Health Care Management Science (2020) 23:427–442
https://doi.org/10.1007/s10729-019-09490-4

A privacy protection method for health care big data management based on risk access control

Mingyue Shi[1,2] · Rong Jiang[1,2] · Xiaohan Hu[1,2] · Jingwei Shang[1,2]

Received: 12 March 2019 / Accepted: 1 July 2019 / Published online: 23 July 2019

Abstract
With the rapid development of modern information technology, the health care industry is entering a critical stage of intelligence. Faced with the growing health care big data, information security issues are becoming more and more prominent in the management of smart health care, especially the problem of patient privacy leakage is the most serious. Therefore, strengthening the information management of intelligent health care in the era of big data is an important part of the long-term sustainable development of hospitals. This paper first identified the key indicators affecting the privacy disclosure of big data in health management, and then established the risk access control model based on the fuzzy theory, which was used for the management of big data in intelligent medical treatment, and solves the problem of inaccurate experimental results due to the lack of real data when dealing with actual problems. Finally, the model is compared with the results calculated by the fuzzy tool set in Matlab. The results verify that the model is effective in assessing the current safety risks and predicting the range of different risk factors, and the prediction accuracy can reach more than 90%.

Keywords Health care big data · Health care management science · Privacy protection · Risk · Access control

1 Introduction

With the development of health care science and the renewal of management concepts, the fine management of the health care field has received more and more attention, and health care composition such as health care quality, doctor-patient communication, patient information, etc. Part of the management also goes from extensive to detailed, from traditional to innovative [1]. Health care digitization has also become the trend of the times. Many countries are actively promoting the development of health care information and intelligence [2]. Telemedicine and smart health care have become emerging fields in the health care industry. Due to the progress of science and technology, hospitals produce a large amount of data every day, including electronic health care records, pictures and clinical test data, etc. Big data has become the main driving force for the transformation of health care industry model, and the development of health care industry also makes it possible for big data to participate in information management [3].

The entry of big data into the health care field makes it difficult for traditional software or hardware to manage large and complex health care information. The health care management is human-centered and based on health care institutions, not only related to the internal resources of the hospital, management operation mechanism and service mode, but also it involves the security management of information related to patients, and the development of cloud platform provides scalable management space for large-scale health care information, reducing the cost of resource management. But in the cloud environment, patients' privacy protection issues are faced challenge. Compared with the traditional IT system, the risk points in the cloud environment have changed. For example, the traditional security boundary disappears, the cloud is exposed to the open network, the types and numbers of users are large, and the liquidity is high. In addition, in the cloud environment, the ownership management and use rights of resources are separated, users cannot directly control resources [4], and data security issues are gradually exposed, which has become an important research field in health care management science [1].

✉ Rong Jiang
jiang_rong@aliyun.com

[1] School of information, Yunnan University of Finance and Economics, Kunming, China

[2] Key Laboratory of Service Computing and Safety Management of Yunnan Provincial Universities, Kunming, China

Health care management is a special field. Its particularity is that it studies "people" [5]. All health care behaviors and their results are based on human information. The information leakage in the health care big data environment is not only the data itself, but more serious is that the hacker steals patients' social security accounts and personal finance, etc. by mining the hidden information behind the data, endangering patients' personal and property safety and even brings serious moral and ethical issues to hospitals [2].

The Symantec Internet Security Threat Report 2016 released the top 10 industries with the most data leak. Only the top three industries are introduced here: the first is the health care industry, with 116 data breaches and the proportion of incidents is 37.2%. The second is retail industry, where the number of data breaches is 34, and the proportion of events is 10.9%. The third is the education industry, where the number of data breaches is 31, and the proportion of events was 9.9%. The comparison of these data will find that the health care field has become the biggest victim, with a much higher percentage of data breaches than the second largest industry. Therefore, it is a matter of pride to do "Internet +" in the health care field, but the scientific management of health care data, especially the protection of patient privacy in the health care big data environment is imperative.

The rest of this paper is organized as follows. Section 2 describes the progress and shortcomings of the work related to the protection of health care big data privacy. Section 3 introduces the relevant theories and principles, then determines key indicators and quantifies the risks. In Section 4, the simulation experiment was carried out, and the validity and accuracy of the model were verified by the fuzzy logic tool set in Matlab. Section 5 summarizes this paper.

2 Related work

The information management problem involved in health care big data mainly refers to the protection of patient privacy. In terms of privacy protection, predecessors introduced various methods, such as: data desensitization technology, anonymous protection technology, data watermarking technology, data traceability technology, role mining, access control technology and later risk-based intelligent access control technology, each model and technology has its specific application scenarios [6]. With the advent of the era of big data, access control technology has also presented new features: judgment based on diversification, fuzzy (or uncertain) decision results and multiple access control technologies are integrated. This section analyzes the more representative methods of privacy protection and compares them with the methods in this paper.

At present, access control technology is a hot research topic in health care big data privacy protection, but the traditional access control technology is too demanding to adapt to this complex cloud environment. So some people introduced the concept of risk into admission control, a report published by JASON in 2004 was the first to introduce the concept of risk into the field of access control [7], the report gave guidelines for risk information system should meet: quantifying risk, establishing acceptable risk level, and ensuring that access is always controlled within acceptable risk level, this theory has made the research on access control based on risk become a hot spot. However, there are two important issues not mentioned in Jason's report: how to formulate corresponding strategies to keep the risk level within acceptable range and how to implement these policies.

According to the investigation, literature [8, 9] specified the risk access control strategy by extending XACML, and proposed the framework for implementing the strategy. Literature [10] studied the risk-adaptive access control model based on fuzzy multi-level, defined the concept of risk band, and divided the risk into different risk levels. The system dynamically controls the risk information flow according to the current operation demand, risk tolerance and environment. However, this method cannot adapt to the health care system in the cloud environment. Literature [11] proposed a risk-based adaptive access control for health care systems, which assigns a certain risk quota to each doctor. As curious doctors consume the risk value more quickly, it is easy to be found by the system administrator. The literature [12] has been improved on the basis of the literature [11], but there are still some shortcomings, such as the factors affecting data privacy leakage and the weight of each factor have not been comprehensively analyzed, only from the behavior of the two types of doctors Analysis; A method and framework for access control based on context-sensitive information is proposed for health care information system in [13]. Although this method has been greatly improved compared with the literature [11, 12], it not only considers the doctor's access behavior, but also analyzes the influence of factors such as resources, environment, subjects, patient symptoms and patient severity on risk. However, it fails to analyze the proportion of each factor in influencing the privacy leakage risk. Literature [14] introduced the risk assessment method under different parameters, but the method is not suitable for health care systems in the cloud environment, because there is no prior data to estimate the expected loss and default probability. Literature [15] established an access control model based on risk assessment, which first assigns permissions to operations based on roles, and then performs access rights allocation based on the sensitivity of the requester's access behavior. Although the model is theoretically a dynamic model, the access delay cannot be determined.

Literature [16] introduced the risk into the cloud-assisted health care system. It first checks the user's trust credentials, determines the user's access rights based on the role, and then evaluates the risk that users may pose. However, the model does not explain how to determine the availability, integrity, sensitivity, and weight of historical access records when calculating the ultimate risk. Literature [17] proposed the risk index weight allocation method, and designed the constrained multiple regression model to achieve the dynamic allocation of index weight. In addition, it improves the accuracy of risk assessment. However, it does not involve dynamic scheduling and elastic computing of cloud platform resources. Literature [18] investigated the big data security and privacy issues in the healthcare industry and discussed ways to solve these problems, but mainly explored anonymization and encryption methods.

The access control strategy based on risk access control research is not only based on experience, but also the subject, object, environment condition and historical access log analysis to calculate the risk value [19], this method has an obvious disadvantage is that the risk factors are all of certain value in the risk assessment, and the final risk level is also a certain value. However, risk represents the possibility of privacy leakage, which is an uncertain factor. Meanwhile, the various indexes affecting the risk are also constantly changing. Therefore, we introduce the concept of fuzzy theory, which is the basic idea of accepting the ambiguity phenomenon. In addition, some factors with unclear boundaries can be quantified into information that can be recognized by the computer. The literature [20] uses FCM and fuzzy rule-based techniques to calculate the risk value of IT at a specific stage in the patient-related visit path. The technology takes into account human intuitive perceptions, blurring patient information and risks, and helping healthcare professionals manage risk. Literature [21] proposed a risk assessment method based on fuzzy model, which considers the uncertainty analysis in risk assessment, including data sensitivity, behavior sensitivity and historical access risk. Literature [22] combined fuzzy theory, artificial neural network, wavelet analysis and quantum group optimization algorithm to propose the risk quantization method of wavelet fuzzy neural network. The fuzzy comprehensive evaluation is used to quantify the attribute information of the subject and the object as the input of the wavelet neural network, and the output is the quantified risk value. In addition, there are many researches on fuzzy theory in recent years, such as: Literature [23] proposed an information security management method. The author combines fuzzy logic theory with FMEA to analyze the security issues of access information and systems, communication security, infrastructure, and security management in information security. With the help of fuzzy set theory and probabilistic risk assessment technology, literature [24] has effectively dealt with such problems as uncertain factors and difficult quantification in information security risk assessment of industrial control systems. Literature [25] proposed a fuzzy extension technology (FBAC) on the basis of ABAC to improve the flexibility of authorization under special circumstances with the help of the fuzzy theory. At the same time, the degree of policy matching can be evaluated to achieve unattended exception authorization. Literature [26] evaluated the failure risk of underground water supply pipelines and developed a hierarchical reasoning system based on fuzzy theory. In addition, a heuristic based membership function determination (HBMFD) method was proposed. The literature [27] combined the analytic hierarchy process with TOPSIS (technique for order preference by similarity to ideal solution) to propose a risk assessment method based on AHP-TOPSIS and fuzzy sets, which successfully processed the uncertainty of expert subjective judgment in risk assessment. Literature [28] proposed an information security risk assessment model based on the combination of Event Tree Analysis and fuzzy decision theory to determine the ranking of alternatives based on the criticality. Literature [29] analyzed the failure modes and influencing factors of fuel cells in marine energy systems, and proposed a novel type-2 fuzzy logic, which can reduce the calculation time. Finally, experiments show that the method is effective in calculating risk. Literature [30] proposed a fuzzy logic theory based on expert judgment to realize risk-adaptive IoT access control model. The model performed a security risk assessment on the access request based on the context information of the requesting user, and solved the flexibility and dynamics of the Internet authorization.

