Research on Distributed Data Stream Mining of Financial Risk Based on Double Privacy Protection

Yuhao Zhao

School of Business, Stevens Institute of Technology, Castle Point on Hudson, Hoboken, NJ, U.S. 07030

Email: jiwusuochu041172@163.com

ABSTRACT：With the advancement of network technology and large-scale computing, distributed data streams have been widely used in the application of financial risk analysis. However, while data mining reveals financial models, it also increasingly poses a threat to privacy. Therefore, how to prevent privacy leakage during the efficient mining process poses new challenges to the data mining technology. This article is mainly aimed at the current privacy data leakage in financial data mining, combined with existing data mining technology to study data mining and privacy protection. First, a data mining model for dual privacy protection is defined, which can better meet the characteristics of distributed data streams while achieving privacy protection effects. Secondly, a privacy-oriented data stream mining algorithm is proposed, which uses random interference technology to effectively protect the original sensitive data. Finally, the analysis and discussion of the algorithm in this paper through simulation experiments show that the algorithm is feasible and effective, and can better adapt to the distributed data flow distribution and dynamic characteristics, while achieving better privacy protection effects, effectively Reduced communication load.

INDEX TERMS：Financial Risk, Double Privacy Protection, Distributed Data, Random Interference

I. INTRODUCTION

In recent years, the rapid development of information processing technology and storage technology has enabled relevant financial institutions to collect large amounts of data for data mining. In the process of data mining, multiple financial data owners may be required to publish or share the data they own [1-2]. However, directly publishing and sharing the original financial data will lead to the disclosure of personal privacy information. In this case, the data owner is in a dilemma [3]. On the one hand, the privacy of personal data needs to be protected, and on the other hand, data availability needs to be guaranteed for data mining tasks. In order to solve this problem, the publication of privacy-protected data for data mining came into being and has become a very active research field [4-5].

This field mainly studies how to publish data that does not disclose private information, while ensuring that the published data can be used for data mining. At present, a lot of research work has emerged for different data types, different application scenarios, and different attack models. Agrawal et al. first proposed an encrypted data perturbation method for building decision tree classifiers. This method assumes that each client has a numeric attribute, and the data mining service needs to learn the distribution of these attribute values to build a classifier model [6]. Argueta et al. questioned the availability of encryption random noise, and pointed out that addition noise can be filtered out in many cases, thus leading to privacy disclosure [7]. For more specific verification, a random matrix-based spectral filtering method is proposed to reconstruct the original data from the perturbed data. Experimental results show that the reconstructed data is very close to the original data. Huang et al. pointed out that the main factor that determines the accuracy of reconstructed data is the correlation between the attributes of the original data [8]. The research results show that the greater the correlation between the attributes of the original data, the higher the accuracy of the reconstructed data, so it will lead to the leakage of more private information. They further proposed two reconstruction methods based on data correlation, using principal component analysis technology and Bayesian estimation technology, respectively. Among them, the method based on data perturbation is easy to implement and has a strict mathematical theory foundation, which has strong practicability and reliability. However, there are still some problems to be solved [9-10].

This paper takes financial risk as the research object, and uses distributed privacy protection and random interference to design distributed data stream mining technology. By accessing the financial risk information platform, personal privacy data can be dynamically and accurately transmitted to the database, and real-time tracking and analysis of information can be achieved. This technology can improve the information transparency and data security of financial risk data mining, and provide an effective and implementable solution for promoting the development of financial undertakings.

II. FINANCIAL RISK

The spillover effect of financial risks is inseparable from the study of individual specific risks. A single specific risk is the basis and source of financial risk spillover effects [11]. Therefore, when studying the spillover effect of financial risks, we should first study the specific financial risks. To study the risk spillover effect between the banking industry and the securities industry, first of all, the risks faced by the banking industry and the securities industry should be studied. On this basis, the mechanism of the risk spillover effect between the banking industry and the securities industry is elaborated [12]. This article will make a further theoretical analysis of the risk theory and the spillover effect between the banking industry and the securities industry. Regarding the causes of financial risks, many theories give their explanations. Schematic diagram of financial risk concept is shown in Figure 1.

