Research on Cloud Health Privacy Information Protection Algorithm Based on Data Mining

Wennan Wang[1(✉)], Shiyang Song[2], Linkai Zhu[3], Junyu Su[4], Te Guo[2], and Jinhai Tang[2]

[1] Academy of Management, Guangdong University of Science and Technology, Dongguan 523083, China
wwwwennan@sina.com

[2] Alibaba Cloud Big Data Application College, Zhuhai College of Science and Technology, Zhuhai 519099, China

[3] Information Technology School, Hebei University of Economics and Business, Shijiazhuang 050061, China

[4] Institute of Data Science, City University of Macau, Macau 999078, China

Abstract. In the process of selecting privacy evaluation indicators, the original algorithm did not set the degree of protection risk, which resulted in a little error in the protection of privacy information, which affected the efficiency of data operation. A data mining-based cloud health privacy information protection algorithm was studied. Build a differential privacy information protection model, establish the connection relationship between users and information, and use trapezoidal distribution membership function to determine the sensitive attributes of privacy information. Set up health privacy information protection indicators on the cloud, evaluate the risks of privacy information protection, and classify them into three levels: high, medium and low. Based on data mining, the information privacy mode is extracted, and the minimum support threshold is used to judge the support of private information in the database. Set the privacy information protection algorithm in the way of data support estimation. Experimental results: in the given data set, the support error of the algorithm in this paper is within 1.5%, and the error of the traditional algorithm is more than 5%. With the increasing size of the data set, the execution time of the traditional algorithm is much higher than that of the algorithm in this paper, which shows that the design is effective.

Keywords: Data mining · Information protection · Health privacy · Protection algorithm · Data set · Support

1 Introduction

Health information is a record of a resident individual's various health conditions in the entire life process, including records of information resources such as diseases and medical visits, and is a long-term dynamic process record. Health information involves residents' diagnosis and treatment information in hospitals, such as various inspection

S. Wang (Ed.): IoTCare 2022, LNICST 501, pp. 20–34, 2023.
https://doi.org/10.1007/978-3-031-33545-7_2

reports, etc., as well as residents' core health records. Health information records contain a large amount of data and are widely used in various fields. Users are most concerned about data security and privacy protection, including data leakage, loss and tampering during data storage, transmission and processing. With the rapid rise of network technology, the continuous development of database technology and the wide application of large-scale database management systems, a large amount of data information has been generated. A wealth of knowledge is hidden in these massive data, and people are eager to obtain useful knowledge from these data [1]. As a powerful data analysis tool, data mining can automatically extract from a large amount of heterogeneous data hidden, unknown and valuable knowledge, and realize the transformation from "data grave" to knowledge wealth, this knowledge is expressed as rules or patterns. They can provide valuable intellectual information for business decision-making, scientific exploration, and medical research.

The knowledge discovered by data mining can not only be used to derive sensitive information from non sensitive information, but also some knowledge discovered by data mining may be sensitive information itself, involving national security, trade secrets and personal privacy. With the continuous progress of information technology, information has gradually evolved into a kind of commodity, which can be collected and stored in database equipment for others to use with or without compensation. This information is easily collected and processed, and finally sold to some intermediary companies and marketing companies [2]. But what follows is that a large amount of personal information becomes more transparent, which makes personal safety feel damaged, and thus brings great harm to personal personality. In the current information age, privacy protection is much more complex than traditional privacy protection. How to prevent the leakage of users' privacy information and reduce the damage and loss of information assets caused by objective or human factors while publishing and applying big data is a widely concerned issue in the field of big data research, which will directly affect the safe application of big data. Based on data mining technology, this paper studies the cloud health privacy information protection algorithm. The overall research technical route of the algorithm is as follows:

(1) First, through the differentiated privacy information protection model, the connection between users and information is analyzed. The membership function is used to calculate the membership of information, and the sensitive attributes of privacy information are obtained.
(2) According to the sensitive attributes of privacy information, build indicators of cloud health privacy information protection, and assess the risk of privacy information protection.
(3) Set the minimum support threshold of privacy information protection, and judge the support of privacy information in the data through data mining methods. Set privacy information protection algorithm in the way of data support estimation.
(4) In the test experiment, the performance of the text algorithm is tested with the support error and operation time as the test indicators.