In summary, compared with other research methods, fuzzy technology has the following advantages: firstly, fuzzy logic technology allows the imprecise definition of data, and at the same time, it can model the nonlinear function of arbitrary complexity [31–36]. Secondly, the fuzzy technology includes an expert experience, and more importantly, the fuzzy technology is dynamic. At last, in the health care system, patient information may be confidential, but some of the information is non-confidential. Therefore, this paper applies fuzzy theory to risk access control with the help of expert system, the patient's relevant health care information and privacy risk can be fuzzy evaluated, and the uncertainty related to risk assessment is solved [37]. The main contributions of this paper are as follows:

(1) Under the premise of the risk factors affecting the privacy leakage of health care big data, three key condition attributes are extracted by using the attribute reduction and discernibility matrix in the rough set theory: access behavior sensitivity, resource sensitivity and historical access risk.
(2) The risk access control model based on fuzzy theory is established, three key indexes are fuzzy treated, and the membership function between each index and the related fuzzy set is determined.
(3) According to the rule base, the relevant indicators are evaluated by rules, all possible results are enumerated and aggregated to obtain fuzzy sets.

(4) In order to obtain the final risk value, the fuzzy set is defuzzified by the central method.

(5) Finally, in order to verify that the model in this paper is effective and highly accurate, the comparative analysis is carried out with the fuzzy logic tool set in Matlab. The results show that the model is effective and the accuracy is over 90%.

3 Risk access control model based on fuzzy theory

The entities in the health care system in the cloud environment are shown in Fig. 1, mainly including information owners, information providers, information users and third-party platforms. With the support of cloud services, telemedicine, mobile health care, intelligent health care, cross-platform and cross-regional health care treatment, etc. have been realized, and patients can achieve health care services such as registration and health care treatment without going to the hospital. At the same time, in order to facilitate the doctor to timely understand the patient's health care history and illness, patients' health care records are stored on the cloud platform in electronic form, saving space and cost and improving the efficiency of health care treatment. However, as data is stored on third-party platforms, cloud service providers control the storage and operation environment of data, so patients cannot directly manage their own data, which brings great challenges to the security issues such as data availability, integrity and confidentiality. If unauthorized users make illegal modifications to the health care data, it will lead to the doctor's wrong diagnosis, or interception during the health care data transmission will greatly threaten the integrity of the data. In addition, some users may copy, modify and steal the sensitive health care data of patients, and the geographic location of patients may be tracked in telemedicine. Therefore, the confidentiality of data will not only violate the privacy of patients, it may also pose a hazard to the patient's personal safety. Finally, an attack on the availability of health care data under cloud services may result in the denial of services within some of the rights. If the necessary request services are not met in an emergency, the patient's life will be threatened.

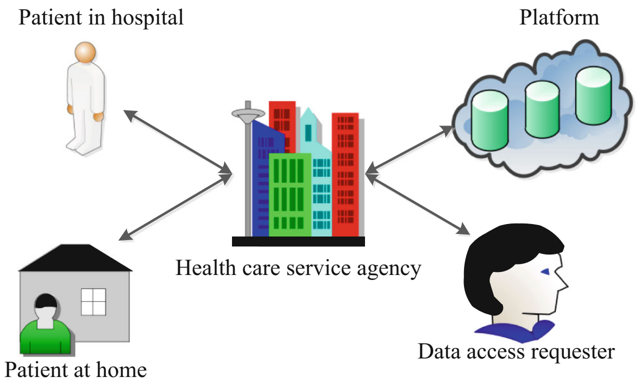


Fig. 1 Health care system structure under cloud service

Analysis based on relevant literature [20, 21, 38] this paper establishes a fuzzy logic system with the help of fuzzy rules and fuzzy toolbox to solve the problem of privacy leakage in health care big data environment. The specific fuzzy logic system principle is shown in Fig. 2.

3.1 Related theory and principle

This section will introduce the relevant theories of fuzzy logic theory and the specific contents of the four modules involved in fuzzy logic system: fuzzification, rule evaluation, rule aggregation and defuzzification.

3.1.1 Fuzzification

In the actual processing of the problem, the data we collect is usually clear, and the fuzzy logic system is based on the processing of fuzzy sets. Therefore, fuzzification is to map these clear data into the fuzzy set in the fuzzy logic system. In other words, the collected clear value x is changed to a certain proportion, and it is mapped to a real value on the fuzzy domain N.

The real value may belong to several fuzzy subsets in the fuzzy domain at the same time, and the membership degree of the real value belonging to each related fuzzy subset is calculated, which is called fuzzification. To make it easier to understand, an example is given to illustrate: when an access requester requests access information, the corresponding resource sensitivity is 0.2. The real value 0.2 belongs to the membership of fuzzy set NS is 0.6, while the membership of fuzzy set S is 0.3. In this way, calculate the membership degree of 0.2 belonging to each fuzzy subset, and the process is to obscure 0.2.

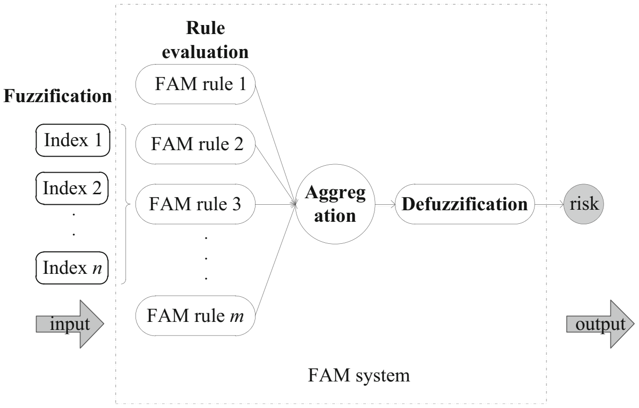


Fig. 2 Schematic diagram of fuzzy logic system

3.1.2 Rule evaluation

The premise of the rule evaluation is that the corresponding rule base has been established, and all the combinations are enumerated according to the fuzzy sets to which the indicators belong. The following is introduced through a simple example: assuming that the real values of x, y and z are fuzzified, the results are as follows: x ∈ {a, b},y ∈ {c, d}, z ∈ {e, f}, where a , b , c , d , e , f represent different fuzzy sets. From this, all the rule evaluation results are as follows.

Rule 1: If (x is a) and (y is c) then (z is e)
Rule 2: If (*x* is *a*) and (*y* is *d*) then (*z* is *e*)
Rule 3: If (*x* is *a*) and (*y* is *c*) then (*z* is *f*)
Rule 4: If (*x* is *a*) and (*y* is *d*) then (*z* is *f*)
Rule 5: If (*x* is *b*) and (*y* is *c*) then (*z* is *e*)
Rule 6: If (*x* is *b*) and (*y* is *d*) then (*z* is *e*)
Rule 7: If (*x* is *b*) and (*y* is *c*) then (*z* is *f*)
Rule 8: If (*x* is *b*) and (*y* is *d*) then (*z* is *f*)

3.1.3 Rule aggregation

Rule aggregation is to aggregate the results of rule evaluation, but in the process of rule aggregation, we need to split the membership function of each factor and then aggregate all the results together [39]. An example of rule aggregation is shown in Fig. 3:

3.1.4 Defuzzification

In fact, defuzzification is the inverse process of fuzzification. After being processed by the fuzzy logic system, the output is a fuzzy set. Because it is a result of multiple fuzzy control rules, its membership function is irregular and segmented. Defuzzification means that the result is equivalent to a certain value, that is, to map it to a representative value through a certain relationship. The whole process is called defuzzification. At present, the commonly used defuzzification methods are as follows: maximum mean method, area center method, and maximum membership degree method. Each method has its own advantages and disadvantages. The analysis of specific methods is not the focus of this paper. Just follow the principle of "justified, easy to calculate and continuous".

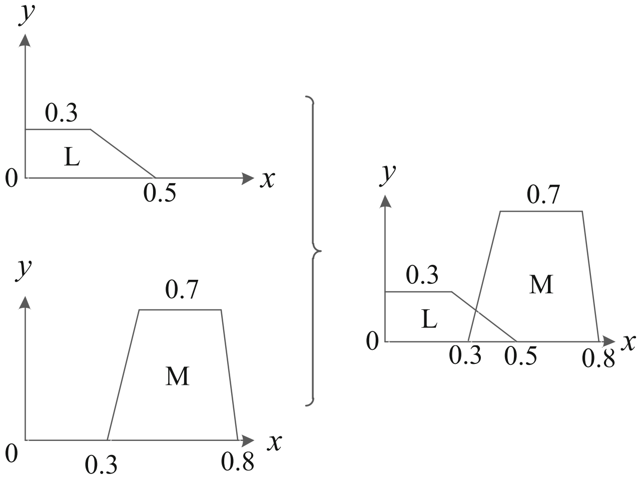


Fig. 3 Example of rule aggregation

3.2 Determining risk indicators

The access of the user in the health care system is recorded as a five-tuple(a, p, e, s, r)in which,

a (action)	Indicates the user's access behavior sensitivity
p (past)	Indicates the user's historical access risk
e (environment)	Indicates the network environment when the user requests access
s (sensitivity)	Indicates the resource sensitivity of the user requesting access
r (risk)	Indicates the risk of patient privacy data disclosure

The system access record is recorded as $V = \{v_1, v_2, \cdots, v_i\}$, and v_i represents the *i*th access record. The user's access behavior sensitivity, historical access risk, network environment, and resource sensitivity are called condition attributes, and the risk size is called decision attribute [40]. Here's the question: are all condition attributes useful for decision attributes? We need to reduce redundant attributes as much as possible and retain the necessary core attributes. In the rough set, data reduction is a very important research direction. Deleting redundant attributes in a big data environment can greatly improve the efficiency of decision making [41–45]. At present, there are many researches on rough sets. This chapter mainly proposes attribute reduction based on rough sets, and uses discernible matrix to solve this problem [40, 46].

