trial data, the behavior of the residents and the health management data and the population of the government and public health data, along with our country public social and economic life of network data, construct the initial big data resources in the field of health care[1-5].

2. Necessarity of Medical Big Data Secure Sharing

With the coming of the big data era, medical big data become a basic support of Internet medical model. Medical big data are from closed to open and sharing, the data sharing has become a worldwide trend. Through local medical big data sharing to gobal medical big data open, it provides important way to obtain the required data on medical research, promotes the reuse of medical big dat, promotes the innovation of the medical and health services industry. At the same time, medical professors can have a better understanding of the residents' health status and medical resources usage.

Medical big data are the subjective and objective data and related statistical data during the process of clinical diagnosis and treatment on patients. Subjective data reflect the understanding of

Copyright © (2018) Francis Academic Press, UK DOI: 10.25236/iccse.18.022

disease in patients with medical workers and formulate and adjust the whole process of treatment, including thinking of case discussion records, cases discussion record, the superior doctors rounds, consultation opinion, course records, etc. Objective data are in the process of examination and nursing care of patients with objective record, including outpatient medical records, the hospital volunteers, temperature is single, the doctor's advice, assay (test reports), medical imaging data, surgery and anesthesia, pathological data and record nursing records, etc. Medical statistics is for the convenience of hospital management personnel to understand the status of the hospital operation of data, such as average, adverse incident report of such confinement, empty beds, this number, monthly dose, and working registers.

Big data for new drug research and development has brought the dawn. Domestic and foreign pharmaceutical companies are now faced with the huge development predicament, and these difficulties such as clinical trials data problems, can be solved with big data.

Initially formed a group platform of medical big data sharing and publishing

The medical data contain the privacy of patients and hospitals, and can not all shared directly. However, in order to promote the medical applications of big data and the development of medical scientific research and innovation, after technical treatment the medical big data can be public or open. For example, Partners HealthCare Group (PHG) in US integrated health service system can provide part of the personal health record for patients, like patients visit doctors' diagnose notes online. The objective data and statistical data of medicine after technical treatment can be open to the public, medical subjective data after processing can realize open step by step.

Sharing level is not high, the security situation is not optimistic

Medical big data sharing has done more than ten years, but now almost all the hospitals is completely silent in their medical big data. The utilization of medical big data in China is still too low. Big data are helpful for the study characteristics, the health of new diseases in different populations and control global infectious diseases. Big data for safety evaluation of medical devices has brought the Gospel, for medical devices to evaluate its security, effectiveness and evaluation. At present, the medical community usually can only evaluate the effectiveness of medical devices. Due to participating in a clinical trial of a limited number of patients, so it's difficult to assess the safety of the low incidence of risk. Experts say that the emergence of large data made it possible to safety evaluation. Through the huge amounts of data, you can see all of the patients used this product.

The quantity of medical big data is growing so fast. Especially for medical imaging data, it is very fast architecture. Medical data resources, participate in the main body is more complex. In China information sharing is the biggest problem in current medical treatment.

3. Restriction factors of medical big data secure sharing

Medical big data open and sharing need the following steps: data collection, data processing, licensing, and data release and maintain. The restriction of big data security share index mainly include resource layer A1, processing layer A2, visualizaion layer A3 and regulatory layers A4.

(1)Resource layer A1 is mainly responsible for collection and management of medical big data, for secondary utilization of hospital information and base external Internet data basis. In this layer, A1 contains structured data (documents) B1, unstructured data B2, application service resources B3. These medical big data mainly comes from HIS database, EMR database, LIS (such as video surgery), database, multimedia database HRP database, etc.