2 Health Privacy Information Protection Algorithm on the Cloud

2.1 Build a Differential Privacy Information Protection Model

The problem of health information privacy protection includes the user set and the connection relationship between users. In order to protect privacy reasonably and effectively, the data content and social structure are combined to build a differentiated privacy information protection model, that is, based on the social network structure model CVB, the value of users in sensitive attributes is extracted as attribute nodes, and the connection between users and attribute nodes is used, represents the specific attribute value of the user on the sensitive attribute.

The social network in reality is modeled as a simple graph with no authority and no direction, and its mathematical expression is:

$$C = (V, O, B, M) \tag{1}$$

Among them: $V = \{V_1, V_2, \ldots, V_w\}$ represents the user node set. V_q represents the q node, which corresponds to a real user in the network on the cloud. $N \subseteq V \times V$ represents a set of user relationships, any $O_{qp} \in O$ represents a social link between users V_q and V_p, and all relationships are of the same type and are not considered sensitive information.

$B = \{B_1, B_2, \ldots, B_w\}$ represents the set of sensitive attribute value nodes, that is, any $B_q \in B$ represents a specific value of sensitive attribute. For example, in the attribute "disease information", hepatitis and cold are two different attribute values, so two attribute value nodes are formed in the social graph model. M represents the mapping relationship between the sensitive attribute value and the user node. The dotted line indicates that the user has this attribute value, that is, the sensitivity function of the attribute value B_q. The mapping relationship is:

$$M\left(B_q\right) : B_q \rightarrow M_q \tag{2}$$

Therefore, the attribute tag of health user information on the cloud consists of two parts: sensitive attribute value and attribute value sensitivity, namely:

$$e_{V_q} = \left(V_q, B_q, V_q, M_q\right) \tag{3}$$

In the formula: e is the attribute label [3]. e_{V_q} represents the sensitive attribute value of user node V_q and the sensitivity of the attribute value. From this, a privacy information protection model with simple attributes is constructed, as shown in Fig. 1.

Figure 1 shows a simple privacy information protection model, including three user nodes $V = \{V_1, V_2, V_3\}$, three attribute nodes $B = \{B_1, B_2, B_3\}$, and the three social relationships in the network are O_1, O_2, and O_3. The attribute label of user V_1 is $e_{V_1} = (B_1, M_1)$, and the corresponding sensitivities are $e_{V_2} = (B_2, M_2)$, $e_{V_3} = (B_3, M_3)$.

In order to reduce the information loss caused by anonymous algorithm and improve the data utility of publishing graph, considering the sensitivity of attribute values has differences, the attribute value sensitivity function is proposed. By measuring the sensitivity of sensitive attribute values, the sensitive attribute values are divided into three

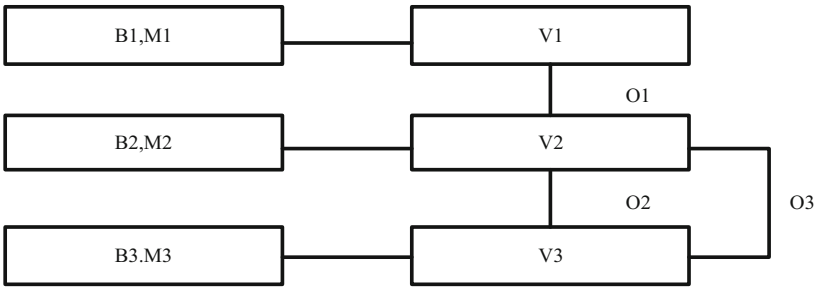

Fig. 1. Privacy information protection model for simple attributes

privacy levels, so that all values under the sensitive attribute category can be differentiated privacy protected.