Definition 1 (discernible matrix): Set $S = \{V, C\}$as the health care information system, where $V = \{v_1, v_2, \cdots, v_i\}$ is called the domain, C denotes the conditional attribute set, D denotes the decision attribute set, c(v) denotes the value of record v on the conditional attribute c, and the matrix is defined as follows:

$$w_{ij} = \begin{cases} \{c|c\in C\}; & c(v_i)\neq c(v_j) \text{ and } D(v_i)\neq D(v_j) \\ \quad 1 \quad ; & c(v_i) = c(v_j) \text{ and } D(v_i)\neq D(v_j) \\ \quad 0 \quad ; & D(v_i) = D(v_j) \end{cases}$$

Analysis of the above formula we will find the following rules:

(1) In the discernible matrix, if there is only one attribute, the attribute is the core attribute we are looking for, which can uniquely distinguish different decision attributes.
(2) Other attributes outside core attributes need to be extracted from non-core attribute combinations.

The following takes some information in the health care information system as an example to extract the core factors affecting patient privacy leakage by using the discernible matrix and mathematical logic principle, as shown in Table 1:

According to the definition of discernible matrix and Table 1, the following matrix form is obtained:

$$\begin{array}{cccccccccccccc}
0 & 0 & c_1 & c_1c_2 & c_1c_2c_3 & 0 & c_1c_2c_3c_4 & 0 & c_2c_3 & c_1c_2c_3 & c_2c_3c_4 & c_1c_2c_4 & c_1c_3 & 0 \\
0 & 0 & c_1c_4 & c_1c_2c_4 & c_1c_2c_3c_4 & 0 & c_1c_2c_3 & 0 & c_2c_3c_4 & c_1c_2c_3c_4 & c_2c_3 & c_1c_2 & c_1c_3c_4 & 0 \\
 & & 0 & 0 & 0 & c_1c_2c_3c_4 & 0 & c_1c_2 & 0 & 0 & 0 & 0 & 0 & c_1c_2c_4 \\
 & & & 0 & 0 & c_2c_3c_4 & 0 & c_1c_2 & 0 & 0 & 0 & 0 & 0 & c_4 \\
 & & & & 0 & c_4 & 0 & c_1c_3 & 0 & 0 & 0 & 0 & 0 & c_2c_3c_4 \\
 & & & & & 0 & c_1 & 0 & c_1c_4 & c_2c_4 & c_1c_2 & c_1c_2c_3 & c_1c_2c_4 & 0 \\
 & & & & & & 0 & c_1c_2c_3c_4 & 0 & 0 & 0 & 0 & 0 & c_1c_2c_3 \\
 & & & & & & & 0 & c_2c_3 & c_1c_3 & c_3c_4 & c_1c_4 & c_1c_2c_3 & 0 \\
 & & & & & & & & 0 & 0 & 0 & 0 & 0 & c_1c_2c_3c_4 \\
 & & & & & & & & & 0 & 0 & 0 & 0 & c_3c_4 \\
 & & & & & & & & & & 0 & 0 & 0 & c_1c_3 \\
 & & & & & & & & & & & 0 & 0 & c_1c_4 \\
 & & & & & & & & & & & & 0 & c_1c_2c_3c_4 \\
 & & & & & & & & & & & & & 0
\end{array}$$

By definition, the discernible matrix is a symmetric matrix, and the main diagonal element is 0.The core attributes are $\{c_1, c_4\}$, so the attribute combinations that do not contain core attributes may be $\{c_2\}$, $\{c_3\}$, $\{c_2, c_3\}$, observe matrix found only $\{c_2, c_3\}$ combinations did not contain core attributes. Therefore, the original decision table attribute can be simplified as $\{c_1, c_2, c_4\}$, $\{c_1, c_3, c_4\}$, which combination of attributes to choose depends on actual needs, if the goal is the simplest rule, the rules obtained by taking $\{c_1, c_2, c_4\}$ as conditional attribute sets are as follows:

(1) (c_1, write) ⇒ Risk of privacy leakage = high risk
(2) (c_1, read) and (c_2, excellent) ⇒ Risk of privacy leakage = low risk
(3) (c_1, copy) and (c_4, sensitive) ⇒ Risk of privacy leakage = high risk

Table 1 Some information in the health care system

V	Conditional attribute set C				Decision attribute D
	Access behavior c1	Network environment c2	Historical access risk c3	Resource sensitivity c4	Risk of privacy leaks
1	read	excellent	low	sensitive	low risk
2	read	excellent	low	not sensitive	low risk
3	write	excellent	low	sensitive	high risk
4	copy	general	low	sensitive	high risk
5	copy	good	high	sensitive	high risk
6	copy	good	high	not sensitive	low risk
7	write	good	high	not sensitive	high risk
8	read	excellent	low	sensitive	low risk
9	read	good	high	sensitive	high risk
10	copy	general	high	sensitive	high risk
11	read	general	high	not sensitive	high risk
12	write	general	low	not sensitive	high risk
13	write	excellent	high	sensitive	high risk
14	copy	general	low	not sensitive	low risk

Table 2 Input variables and ranges

Risk indicator	Category	Symbol	Normalized
access behavior sensitivity (a)	Low	L	[0,0.45]
	Middle	M	[0.4,0.6]
	High	H	[0.58,1]
resource sensitivity (s)	Not Sensitive	NS	[0,0.3]
	Sensitive	S	[0.25,0.55]
	High Sensitive	HS	[0.5,1]
historical access risk (p)	Low	L	[0,0.4]
	Middle	M	[0.37,0.65]
	High	H	[0.6,1]
risk (r)	Ignore	N	[0,0.2]
	Low	L	[0.1,0.4]
	Middle	M	[0.37,0.5]
	High	H	[0.48,0.8]
	Unacceptable	UH	[0.65,1]

(4) (c_1, copy) and (c_4, not sensitive) ⇒ Risk of privacy leakage = low risk
(5) (c_1, read) and (c_2, general) and (c_4, sensitive) ⇒ Risk of privacy leakage = low risk
(6) (c_2, good) and (c_4, sensitive) ⇒ Risk of privacy leakage = high risk
(7) (c_2, general) and (c_4, not sensitive) ⇒ Risk of privacy leakage = high risk

The rules obtained by taking $\{c_1, c_3, c_4\}$ as conditional attribute sets are as follows:

(1) (c_1, write) ⇒ Risk of privacy leakage = high risk
(2) (c_1, read) and (c_3, low) ⇒ Risk of privacy leakage = low risk
(3) (c_1, copy) and (c_4, sensitive) ⇒ Risk of privacy leakage = high risk
(4) (c_1, copy) and (c_4, not sensitive) ⇒ Risk of privacy leakage = low risk
(5) (c_1, read) and (c_3, high) ⇒ Risk of privacy leakage = high risk

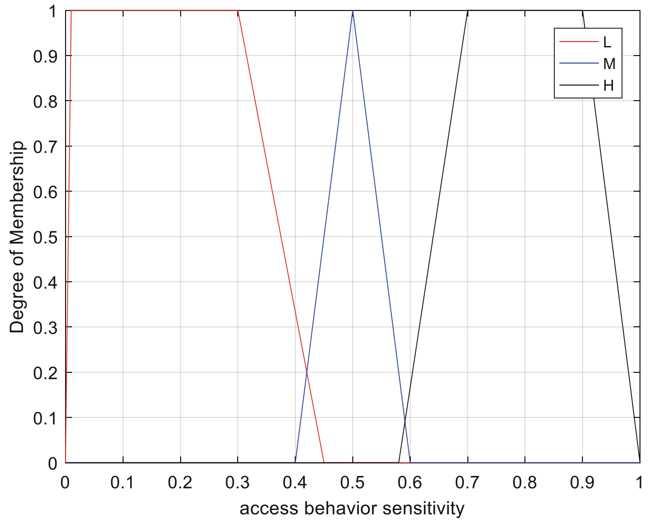


Fig. 4 The graph of the relationship function of behavior sensitivity

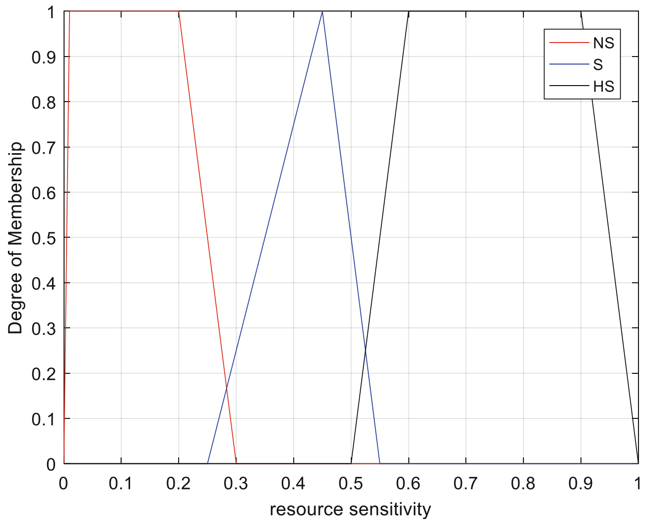


Fig. 5 The graph of the relationship function of resource sensitivity

Thus, $\{c_1, c_2, c_4\}$ can get 7 rules as the set of conditional attributes, and $\{c_1, c_3, c_4\}$can get 5 rules as the set of conditional attributes, so the simplest conditional attributes set of the original decision table is$\{c_1, c_3, c_4\}$.