(2)Processing layer A2 is the core of the medical big data sharing system. It is mainly responsible for data processing and processing of big data review. Through processing of the raw big data, it forms a new medical big data set which do not disclose personal privacy and sensitive information, but maintain distribution characteristics of the big data. At the same time it guarantee availability and safety of medical big data, it realize the balance between personal privacy and data open. After processing of the data by audit, in accordance with the basic principles of data open and release requirements can be directly submitted to visuliazion layer A3. Otherwise, all these data will return for reprocessing. In addition, even if remove identifier usually cannot protect the privacy, the attacker can get the patient's privacy by linking operation data, common disturbance, generalization,

anonymous, or distributed clustering privacy protection, data processing method of reducing the effect of data mining. In this layer, A2 contains data processing B4 data privacy protection B5, data privacy protection B6.

(3)Visualiazion A3. Medical big data will show the same content through the rich show forms for all users and at all personalized service levels, which include the residential health care professionals, hospital administration portal and some third parties. Through the residents health portal, the residents can visit their own electronic medical records, get medical and health information, make a appointment, ask for the doctors' recommend, have a healthy diet and related supporting services. Health care professionals can get professional information service function. Hospital administration can through hospital administrative management portal to understand hospital macro information in time, and implement comprehensive medical management. Scientific researchers and third parties can get related medical information services through the portal. In this layer, A3 contains patients data security B7, hospital service function security B8.

(4) Regulatoy layer A4. Medical big data secure sharing can not develop without the the laws, regulations and policies from government. In medical big data sharing, we must face in information security, privacy, and some other challenges. Such as the personal information in the electronic medical record is the focus of the medical dispute, is also the key of data sharing. We need the relevant legal safeguard patients privacy don't be leaked out. And medical big data sharing providers are all medical institutions, the government need make the data sharing rules, a series of related policy and legislations to strengthen and promote medical big data secure sharing. At the same time in order to secure sharing medical big data, need the professional standard of medical big data format, type and related process. In this layer, A4 contains factors: no uniform technical standard B9, lack of legal supervision B10.

4. Analytic Hierarchy Process

The foundation of the Analytic Hierarchy Process(AHP) is a set of axioms that carefully delimits the scope of the problem environment. It is based on the well-defined mathematical structure of consistent matrices and their associated right eigenvector's ability to generate true or approximate weights. The AHP methodology compares criteria, or alternatives with respect to a criterion, in a natural, pairwise mode. To do so, the AHP uses a fundamental scale of absolute numbers that has been proven in practice and validated by physical and decision problems experiments.

AHP algorithm basic steps[6]:

Step 1: Define the problem and determine the kind of knowledge sought. Structure the decision hierarchy from the top with the goal of the decision, then the objectives from a broad perspective, through the intermediate levels (criteria on which subsequent elements depend) to the lowest level (which usually is a set of the alternatives).

Step 2: Construct a set of pairwise comparison matrices. Each element in an upper level is used to compare the elements in the level immediately below with respect to it.

Use the priorities obtained from the comparisons to weight the priorities in the level immediately below. Do this for every element. Then for each element in the level below add its weight values and obtain its overall or global priority. Continues this process of weighting and adding until the final priorities of the alternatives in the bottom most level are obtained.

Judgment matrix setup: values from 9 to 1 are used to show factors important level (or decision makers' preferences) from the most to the least. This scaling method is recommended by psychologists who believe that people can distinguish no more than $7(\mp 2)$ objects. Table 1 presents the definitions of different values used in judgment matrix. Usually, values in the judgment matrix are assigned by experts or acquired from surveys.

C is $\mathrm{n} \times \mathrm{n}$ matrix, has following properties:

$\forall i, j \in N \triangleq \{1,2, \ldots, \mathrm{n}\}$,

$$\begin{cases} b_{ij} > 0 \\ b_{ii} = 1 \\ b_{ji} = {}^{1}/_{b_{ij}} \end{cases} \quad (1)$$

Table 1. Satty Nine Evaluation System

Intensity of Importance	Definition	Explanation
1	Equal importance	B_i and B_j contribute equally to the objective
3	Moderate importance	Experience and judgement slightly favor B_i over B_j
5	Strong importance	Experience and judgement strongly favor B_i over B_j
7	Very strong or demonstrate importance	An activity is favored very strongly over B_i, its dominance demonstrated in practice
9	Extreme importance	The evidence favoring B_i over B_j is of the highest possible order of affirmation
2,4,6,8	For compromise between the above values	Sometimes one needs to interpolate a compromise judgment numerically because there is no good word to describe it

Step 3: Arrangement of relative importance of criteria: this step determines factors' weight based on the relative importance of the layer of *i*+1 to the layer of *i* in the terms of element *j*. This calculation requires inputs from judgment matrix's maximum eigenvalue λ_{max} and its corresponding eigenvector, which could be acquired by polymer root algorithm.