Different attribute values have different degrees of privacy. When users enter their own health information, for example, the privacy degree of HIV is much greater than that of the common cold. Therefore, the privacy degree of sensitive attribute values is defined as the attribute value sensitivity, and accordingly The user's information privacy level is divided [4]. This time, a trapezoidal distribution membership function is used to determine the sensitivity of the sensitive attribute, and the thresholds r and t. The function value range is [0, 1]. Then exists:

$$\varsigma(u) = \begin{cases} u, u \leq r \\ \frac{u-r}{t-r}, r < u \leq t \\ 1, u > t \end{cases} \tag{4}$$

$$\varsigma \begin{cases} u \leq r, \varsigma = low \\ r < u \leq t, \varsigma = middle \\ u > t, \varsigma = high \end{cases} \tag{5}$$

In the formula: the sensitivity of sensitive attribute is $\varsigma(u)$. Privacy protection thresholds r and t are two constants, which reflect the extent to which the respondent can accept privacy disclosure. Their size is set by the data publisher according to different application backgrounds or the privacy protection requirements of the respondent. By default, $r = 0.5, t = 1$. In addition, data publishers can set privacy protection thresholds in real time according to the dynamic status of the network environment. The lower the threshold, the higher the demand for privacy protection.

The user's disease information "HIV" is not sensitive information in the AIDS friendly communication network, but is high-level sensitive information in ordinary social groups. Accordingly, users' information privacy levels are divided into three categories: high, middle, and low. There are certain differences in the protection forms of the three information privacy levels, which are processed through the principle of anonymity protection. The process is shown in Fig. 2 below.

In the above figure, the publisher initiates a publishing request in the network service to request the information of users in the database to be published. The protection model obtains the information of the publisher from the database, and combines the privacy

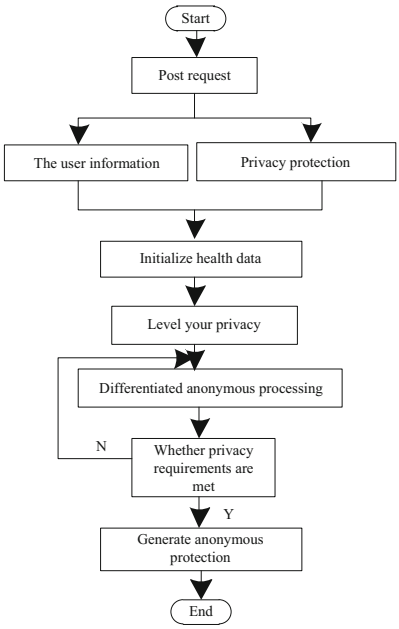

Fig. 2. Anonymous processing process of user health privacy information

protection requirements set by the publisher in the account to obtain the set of privacy information and the corresponding privacy protection requirements.

Generally, a simple form of anonymity is used to replace the user's unique identity with serialized symbols, build a social graph, obtain the sensitivity of each attribute value in the privacy information according to the sensitivity function, divide the privacy level, store the level in the user's attribute tag, and adopt the corresponding anonymity strategy. However, how to choose anonymous processing criteria requires selecting specific indicators in the protection model.

2.2 Setting up Health Privacy Information Protection Indicators on the Cloud

In order to ensure that the health information on the cloud is in a safe state, a risk assessment method is used to protect the privacy information, and a risk early warning protection index is proposed. Qualitative or quantitative evaluation of possible risk events, and privacy risk situation assessment based on the release of health information, can find out the main risk factors of privacy leakage, the state of each risk factor and the development trend, which is the design of health information release. Prerequisites for privacy-preserving algorithms.

With reference to ITSEC and in combination with the characteristics of health information release and its application mode, privacy assets, privacy threats and privacy vulnerabilities are combined to form the possible sources of privacy risk. The three are

taken as the first level elements for evaluating privacy risk, and the second level evaluation elements and third level privacy risk evaluation factors are selected for them to design a privacy information protection system [5]. See Table 1 for details.