3.3 Risk quantification

In section 3.2, three key indicators affecting privacy leakage risk have been identified, namely: access behavior sensitivity, resource sensitivity, and historical access risk. (1) Access behavior sensitivity: for example, if a doctor

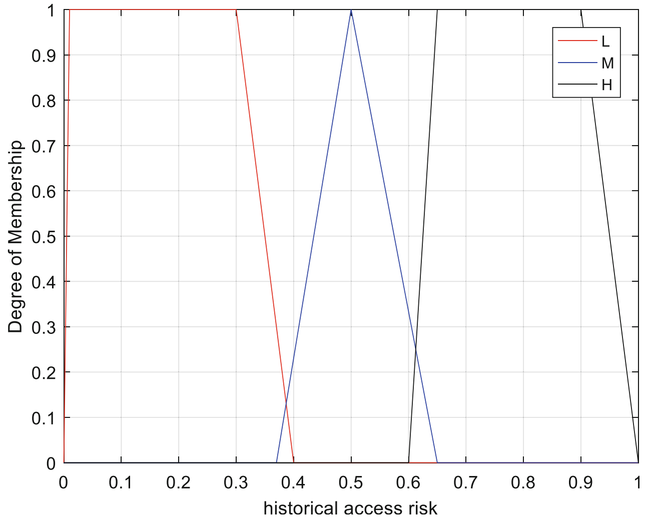


Fig. 6 The graph of the relationship function of historical access risk

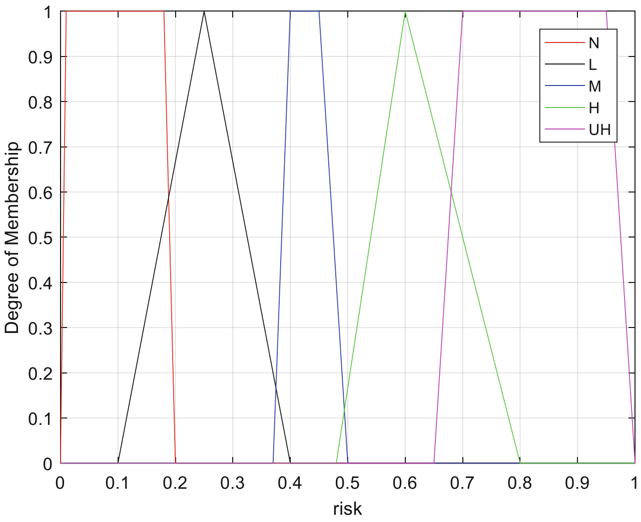


Fig. 7 The graph of the relationship function of risk

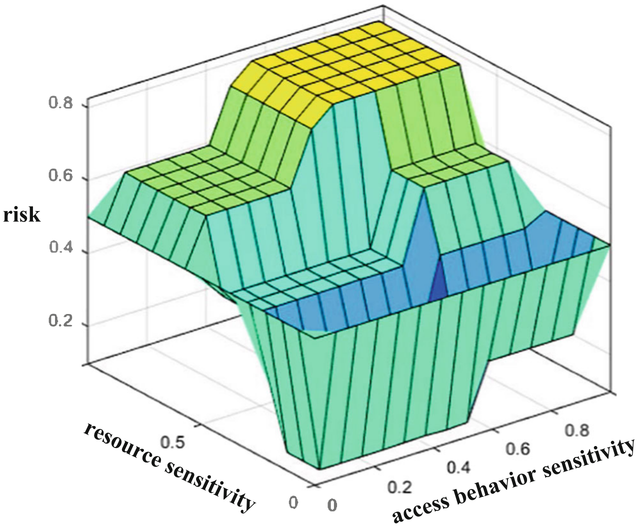


Fig. 8 The regular interface between access behavior sensitivity and resource sensitivity

"writes" a patient's health care data, the confidentiality, integrity and availability of the data will be affected, but if it is a "read" operation, it will affect the "confidentiality" of the data and availability will not be affected. (2) Resource sensitivity: The risk of data leakage is directly related to its own sensitivity. For example, the name and family address of an HIV patient are very private information. If it is leaked, it will cause great harm to the hospital and the patient itself. And gender, age, etc. are relatively less sensitive. (3) Historical access risk: the greater the risk value of privacy leakage in a user's historical access record, the greater the risk of privacy data leakage in the future.

In a fuzzy system, the input and associative outputs are some fuzzy evaluation values rather than exact values. Firstly, the input variables and associative output are

Table 3 Rule base

Fuzzy associative memory rules	
1	If (a is L) and (s is NS) and (p is L) then (r is N)
2	If (a is M) and (s is NS) and (p is L) then (r is N)
3	If (a is H) and (s is NS) and (p is L) then (r is N)
4	If (a is L) and (s is S) and (p is L) then (r is L)
5	If (a is M) and (s is S) and (p is L) then (r is L)
6	If (a is H) and (s is S) and (p is L) then (r is L)
7	If (a is L) and (s is HS) and (p is L) then (r is M)
8	If (a is M) and (s is HS) and (p is L) then (r is M)
9	If (a is H) and (s is HS) and (p is L) then (r is M)
10	If (a is L) and (s is NS) and (p is M) then (r is N)
11	If (a is M) and (s is NS) and (p is M) then (r is N)
12	If (a is H) and (s is NS) and (p is M) then (r is L)
13	If (a is L) and (s is S) and (p is M) then (r is M)
14	If (a is M) and (s is S) and (p is M) then (r is M)
15	If (a is H) and (s is S) and (p is M) then (r is H)
16	If (a is L) and (s is HS) and (p is M) then (r is H)
17	If (a is M) and (s is HS) and (p is M) then (r is UH)
18	If (a is H) and (s is HS) and (p is M) then (r is UH)
19	If (a is L) and (s is NS) and (p is H) then (r is L)
20	If (a is M) and (s is NS) and (p is H) then (r is M)
21	If(a is H) and (s is NS) and (p is H) then (r is H)
22	If (a is L) and (s is S) and (p is H) then (r is UH)
23	If (a is M) and (s is S) and (p is H) then (r is UH)
24	If (a is H) and (s is S) and (p is H) then (r is UH)
25	If (a is L) and (s is HS) and (p is H) then (r is UH)
26	If (a is M) and (s is HS) and (p is H) then (r is UH)
27	If (a is H) and (s is HS) and (p is H) then (r is UH)

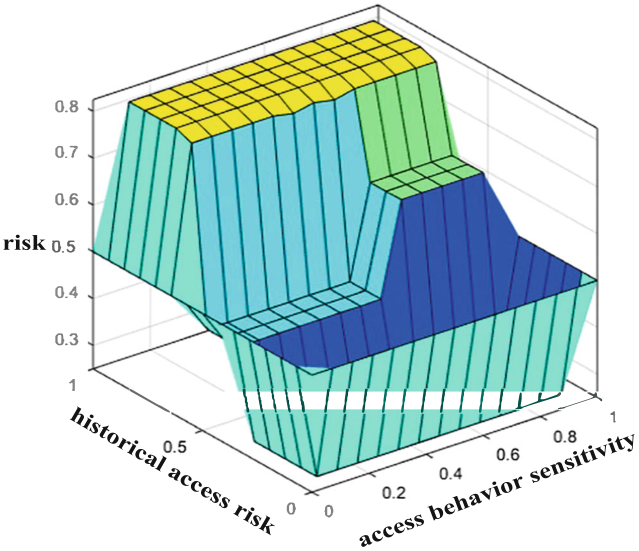


Fig. 9 A regular interface between access behavioral sensitivity and historical access risk

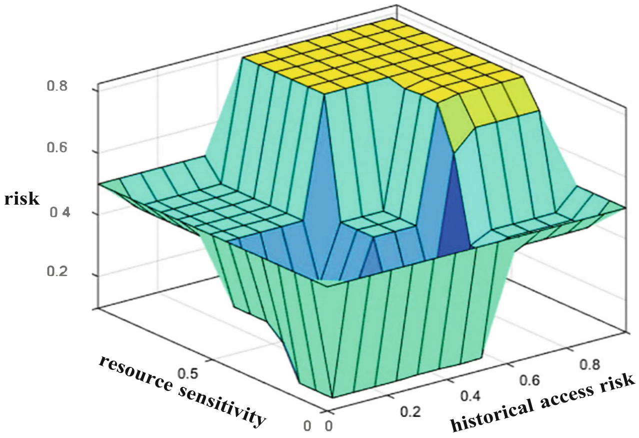


Fig. 10 A regular interface between resource sensitivity and historical access risk

Table 4 The degree of membership corresponding to the risk index

Risk indicator	Membership	
a = 0.43	L	M
	0.13	0.3
s = 0.28	NS	S
	0.2	0.15
$p = 0.25$	L	
	1	

fuzzified, and the index factors and risk level are roughly divided into different categories. Access behavior sensitivity is roughly divided into three categories: {Low, Middle, High}, resources sensitivity roughly divided into {Not Sensitive, Sensitive, High Sensitive}, historical access risk roughly divided into {Low, Middle, High}, and the output of risk is divided into five categories {Ignore, Low, Medium, High, Unacceptable}. In fuzzy logic, these categories are called fuzzy sets, in which each index and risk belong to a fuzzy set with a membership degree between 0 and 1, and there is no clear demarcation point between each fuzzy set, so that it can improve the fault tolerance rate.