Step 4: Consistency check: Consistency index (*CI*)

$$CI = \frac{\lambda_{max} - n}{n-1} \quad (2)$$

is used to evaluate judgment matrix's consistency. When$\lambda_{max} = n$,$CI = \frac{\lambda_{max}-n}{n-1} = 0$, complete consistency is achieved, and a large CI indicates lower judgment matrix's consistency. Generally, judgment matrix is consistent when $CI < 0.10$, or factors' weights needs to be reevaluated until reached new consistency. Judgment matrix's consistency has a correct this phenomenon. The corrected consistency is expressed in Eq. 2.

Step 5: Global ranking: all the factors within the same layer are assigned weights based on their relative importance to target layer. Ranking is carried from the highest layer to the lowest layer. Global ranking's weighted consistency(Eq. 3) is checked by *CI*, and the threshold value is also 0.10.

$$CR = CI/RI \quad (3)$$

Where λ_{max} is the maximum eigenvalue.

Table 2. Averaged random consistency index (*RI*)

Dimension of matrix(n)	1	2	3	4	5	6	7	8	9
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45

Table 3. Recursive calculation method of hierarchical total ordering

Layers	A_1	A_2	⋯	A_m	B layer hierarchical total ordering weight
	a_1	a_2	⋯	a_m	
B_1	c_{11}	c_{12}	⋯	c_{1m}	$b_1 = a_1c_{11} + a_2c_{12} + \cdots + a_1c_{1m}$
B_2	c_{21}	c_{22}	⋯	c_{2m}	$b_2 = a_1c_{21} + a_2c_{22} + \cdots + a_mc_{2m}$
⋮	⋮	⋮		⋮	⋮
B_n	c_{n1}	c_{n2}	⋯	c_{nm}	$b_n = a_1c_{n1} + a_2c_{n2} + \cdots + a_mc_{nm}$

5. Case Study

5.1 The Priciple of Medical Big Data Secure Sharing

(1) The principle of equality. Institutions or individuals have the equal right during medical big data sharing, that is, all organizations or individuals open to use Shared medical big data.

(2) Patient privacy protection principles. Medical big data are not only have 4V characteristics, but also it has its own biggest feature which is the patients' privacy. Compare to the traditional data security, it should also pay attention to the risk of medical big data correlation. Patient information once released out, in violation of the provisions of the patient information secrecy, not only bring the possibility of legal action, but also destroy the patient's trust in the doctors.

(3) The principle of standardization. Standardization is open formats such as XML, CSV and RDF (human and machine readable) release data as much as possible, or by a CGI script converts the data to a relational data standard form, form structured data. Open data file format is mainly on the network at present XML, RDF, CSV, Excel spreadsheets, such as JSON (Java Script Object Notation), text file (file format) classic, plain text files (TXT), scan images, proprietary format, such as HTML, including structured file formats are JSON, XML, RDF and CSV.

(4) The principle of timeliness. The core of the medical data open is to provide a better service for patients, improve data reuse value. But preparation and release data need time, published data shall establish priorities, such as the SARS epidemic contagious and data should be released timely, comprehensive, accurate, and are described in simple, clear and full specifications, convenient to carry out the investigation and study, reduce the expansion of the outbreak.

5.2 Class Hierarchy Model of Medical Big Data Secure Sharing

From previous analysis, decision-making goal includes four criteria layers. Each criteria layer includes two or three major categories factors (Fig. 1).