Table 1. Health privacy information protection system on the cloud

First-level indicator	Secondary indicators	Three-level indicator
Privacy assets	Degree of confidentiality	Data encryption
		Quarantine data
		Manage keys
		Data confidentiality
	Completeness	Backup data
		Destroy data
		Upgrade software
	Available value	Migrate data
		Identify risks
Privacy threat	Technical risk	Deliberate attack
		Network control
		Vulnerability Handling
	Human risk	Staff member
		Verify identidy
		Wrong operation
Privacy vulnerability	Organizational vulnerability	Approval system
		Network system
		Liability Regulations
	Technical vulnerability	Service Content
		Degree of access
	Other vulnerability	Regulations
		Privacy treatment
		Risk report

On the basis of the above indicators, through the research and analysis of a large number of privacy leaks and authoritative reports at home and abroad, and referring to the risk assessment indicators of the world's most authoritative IT research and consulting company Gartner, European network and information security agency ENISA, cloud computing security alliance CSA and other world authoritative organizations, sort out and summarize the privacy risk assessment indicators.

On the basis of fuzzy mathematics, Vague set theory, and intuitionistic fuzzy theory, a method for analyzing and processing uncertain information is developed. It uses the

connection number to reflect the fuzzy, certain, uncertain and other phenomena between things and the changes between them. With the help of ternary or multivariate connection number, the various factors affecting privacy risk can be concentrated, effectively classified and reasonably described. Among them, factors that have a positive and obvious impact on system privacy risk can be expressed as support; Factors unrelated to privacy risk can be expressed as opposition; Other factors with uncertain impact on privacy risk can be expressed as a concentration is the uncertainty, or is further refined into the partial identical component, the neutral component or the partial inverse component of the uncertainty.

From the perspective of privacy risk assessment, "homeopathy" indicates that the result of privacy risk assessment tends to be in the same trend state with the ideal standard risk, that is, it is in "low risk"; "Balance of power" reflects that the privacy risk assessment results and the ideal standard risk are close to each other, that is, they are at "medium risk"; "Counter trend" indicates that the result of privacy risk assessment tends to be opposite to the ideal standard risk, that is, it is at "high risk". The results of privacy risk assessment can be simply "clustered" through potential value. The risk indicators of the above privacy protection are evaluated, and their levels are shown in Table 2.

Table 2. Health and privacy protection index levels on the cloud

Index	Influencing factors	Privacy assets	Privacy threat	Privacy vulnerability
Degree of confidentiality	Data encryption	High risk	Medium risk	High risk
	Quarantine data	Low risk	Low risk	High risk
	Manage keys	High risk	High risk	High risk
	Data confidentiality	Medium risk	High risk	High risk
Completeness	Backup data	High risk	High risk	Medium risk
	Destroy data	Low risk	Medium risk	High risk
	Upgrade software	High risk	High risk	Low risk
Available value	Migrate data	High risk	Low risk	High risk
	Identify risks	High risk	High risk	Low risk
Technical risk	Deliberate attack	High risk	High risk	High risk
	Network control	High risk	Low risk	High risk
	Vulnerability Handling	Low risk	High risk	Medium risk
Human risk	Staff member	Low risk	Low risk	High risk
	Verify identidy	Medium risk	Low risk	High risk

(*continued*)

Table 2. (*continued*)

Index	Influencing factors	Privacy assets	Privacy threat	Privacy vulnerability
	Wrong operation	High risk	Low risk	Medium risk
Organizational vulnerability	Approval system	High risk	High risk	Low risk
	Network system	Low risk	Low risk	Low risk
	Liability Regulations	Medium risk	High risk	High risk
Technical vulnerability	Service Content	High risk	High risk	High risk
	Degree of access	High risk	High risk	Low risk
Other vulnerability	Regulations	Low risk	Medium risk	Medium risk
	Privacy treatment	High risk	High risk	Low risk
	Risk report	High risk	High risk	Low risk

With the help of many experts, we will evaluate the level of risk factors in the current privacy in the process of health privacy data release, so as to complete the construction of privacy risk assessment index system. According to the principle of maximum membership, the results are divided into corresponding evaluation levels, corresponding to different risk levels, and information privacy patterns are extracted based on data mining.