As shown in Table 2, the input variables and ranges are given. According to Table 2, the relation function (membership degree function) corresponding to each indicator set and risk can be determined. This function returns the membership degree of the variable in the fuzzy set. As mentioned above, this paper mainly uses fuzzy tools to process fuzzy sets. There are many shapes of membership function in fuzzy tools. Here select the simplest triangle and trapezoid with the most abundant expert knowledge, as shown in Figs. 4, 5, 6, and 7.

After determining the relationship function of risk and each indicator, the next step is to determine the rule base to associate risk with indicators at different levels. It is known from experience that under certain conditions of

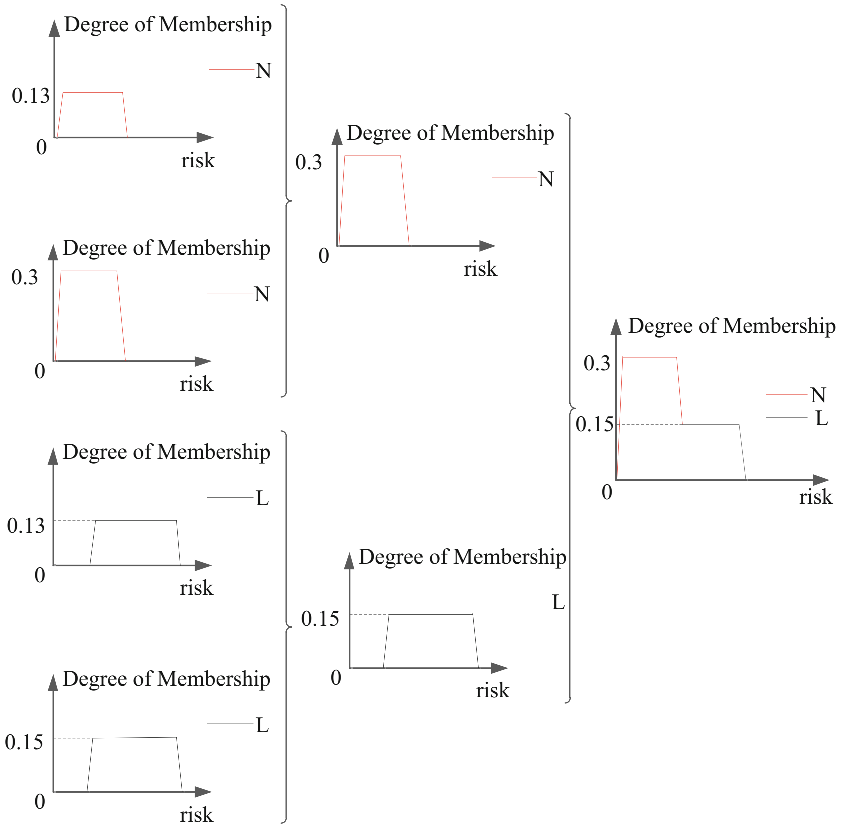
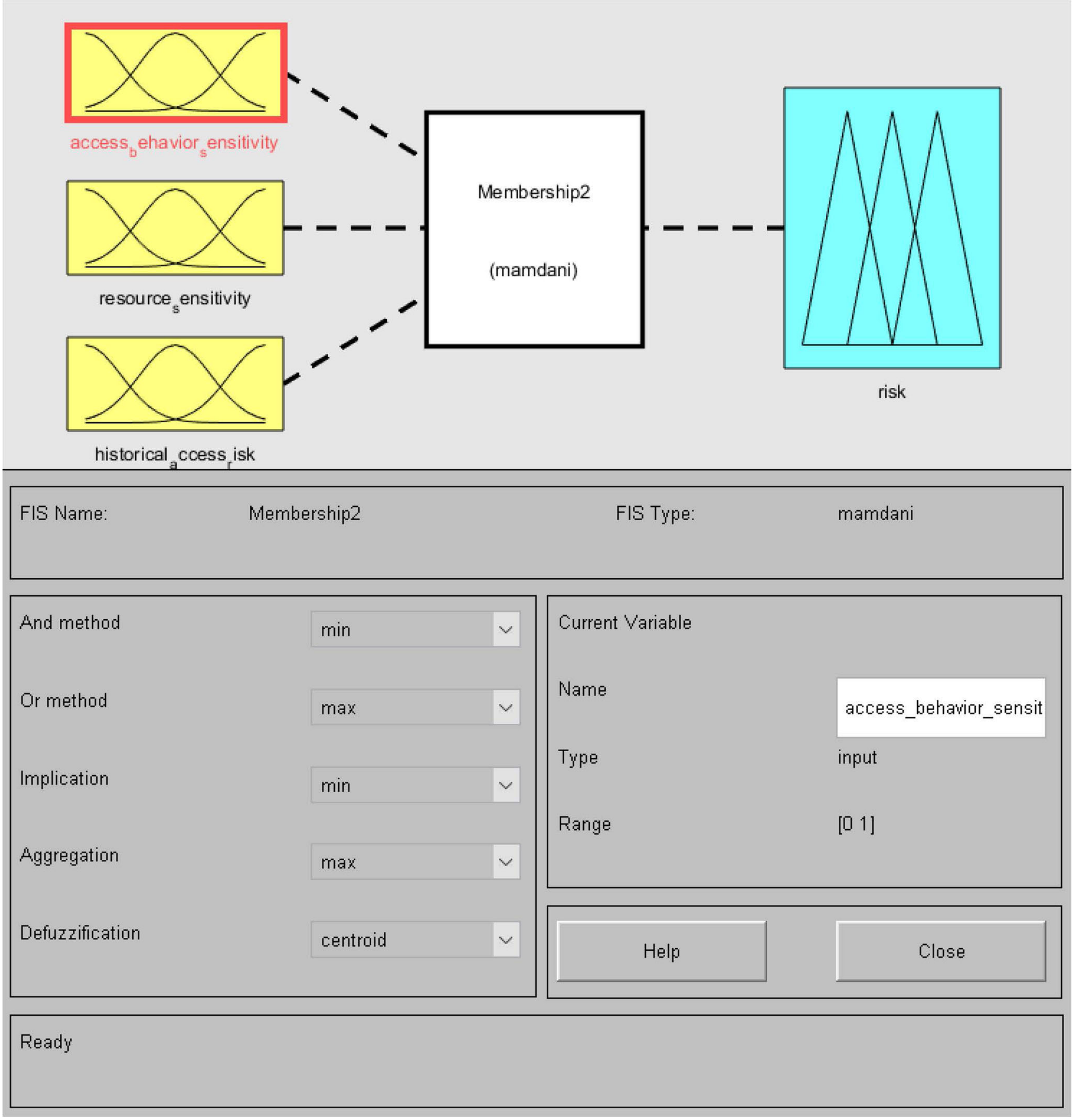


Fig. 11 The process of rule aggregation



FIS Name: Membership2

FIS Type: mamdani

And method: min

Or method: max

Implication: min

Aggregation: max

Defuzzification: centroid

Current Variable

Name: access_behavior_sensit

Type: input

Range: [0 1]