Step1: Through the identification of medical data sharing security risk factors, establish the corresponding class hierarchy model of secure medical data sharing. Here

A_i, B_j($i=1,2,3,4; j=1,2,\cdots,10$) means the factors in all layers,show as Fig.1.

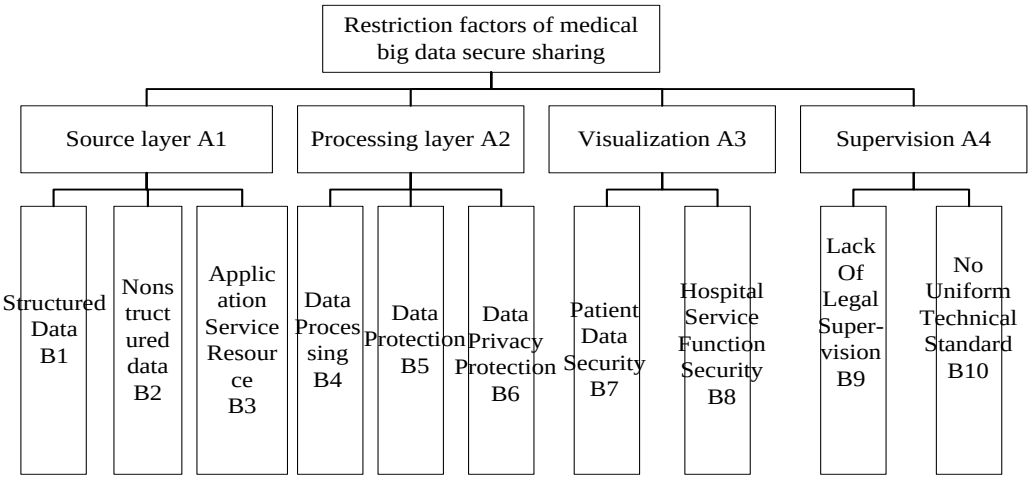

Fig. 1. Hierarchy model of medical big data secure sharing

Step2: According to the survey and fuzzy Delphi method, construct hierarchy judgment matrix *C* and check the consistency.

$C = \begin{pmatrix} 1 & 1/6 & 1/4 & 1/9 \\ 6 & 1 & 2 & 1 \\ 4 & 1/2 & 1 & 1/3 \\ 9 & 1 & 3 & 1 \end{pmatrix}$, here the matrix eigenvalue is $\lambda_{max} = 4.031$,substitute it into Eq. (2) and Eq. (3), achieved *CR*=0.0116<0.1, denotes the matrix satisfies the consistency.

Step3: Arrangement of relative importance of criteria. Calculate the judgment matrix's maximum eigenvalue's corresponding eigenvector, which could be acquired by polymer root algorithm. This step determines factors' weights based on the relative importance of elements A_1, A_2, A_3, A_4 in terms of the goal target layer, $u = (0.0491, 0.3511, 0.1700, 0.4298)$

Step 4: Similarly,calculate factors' weights based on the relative importance of elements in terms of the criteria layer, achieve the following:

1) Paired Comparisons. Compare B_1, B_2, B_3, achieve judgment matrix

$D_1 = \begin{pmatrix} 1 & 3 & 6 \\ 1/3 & 1 & 4 \\ 1/6 & 1/4 & 1 \end{pmatrix}$, where the max eigenvalue $\lambda_{max} = 3.0536$, substitute into formula (2) and (3), achieved *CR*=0.0516<0.1. Indicates the judgment matrix satisfies the consistency.

The weights vector of factors B_1, B_2, B_3 in terms of criteria factor $A_1$is $v_1 = (0.6442, 0.2706, 0.0852)$.

2) Paired Comparisons. Compare B_4, B_5, B_6, achieve the judgment matrix

$D_2 = \begin{pmatrix} 1 & 3 & 5 \\ 1/3 & 1 & 2 \\ 1/5 & 1/2 & 1 \end{pmatrix}$, where the max eigenvalue$\lambda_{max} = 3.0037$,substitute into formula (2) and (3), achieved *CR*=0.0036<0.1 Indicates the judgment matrix satisfies the consistency.