2.3 Extracting Information Privacy Patterns Based on Data Mining

Mining itself is also an effective means of knowledge expression. The goal is to retain as many non sensitive patterns as possible in the result data set without disclosing sensitive patterns, especially some non sensitive patterns containing important information, so as to improve the availability of the result data set, that is, frequent pattern mining can be used to find non sensitive patterns containing important information. Figure 3 shows the process of processing the original data set to get the result data set.

The original dataset contains all frequent patterns, while the resulting dataset contains only non-sensitive patterns. First, on the original data set, frequent pattern mining is performed to obtain the corresponding set of frequent patterns. Through analysis, the data owner determines which of the frequent patterns obtained by mining contain sensitive information, which are called sensitive patterns or private patterns [6]. Then, according to the frequent patterns obtained by mining, some of which have been identified as sensitive patterns by the data owner, the original data set is processed by applying the privacy protection algorithm to obtain a new result data set.

Mining schemas typically target datasets stored in transactional databases, also known as transactional databases. Health privacy information on the cloud is a typical transaction database. A transactional database is usually described as follows: Let $S = \{S_1, S_2, ..., S_N\}$ be a set of N items [7]. A transaction record D, also called a transaction record, is a subset of the items in S. Table 3 is a simple transaction database.

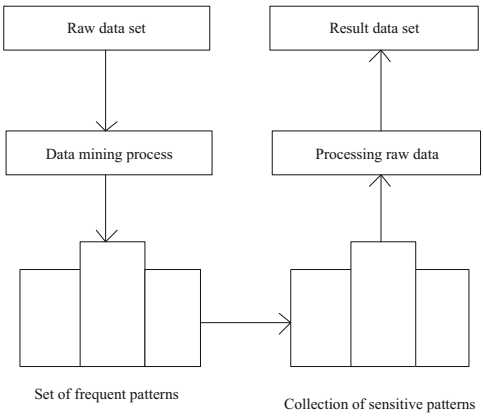

Fig. 3. Schematic diagram of the privacy protection process in frequent pattern mining

Table 3. Simple transaction database

Health information records	Item in info record
S1	z, x, c
S2	v, z, b, x
S3	v, z, b, x
S4	z, b, x, n
S5	v, z, b, m

In the table above, each transaction record has a unique identifier. A transaction database is a collection of transaction records. In this way, the database schema is defined, and the schema is called an item set, which is set as F, which is a subset composed of items in S, and F is included in transaction record D. If F contains G items, then the length of F is called G, denoted as:

$$|F| = G \tag{6}$$

Therefore, for convenience, we abbreviate the mode like $F = \{c, n, m\}$ as:

$$F = cnm \tag{7}$$

When there are two modes, rewrite them as F and H. If F is a subset of H, then F is said to be a sub-pattern of H. Correspondingly, H is said to be a supermode of F. Obviously, for a pattern of length 1, its subpatterns do not exist. And when there are two modes J and K, if J is a sub-mode of K, and there is no mode L, such that J is a sub-mode of L and L is a sub-mode of K, then J can be said to be a direct sub-mode of K, K is the direct supermode of J.

On this basis, to obtain the sensitivity support, a transaction record D supports mode F if and only if the transaction record D contains mode F [8]. If there are N multiple

transaction records in transaction database S, and the number of transaction records including F is Z, then the absolute support degree of mode F on S is Z, which is denoted as $ZF_S(F)$. The relative support, referred to as support, is Z/N, and denoted as $Z \cup F_S(F)$.