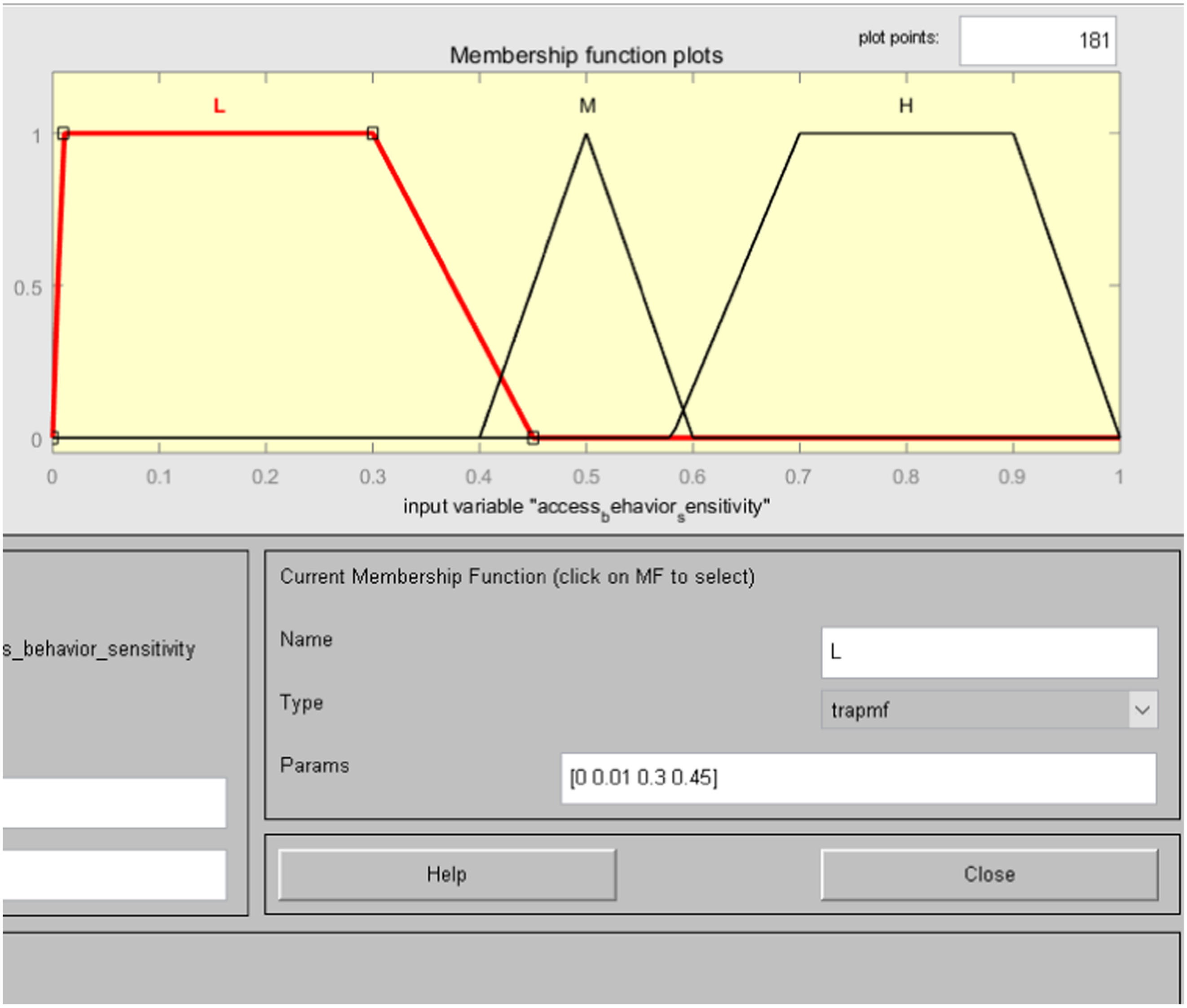
Ready

Fig. 12 Fuzzy design of input and output variables

other indicators, historical access risk *p* is positively correlated with risk *r*. Combine the access behavior sensitivity and resource sensitivity for more complicated and detailed analysis, and get all the possible situations as shown in Table 3, resulting in a total of 27 rules [21]. Finally, by fixing one of the indicators, the impact of the other two indicators on the risk is analyzed, and the three-dimensional graph shown in Figs. 8, 9, and 10 are generated to facilitate the intuitive analysis of the performance of the fuzzy system.

4 Simulation experiment

Figure 2 shows the schematic diagram of the fuzzy logic system established by means of fuzzy rules and fuzzy toolbox. It



Current Membership Function (click on MF to select)

Name: L

Type: trapmf

Params: [0 0.01 0.3 0.45]

Fig. 13 Membership function design of input and output variables

is divided into four steps: Fuzzification, Rule evaluation, Rule aggregation and Defuzzification. Next, use an example to analyze how the system implements privacy risk assessment and access control. Suppose the system receives an access request message: *A* requests to read the health care information of a patient. The three indicators corresponding to the request are: $a = 0.43$, $s = 0.28$, $p = 0.25$.

4.1 Experiment procedure

4.1.1 Fuzzification

The access behavior sensitivity, resource sensitivity and historical access risk corresponding to requester *A* are mapped to approximate fuzzy sets respectively.$a = 0.43 \in \{L, M\}$, $s = 0.28 \in \{NS, S\}$,$p = 0.25 \in \{L\}$.

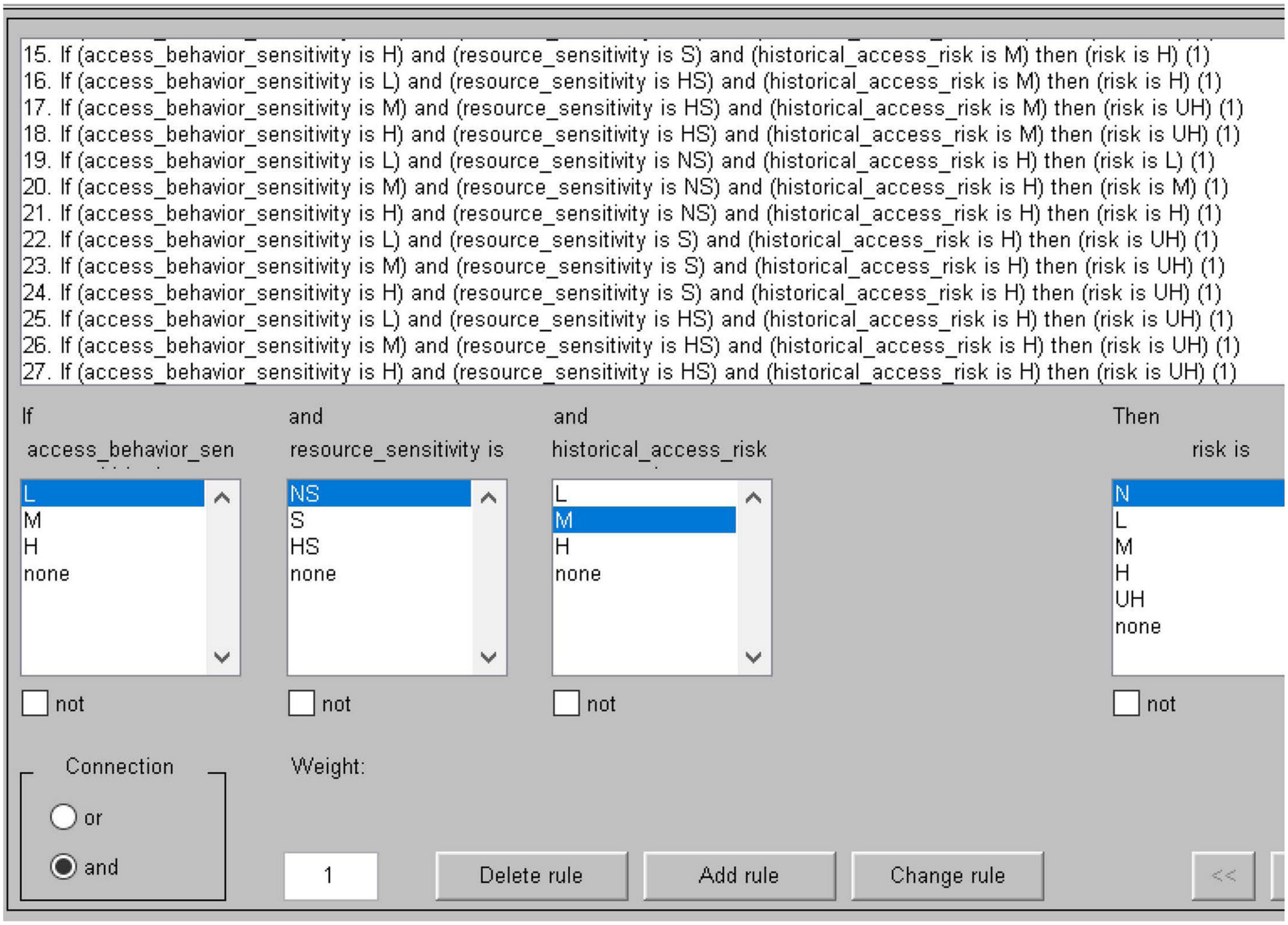
According to the relationship function of each index, the membership degree as shown in Table 4 is obtained.

4.1.2 Rule evaluation

By definition, rule evaluation is to enumerate all the combined results of the fuzzy set to which the indicator belongs. The following rules are obtained from Table 4:

Rule 1: If (a is L) and (s is NS) and (p is L) then (r is N)
Rule 2: If (a is L) and (s is S) and (p is L) then (r is L)
Rule 3: If (a is M) and (s is NS) and (p is L) then (r is N)
Rule 4: If (a is M) and (s is S) and (p is L) then (r is L)

Since it is an AND operation, the min function is selected here as the fuzzy function:



15. If (access_behavior_sensitivity is H) and (resource_sensitivity is S) and (historical_access_risk is M) then (risk is H) (1)
16. If (access_behavior_sensitivity is L) and (resource_sensitivity is HS) and (historical_access_risk is M) then (risk is H) (1)
17. If (access_behavior_sensitivity is M) and (resource_sensitivity is HS) and (historical_access_risk is M) then (risk is UH) (1)
18. If (access_behavior_sensitivity is H) and (resource_sensitivity is HS) and (historical_access_risk is M) then (risk is UH) (1)
19. If (access_behavior_sensitivity is L) and (resource_sensitivity is NS) and (historical_access_risk is H) then (risk is L) (1)
20. If (access_behavior_sensitivity is M) and (resource_sensitivity is NS) and (historical_access_risk is H) then (risk is M) (1)
21. If (access_behavior_sensitivity is H) and (resource_sensitivity is NS) and (historical_access_risk is H) then (risk is H) (1)
22. If (access_behavior_sensitivity is L) and (resource_sensitivity is S) and (historical_access_risk is H) then (risk is UH) (1)
23. If (access_behavior_sensitivity is M) and (resource_sensitivity is S) and (historical_access_risk is H) then (risk is UH) (1)
24. If (access_behavior_sensitivity is H) and (resource_sensitivity is S) and (historical_access_risk is H) then (risk is UH) (1)
25. If (access_behavior_sensitivity is L) and (resource_sensitivity is HS) and (historical_access_risk is H) then (risk is UH) (1)
26. If (access_behavior_sensitivity is M) and (resource_sensitivity is HS) and (historical_access_risk is H) then (risk is UH) (1)
27. If (access_behavior_sensitivity is H) and (resource_sensitivity is HS) and (historical_access_risk is H) then (risk is UH) (1)

If access_behavior_sen: L

and resource_sensitivity is: NS

and historical_access_risk: M

Then risk is: N

Connection: ○ or ● and

Weight: 1

Fig. 14 Edit rule base

Rule 1: N = min(0.13,0.2,1) = 0.13
Rule 2: L = min(0.13,0.15,1) = 0.13
Rule 3: N = min(0. 3,0.2,1) = 0. 3
Rule 4: L = min(0. 3,0.15,1) = 0.15