The weights vector of factors B_4, B_5, B_6 in terms of criteria factor A_2 is $v_2 = (0.6483, 0.2297, 0.1220)$

3) Paired Comparisons. Compare B_7, B_8 achieve judgment matrix

$$D_3 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \text{where the max eigenvalue}$$

$\lambda_{max} = 2$, substitute into formula (2) and (3), achieved *CR*=0<0.1. Indicates the judgment matrix satisfies the consistency. The weights vector of factors B_7, B_8 in terms of criteria factor A_3 is $v_3 = (0.5, 0.5)$.

4) Paired Comparisons Compare B_9, B_{10} achieve judgment matrix

$$D_4 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

where the max eigenvalue $\lambda_{max} = 2$ Substitute into formula (2) and (3), achieved *CR*=0<0.1 Indicates the judgment matrix satisfies the consistency. The weights vector of factors B_9, B_{10}in terms of criteria factor A_4 is $v_4 = (0.5, 0.5)$

Step 5: Global ranking. See table 4

Table 4. Risk ordering of medical big data secure sharing

B	A_1	A_2	A_3	A_4	Weight	Ordering
B_1	0.6442				0.0316	8
B_2	0.2706				0.0133	9
B_3	0.0852				0.0042	10
B_4		0.6483			0.2276	1
B_5		0.2297			0.0806	6
B_6		0.1220			0.0428	7
B_7			0.5000		0.0850	4
B_8			0.5000		0.0850	5
B_9				0.5000	0.2149	2
B_{10}				0.5000	0.2149	3

5.3 Results Analysis

Through the above analysis and calculation show that influential factors for medical big data secure sharing, the most factoer is B4 which is medical data processing.

6. Summary

In this paper, we discussed the restriction factors of medical big data secure sharing. Medical big data sharing has many personal and difficult problems. The research built up the restriction factor system via Delphi method, and calculate the weight of each factor by analytic hierarchy process(AHP), then the top node and ordering of all the factors are gotten. Base on the designed index system, the paper assess the restriction of medical big data secure sharing by fuzzy comprehensive analysis method. The experiment illustrates this index system can effectively quantify and assess the security level of medical big data sharing. In the future work, we will focus on the following issues:

(1)Strengthen the top design of medical big data secure sharing
(2) Improve the Technology system of medical big data secure sharing
(3) Design content standard of medical big data secure sharing
(4) Improve the mechanism of medical big data secure sharing
(5) Stengthen legal guarantees of medical big data secure sharing

Acknowledgement

This research was financially supported by the Ministry of Education of Humanities and Social Science project (NO. 14YJCZH020), National Social Sciences Program Fund (NO. 16BFX085).

References

[1] Xu K.T. Research of Architecture of Large Data Based Wisdom Mobile Medical Information System. Master Thesis. Qingdao Unversity of Science and Technology. 2015.

[2] Liu N., Chen M. Medical Data Open Method and Strategy Research. Chinese Hospital Management. 2015, 35(9): 37-39.

[3] Liu H, Li X. Research on Long-term Incentive Effect of Local Government Environmental Protection[J]. Journal of Applied Science and Engineering Innovation, 2018, 5(2): 42-46.

[4] Wang D.M.Research on Optimization of Apriori Algorithm Based on Cloud Computing and Medical Data. Beijing University of Posts and Telecommunication. 2015.

[5] Li J., Guo H.H., Guo M.J. and Dai T. A Quantitative Analysis On the Topic Distribution and Openness Degree of Health and Medical Data in the Open Government Practice of the US and UK. Library and Information Service. 2015, 10, 59(20): 132-137.

[6] Zhang S, Liang C. Research on Construction Schedule Control Based on Critical Chain Method and BIM[J]. Journal of Applied Science and Engineering Innovation, 2018, 5(2): 47-50.