When the private information database S is clear in the context, we also abbreviated the support degree $Z \cup F_S(F)$ and support number $ZF_S(F)$ of pattern F as $Z \cup F(F)$ and $ZF(F)$, respectively. Therefore, given the transaction database and the minimum support threshold, set it to λ, if the support $Z \cup F(F)$ of pattern F on S is greater than or equal to the minimum support threshold λ, then F is a frequent pattern on S.

Set the minimum support threshold λ to the value of 50%. According to the definition of frequent patterns, the support of all frequent patterns in the transaction database in Table 3 can be given, as shown in Table 4.

Table 4. Frequent mining patterns of single-transaction databases

Pattern length	Frequent pattern
1	(v,50%), (z,100%), (b,70%), (x,70%)
2	(vz,50%), (vb,50%), (vx,30%), (zb,70%) (zx,70%), (bx,50%)
3	(vzb,50%), (vzx,50%), (vbx,30%), (zbx,70%)
4	(vzbx,30%)

The task of frequent pattern mining is to find all frequent patterns whose support is not less than λ and their corresponding support when the minimum support threshold λ is given. Given a transaction database and a minimum support threshold λ, let $X(S, Z \cup F_S(F))$ denote the frequent pattern mining results on S, then:

$$X(S, Z \cup F_S(F)) = \{p, Z \cup F_S(F) | Z \cup F_S(F) \geq \lambda\} \tag{8}$$

In the formula: for convenience, when the transaction database S and the minimum support λ are clear in the context, $X(S, \lambda)$ is also simply marked as X [4]. Therefore, through the frequent patterns in data mining, the privacy pattern of health information on the cloud is given, and with the help of data support, the algorithm of privacy protection of data information is designed.

2.4 Data Support Setting Privacy Information Protection Algorithm

Suppose a database tuple is composed of Q and W, Q indicates that the attribute appears, and W indicates that the attribute does not appear. The probability of each data item remaining at the original value is E, and the probability of flipping is $Q - E$. All database tuples are distorted in the same way to form a new database. Data mining is performed

on databases formed after distortion. For different properties, different values of E can be used to distort. For simplicity, all probabilities E are assumed to be equal in this paper.

Let the real data set matrix be represented as R, the matrix obtained by R after the distortion operation is T, and the distortion probability is E. The number of Q in column U of R is recorded as I_Q^R, the number of W is recorded as I_W^R, the number of Q in column U of T is recorded as I_Q^T, and the number of W is recorded as I_W^T. It can be seen from the data distortion process that:

$$\begin{cases} I_Q^R \times E + I_W^R \times (1-E) = I_Q^T \\ I_W^R \times E + I_Q^R \times (1-E) = I_W^T \end{cases} \tag{9}$$

This leads to:

$$I^R = P^{-1} I^T \tag{10}$$

Of which:

$$P = \begin{bmatrix} E & 1-E \\ 1-E & E \end{bmatrix} \tag{11}$$

$$I^T = \begin{bmatrix} I_Q^T \\ I_W^T \end{bmatrix} \tag{12}$$

$$I^R = \begin{bmatrix} I_Q^R \\ I_W^R \end{bmatrix} \tag{13}$$

In the formula: P is a matrix of order N. From Eq. (9), the true matrix one-item set support I_Q^R can be estimated from the distorted matrix T.

The calculation method of the support of N item set is similar to that of item set. at this time:

$$P = \begin{bmatrix} p_{0,0} & p_{0,1} & \cdots & p_{0,2^N-1} \\ p_{1,0} & p_{1,1} & \cdots & p_{1,2^N-1} \\ \cdots & \cdots & \cdots & \ddots \\ p_{2^N-1,0} & p_{2^N-1,1} & \cdots & p_{2^N-1,2^N-1} \end{bmatrix} \tag{14}$$

$$I^T = \begin{bmatrix} I_{2^N-1}^T \\ \vdots \\ I_Q^T \\ I_W^T \end{bmatrix} \tag{15}$$

$$I^R = \begin{bmatrix} I_{2^N-1}^R \\ \vdots \\ I_Q^R \\ I_W^R \end{bmatrix} \tag{16}$$