4.1.3 Rule aggregation

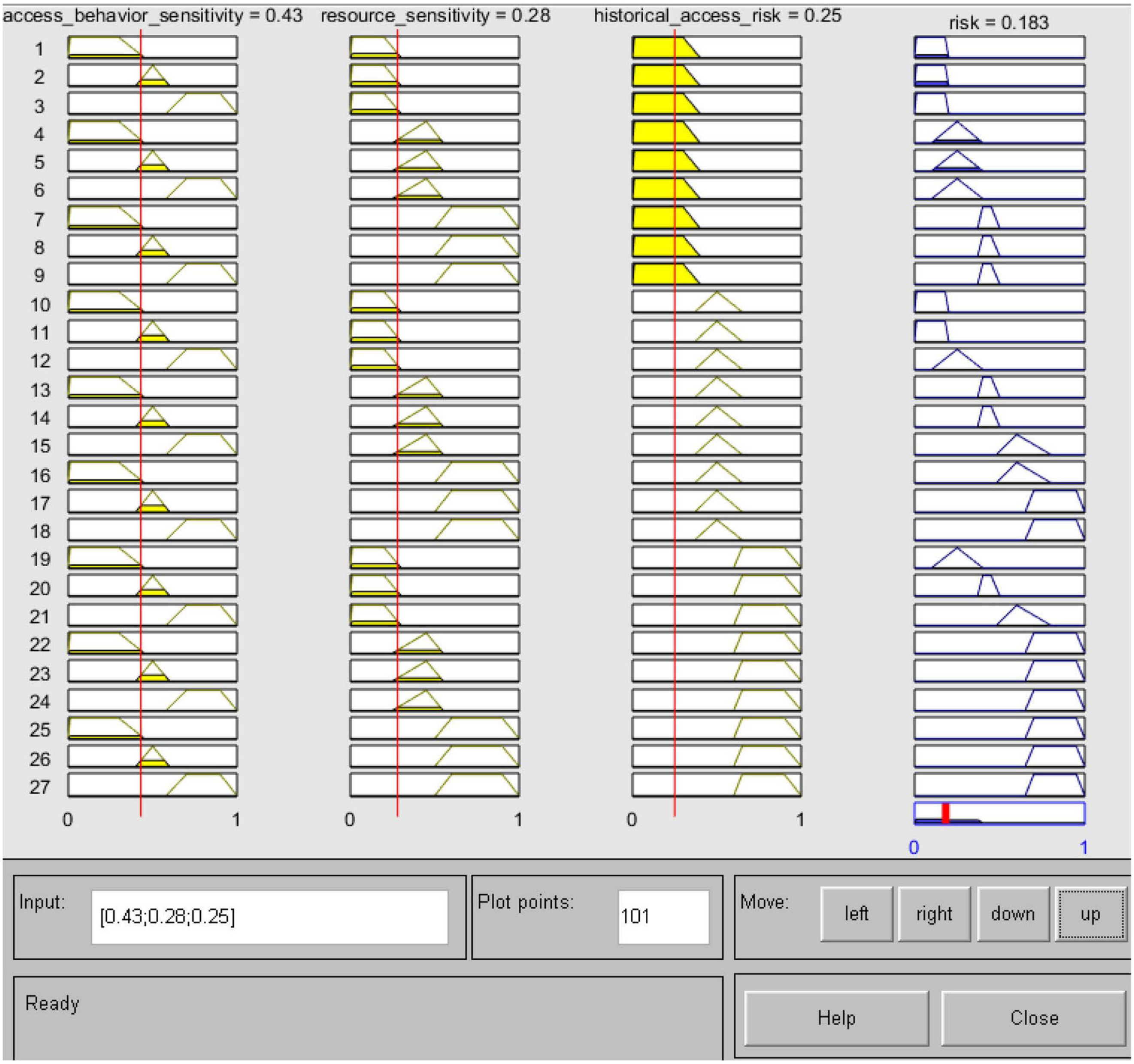
According to the results of the rule evaluation, the fuzzy relationship function of each risk level is tailored, and then the split results are aggregated. Final result is shown in Fig. 11:

4.1.4 Defuzzification

From the defuzzification theory introduced in Section 3.1.4, the final result of defuzzification is to obtain an accurate risk value. We use one of the most accurate, but also the most complex method - the center method, which is the principle to sample different expected values, and then average processing to calculate the sum of the contributions of each sample point to the overall membership. For example, the final risk value obtained by the central method is 0.167, which means that A's access request may cause the health care privacy data leakage risk is very small, or even can be ignored. The risk value determines whether the access request is allowed. If the risk value is within the tolerance range of the system, it is allowed to access or take certain risk mitigation measures to reach the risk threshold set of the system. Conversely, if the risk value is very large and exceeds the range that the system can tolerate, the access request is rejected directly.

4.2 Result analysis

In the end, we analyze the experimental results and verify the validity and accuracy of the model. However, due to the lack of real data in the research on health care big data privacy protection based on fuzzy theory, it is difficult to compare and analyze with other methods. Secondly, in



access_behavior_sensitivity = 0.43 resource_sensitivity = 0.28 historical_access_risk = 0.25 risk = 0.183

Input: [0.43;0.28;0.25]

Plot points: 101

Ready

Fig. 15 The actual results calculated by the fuzzy toolkit

previous research methods, few scholars have introduced the accuracy of risk assessment based on fuzzy theory. Therefore, there is a lack of comparative reference data in terms of accuracy comparison. In this paper, the effectiveness and accuracy of the model are compared and analyzed by means of the fuzzy logic tool set in Matlab. The specific operation process and results are shown in Figs. 12, 13, 14, and 15.

Input: [0.43; 0.28; 0.25]
Output: risk = 0.183

The risk obtained by the model in this paper is 0.167, while the risk calculated by the fuzzy logic tool in Matlab is 0.183. The comparison shows that the accuracy of the model reaches 91.25%. However, in order to avoid the contingency of the results, this paper used the same method to do 50 sets of simulation experiments, the results shown in Fig. 16.

Through the simulation experiment and the comparative analysis results shown in Fig. 16, it can be seen that the model in this paper can effectively evaluate the risk of privacy leakage of health care big data, and can

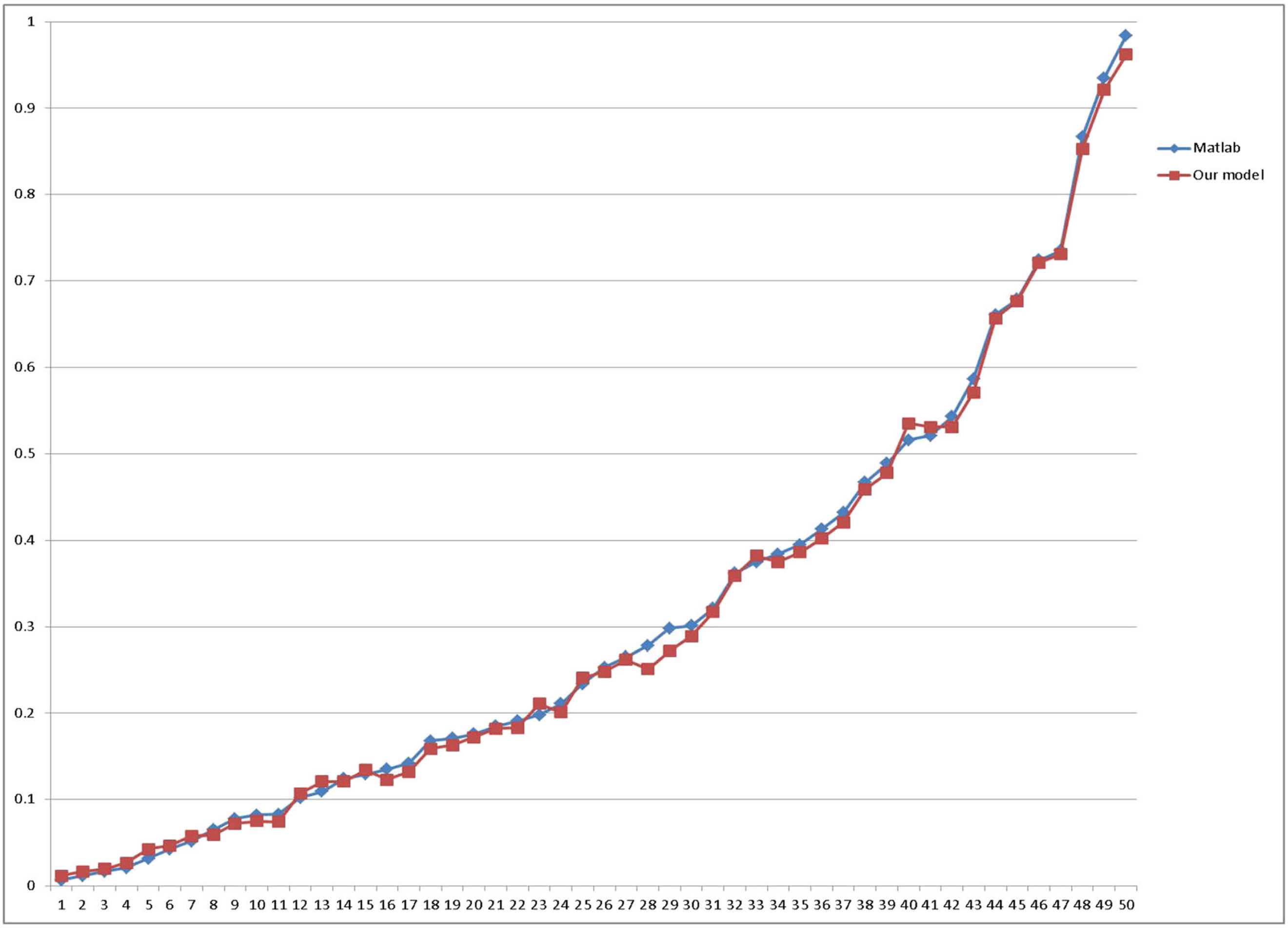


Fig. 16 Comparative test results

accurately predict the possibility of risk under different risk factors.