In the formula: I_A^T is defined as the number of N itemsets in the distorted database, the itemsets are in the form of N-bit binary numbers, and the decimal value A [9] corresponding to the N-bit binary numbers. $p_{i,j}$ is the probability that N itemset j (represented as N-bit binary) is distorted into N itemset i (represented as N-bit binary). The binomial set $p_{0,1}$ represents the probability that the itemset 11 is distorted to 00, which is $(1-E)^2$. Then the support degree of the N itemset in the original data set is $I_{2^{N-1}}^R$. So far, the design of the cloud health privacy information protection algorithm based on data mining is completed.

3 Experimental Studies

3.1 Experiment Preparation

In the above, based on data mining technology, a new algorithm is designed for the privacy protection of health information on the cloud. In order to verify the effectiveness of this method, experimental testing method is used to demonstrate. Taking the noise protection algorithm as the control group, the effectiveness and efficiency of different algorithms are verified by comparing with the method in this paper. In the experimental test, first, given the minimum support threshold, in 50 groups of original data sets, obtain all the health information of users, randomly select 20 sensitive data sets, and take the remaining non sensitive data sets as the release data sets.

The first part uses the support error index to measure the effectiveness of the two algorithms. The larger the support error value, the greater the impact of the algorithm on the result data set, the worse the processing of sensitive data in health information, and it is difficult to complete the precision. Data protection.

Define the calculation method of the support error degree, and assume that there are $f = [f_1, f_2, ..., f_g]$ modes in the published data set, then:

$$k = \sum_{i=1}^{g} \frac{h(f)' - h(f)}{h(f)} \tag{17}$$

In the formula: k is the support error degree. $h(f)$ represents the support of different modes in the result data set, while $h(f)'$ represents the support of different modes in the original data set.

In the second part of the test, the privacy protection threshold is taken as a constant, and the size of the original data set is gradually increased, so as to compare the efficiency of the two algorithms.

3.2 Results and Analysis

In the first part of the test, the number of transactions in the original data set is 200,000, the original data set remains unchanged throughout the process, the minimum support threshold is set to 4%, and the privacy protection threshold is gradually increased from 5% to 40%. In this case, the effectiveness of the two algorithms is compared, and the results are shown in Fig. 4.

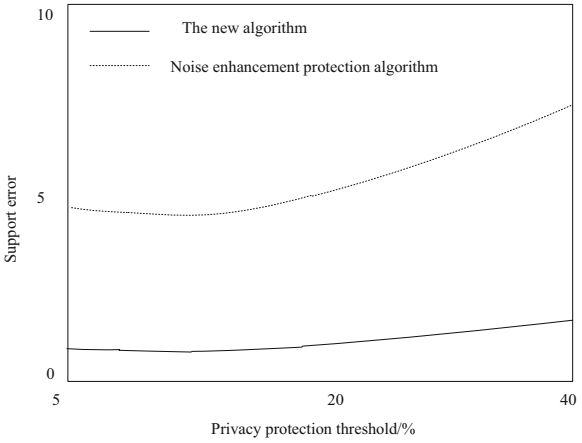

Fig. 4. Compares the support error under different privacy protection thresholds

It can be seen from the figure that a small privacy protection threshold means a higher degree of protection for sensitive modes. For the noise protection algorithm, as the privacy protection threshold increases, more sensitive transactions are processed, so the support error becomes higher. However, the variation range of this method becomes smaller, and the error degree is basically kept within 1.5, which has application effect.

In the second part of the test, keep the privacy protection threshold unchanged, set it to 30%, and the minimum support threshold is 2.0%, gradually increase the size of the original data set, and compare the efficiency of the two algorithms. The results are shown in Fig. 5.

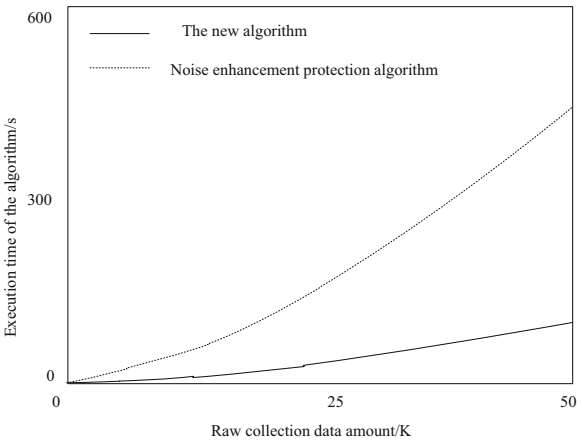

Fig. 5. Compares the execution time of the algorithm under different original data sets

In the figure, when the original data set is the same, the execution time of the algorithm in this paper is less than the execution time of increasing the noise privacy protection

algorithm, and with the continuous increase of the original data set, increasing the execution time of the noise protection algorithm will increase the speed of its growth. It is much larger than the algorithm in this paper, indicating that the method designed this time is more effective.

4 Conclusion

With the development of network information technology and database technology, data mining technology shows an increasingly broad application prospect and plays an increasingly important role in more and more fields. However, as a large number of private data are widely collected and analyzed, the application of data containing sensitive information will pose a threat to personal privacy security. Therefore, this paper designs a new protection algorithm on the basis of data mining technology on how to deal with the privacy protection problem.

On the one hand, it deals with the privacy information contained in the health dataset itself, and on the other hand, it deals with the sensitive knowledge generated after the application of the dataset. The experimental results show that the error degree of this method is significantly reduced, and the error degree of this method is basically kept within 1.5; The execution time of the algorithm in this paper is significantly reduced, and the maximum is only about 150 s.

However, due to the limitation of time and energy, the research results that have been achieved still need to be further improved, and can be further expanded in the process of combining with relevant research directions and fields. Specifically, an algorithm will have more vitality only if it is applied in real life. It will be our important work in the future to apply our algorithm to realistic datasets and solve problems that arise in reality.

Acknowledgment. The authors are grateful to the support of the Guangdong Universities' Innovation Team Project (2021KCXTD015) and Guangdong Key Disciplines Project (2021ZDJS138).

References

1. Liu, Y., Yang, Y.: A Data mining algorithm for matrix and sort index association rules. Comput. Technol. Dev. **31**(2), 54–59 (2021)
2. Jia, X.: Time series data mining algorithm based on multiobjective decision. Electron. Design Eng. **29**(17), 45–49 (2021)
3. Yao, R., Fei, Y., Ding, Y., et al.: Research on engineering data information prediction model based on intelligent data mining algorithm. Electron. Design Eng. **30**(7), 63–67 (2022)
4. Lu, P., Ge, S., Wu, X., et al.: Design of personal health information management and privacy protection system based on ZigBee. Comput. Meas. Control **29**(4), 170–174 (2021)
5. Wei, W., Xin, Z., Wang, S., Zhang, Y.: Covid-19 diagnosis by WE-SAJ. Syst. Sci. Control Eng. **10**(1), 325–335 (2022)
6. Tang, C., Lin, X.: Protocol of privacy-preserving set intersection computation. Netinfo Secur. **1**, 9–15 (2020)
7. Huang, C., Wang, W., Zhang, X., Wang, S. -H., Zhang, Y.-D.: tuberculosis diagnosis using deep transferred EfficientNet. IEEE/ACM Trans. Comput. Biol. Bioinf. (2022)

8. Song, X., Gai, M., Zhao, S., et al.: Privacy-preserving statistics protocol for set-based computation. J. Comput. Res. Dev. **57**(10), 2221–2231 (2020)
9. Zhang, G., Tang, Z., Li, S., She, F.: Measurement and simulation of risk tolerance of medical privacy big data disclosure. Comput. Simul. **38**(12), 480–484 (2021)
10. Wang, J.: Back-propagation neural network learning algorithm based on privacy preserving. Comput. Sci. **49**(z1), 575–580 (2022)