5 Conclusion

The development of health care big data is of great significance for building intelligent health care, and promotes the development of hospital management towards digital, intelligent and scientific. With the advancement of cloud computing technology, the application of regional intelligent health care service platforms will have more room for expansion, and the development prospects of intelligent health care cloud will be even better. However, the application of health care big data also has its own advantages and disadvantages. This paper starts from the essence of health care, management and science, people-oriented, patient-centered, research on information security under the health care big data environment, and establishes a medical system based on fuzzy theory to assess the risk of patient medical information leakage from three aspects: resource sensitivity, access behavior sensitivity and historical access information. Due to the lack of complete data in the actual problem-solving process, it is difficult to ensure the accuracy and reliability of risk analysis. Therefore, modeling is carried out from the ambiguity of risk factors, using fuzzy rule technology to solve the uncertainty in practical application, improving the performance of risk assessment methods, so that hospital managers can do their best in making decisions, and provide prediction and guidance for health care work, continuously improve and improve the intelligent health care management model, further realize refined health care services, and promote the comprehensive progress of hospital comprehensive management capabilities.

Acknowledgements This work was supported by National Natural Science Foundation of China (Nos. 61763048,61263022, 61303234), National Social Science Foundation of China (No. 12XTQ012), Innovation and Promotion of Education Foundation Project of Science and Technology Development Center of Ministry of Education (No. 2018A01042), Science and Technology Foundation of Yunnan Province (No. 2017FB095), the 18th Yunnan Young and Middle-aged Academic and Technical Leaders Reserve Personnel Training Program (No.2015HB038).

References

1. Guan ZJ (2014) Reflections on the Theory and Practice of Medical Management Science[J]. China Medical Management Science 4(4): 5–7 (In Chinese)
2. Niu Y, Yan MM, Zheng H, Yang JJ (2016) A Review of Medical Data Access Control in Cloud Computing Environment [J]. Wisdom Health 2(23):23–27 (In Chinese)
3. Liu WW (2018) Discussion on the Significance of Hospital Archives Management under the Background of Wisdom Medical Management in the Big Data Era[J]. Information Medical 7(19):155–156 (In Chinese)
4. China academy of communications (2018) White paper on cloud computing [R]. (In Chinese)
5. Yu GJ, Yang JH (2015) Medical big data [M]. Shanghai science and technology press. (In Chinese)
6. Li H, Zhang M, Feng DG, Hui Z (2017) Big data access control research [J]. J Comput Sci 40(1):1–11 (In Chinese)
7. JASON (2004) Report: Broader Access Models for Realizing Information DomiCorporation[R]. MITRE Corporation JSR-04-132
8. Chen C, Han W, Yong J (2010) Specify and enforce the policies of quantified riskadaptive access control[C]. In Proceedings of the 14th International Conference on Computer Supported Cooperative Work in Design Shanghai. China: CSCWD
9. Zheng Q, Han WL (2011) XACML based quantitative risk adaptive access control research and implementation [D]. Fudan University, Shanghai (In Chinese)
10. Pau-Chen C, Rohatgi P, Keser C (2007) Fuzzy Multi Level Security An Experiment on QuantifiedRisk Adaptive Access Control[C]. In Proceedings of the IEEE Symposium on Security and Privacy. Oakland, USA: S&P: 222–230
11. Wang Q,Hong J (2011) Quantified risk-adaptive access control for patient privacy protection in health information systems[C]. In Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security. Hong Kong, China: ASIACCS: 406–410
12. Hui Z, Li H, Zhang M, Feng DG (2015) Risk adaptive access control model for medical big data [J]. J Commun 36(12):190–199 (In Chinese)
13. Choi D, Kim D, Park S (2015) A framework for context sensitive risk-based access control in medical information systems[J]. Comput Math Methods Med 34(7):1–9
14. Wawrzyniak D (2006) Information Security Risk Assessment Model for Risk Management[J]. Springer Lect Notes Comput Sci 4083:21–30
15. Dipe N, Hung LX, Zhung Y, Lee S (2007) Enforcing Access Control Using Risk Assessment [J]. Universal Multiserv Netw 2: 419–424
16. Sharma M, Bai Y, Chung S, Dai L (2012) Using Risk in Access Control for Cloud Assisted eHealth[A]. In Proceedings of the 14th International Conference on High Performance Computing and Communications .IEEE computer society 1047–1049
17. Yang HY, Ning YG (2018) Weight allocation method of adaptive risk assessment index for access control of cloud platform [J]. Comput Appl 3(20):1–5 (In Chinese)
18. Abouelmehdi K et al (2018) Big healthcare data: preserving security and privacy [J]. J Big Data 5:1
19. Xie WC, Yang YJ (2013) A risk-based access control framework and related technology research [D]. PLA Information Engineering University, Zhengzhou (In Chinese)
20. Smith E, Eloff J (2000) Cognitive Fuzzy Modeling for Enhanced Risk Assessment in a Health Care Institution [J]. IEEE Intell Syst 5(4):69–74
21. Li J, Bai Y, Zaman N (2103) A fuzzy modeling approach for risk-based access control in eHealth cloud[A]. In Proceedings of the 12th IEE internationl conference on trust , security and privacy in computing and communications 17–21
22. Shi XJ, Yu WH (2018) Access control risk quantization method based on fuzzy neural network [J]. Intell Comput Appl 8(1):1–4 (In Chinese)1
23. Silva MM, Gusmao APH, Poleto T (2014) A multidimensional approach to information security riskmanagement using FMEA and fuzzy theory [J]. Int J Int Manag 34:733–740
24. Shang WL, Gong TY, Chen CY et al (2019) Information Security Risk Assessment Method for Ship Control System Based on Fuzzy Sets and Attack Trees [J]. Secur Commun Netw. https://doi.org/10.1155/2019/3574675
25. Xu Y, Gao W, Zeng QR et al (2018) A feasible fuzzy-extended attribute-based access control technique [J]. Secur Commun Netw. https://doi.org/10.1155/2018/6476315
26. Fayaz M, Ahmad S, Hang L et al (2019) Water Supply Pipeline Risk Index Assessment Based on Cohesive Hierarchical Fuzzy Inference System [J]. Processes 7(128):1–15
27. Ak MF (2018) AHP–TOPSIS integration extended with Pythagorean fuzzy sets for information security risk analysis [J]. Complex Intell Syst 12:1–14
28. Gusmao APH, Silva LC, Silva MM et al (2016) Information security risk analysis model using fuzzy decision theoryAna [J]. Int J Int Manag 36:25–34
29. Bahrebar S, Blaabjerg F, Wang HA et al (2018) A Novel Type-2 Fuzzy Logic for Improved Risk Analysis of Proton Exchange Membrane Fuel Cells in Marine Power Systems Application [J]. Energies 11(721):1–16
30. Atlam HF, Walters RJ et al (2019) Fuzzy logic with expert judgment to implement an adaptive risk-based access control model for IoT [J]. Mob Netw Appl. https://doi.org/10.1007/s11036-019-01214-w
31. Niknam T, Khooban MH (2013) Fuzzy sliding mode control scheme for a class of non-linear uncertain chaotic systems[J]. IET Sci Meas Technol 7:249–255
32. Khooban MH, ShaSadeghi M, Niknam T, Blaabjerg F (2017) Analysis control and design of speed control of electric vehicles delayed model: Multi-objective fuzzy fractional-order controller[J]. IET Sci Meas Technol 11:249–261
33. Khooban MH, Ghaemi M, Hosseini-Sani SK (2014) Direct adaptive general type-2 fuzzy control for a class of uncertain non-linear systems[J]. IET Sci Meas Technol 8:518–527
34. Perez-Dominguez L, Rodriguez-Picon LA, Alvarado-Iniesta A, Luviano Cruz D, Xu Z (2018) MOORA under Pythagorean fuzzy set for multiple criteria decision making[J]. Complexity. https://doi.org/10.1155/2018/2602376
35. Yazdi M (2017) Hybrid probabilistic risk assessment using fuzzy FTA and fuzzy AHP in a process industry[J]. J Fail Anal Prev 17(4): 756–764
36. Gul M (2018) A review of occupational health and safety risk assessment approaches based on multi-criteria decision-making methods and their fuzzy versions[J]. Hum Ecol Risk Assess Int J 24(7):1723–1760
37. Zeng J, An M, Smith NJ (2007) Application of a fuzzy based decision making methodology to contruction project risk assessment [J]. Int J Proj Manag 25(6):589–600
38. Khooban MH, Liaghat A (2017) A time-varying strategy for urban traffic network control: A fuzzy logic control based on an improved black hole algorithm[J]. Int J Bio-Inspired Comput 10:33–42
39. Garg H (2017) Confidence levels based Pythagorean fuzzy aggregation operators and its application to decision-making process[J]. Comput Math Organ Theory 23(4):546–571

40. Chang LY, Wang GY, Wu Y (1999) A method of attribute reduction and rule extraction based on Rough Set theory [J]. Softw J 10(11): 1–5 (In Chinese)
41. Sun L, Xu J, Cao X (2009) Decision table reduction method based on new conditional entropy for rough set theory[J]. IEEE 5: 1–4
42. Zhou L, Jiang F (2011) A Rough Set Approach to Feature Selection Based on Relative Decision Entropy[J]. Springer 6:110–119
43. Garg H (2016) A novel accuracy function under interval-valued Pythagorean fuzzy environment for solving multicriteria decision making problem[J]. J Intell Fuzzy Syst 31(1):529–540
44. Garg H (2016) A new generalized Pythagorean fuzzy information aggregation using Einstein operations and its application to decision making[J]. Int J Intell Syst 31(9):886–920
45. Garg H (2018) Alinear programming method based on an improved score function for interval-valued pythagorean fuzzy numbers and its application to decision-making[J]. Int J Uncertain FuzzinessKnowl Based Syst 26(01):67–80
46. Gul M, Guneri AF (2016) A fuzzy multi criteria risk assessment based on decision matrix technique: a case study for aluminum industry[J]. J Loss Prev Process Ind 40:89–100

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.