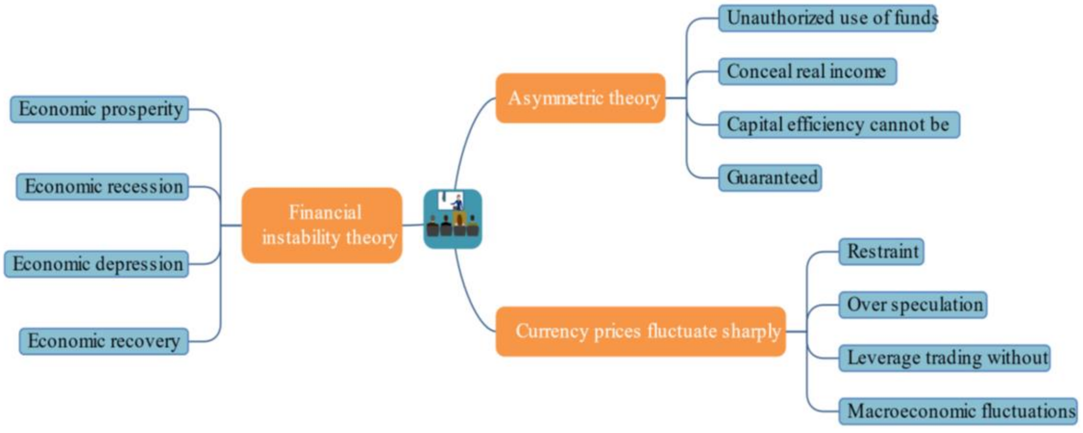


Figure 1 Schematic diagram of financial risk concept

A FINANCIAL INSTABILITY THEORY

The main explanations for the theory of financial instability include the "cyclic explanation" group led by American economists Hyman Minsky and Charles Kinder Berger and the "monetarist explanation" group led by Friedman. The former believes that finance is an important part of various economies, so with the prosperity, recession, depression and recovery rhythm of the economic cycle, there will also be periodic instability in finance, namely financial risk [13]. This is consistent with the interpretation of the New Palgrave Dictionary of Monetary and Financial Affairs. According to the interpretation of the dictionary, the financial instability hypothesis refers to the inherent tendency of private credit creation institutions, especially commercial banks and related lenders, to experience cyclical crises and bankruptcy [14]. Because of the reality of its social life, that is, its close relationship with the real economy, finance has planted the seeds of financial instability during the period of economic and social prosperity.

B ASYMMETRIC INFORMATION THEORY

Information economics scholars believe that the information in the economy is incomplete, that is, asymmetric, and one party of the economic actor knows more information than the other party. Moral hazard and adverse selection are the products of information asymmetry [15-16]. Moral risk refers to the information asymmetry after the transaction occurs. There are three main manifestations of moral risk in the loan market: First, unauthorized changes in the use of funds. The second is that some borrowers with repayment ability conceal their true income status and delay or refrain from returning the loan. This situation often occurs in the absence of sanctions for breach of contract [17]. The third is that after the borrower obtains the funds, he does not pay attention to the use of the funds, so that the use efficiency of the funds cannot be guaranteed, resulting in losses of the borrowed funds.

C. THE THEORY OF SHARP FLUCTUATIONS IN ASSET PRICES

Financial risk directly manifests as the loss of currency, so dramatic fluctuations in asset prices can directly cause instability in the financial system and trigger financial risks. As we all know, the price of financial assets always fluctuates from one peak to another, and this fluctuation is an important source of financial risk. The causes of financial asset price fluctuations are mainly concentrated in the following three aspects: first, there is excessive speculation; second, the uncontrolled use of credit and leveraged transactions; third, macroeconomic fluctuations.

D INTERNATIONAL COMMUNICATION THEORY OF FINANCIAL RISK

The globalization of international capital, the globalization of financial institutions and financial markets not only bring us a more dynamic economic structure, but also bring us a more complex financial risk environment. Financial globalization provides a free-flowing capital carrier for the international spread of finance. In particular, some developing countries have relaxed foreign exchange control in order to meet the needs of their own economic development and the integration of the world economy [18]. At the same time, however, the corresponding financial supervision system has not been formed, which has led to the larger-scale capital transfer becoming a carrier for the spread of financial risks internationally.

III. RELATED WORK

Before delving into the details of our proposed method, we first reexamine the existing methods, such as double privacy protection and random interference.

A. DOUBLE PRIVACY PROTECTION

In collaborative data mining, there are two forms of data distribution, one is horizontal division and the other is vertical division. Horizontally divided data refers to the fact that the data held by each participant contains the same instance, but the data attributes contained in the instance are different. Vertically dividing data means that the data held by each participant contains different instances, but the data attributes are the same. Figure 2 depicts the data perturbation during data level division [19-20]. Suppose A and B are two participants in collaborative data mining. Since the data records are horizontally distributed in two places, A and B, in order to protect the privacy of their data, they need to randomly disturb the data they own before the data is released [21].

Therefore, in order to ensure the uniformity of the projection results, A and B need to use the same random matrix for projection disturbance. In this way, the risk of random matrix leakage is faced. In order to ensure that the original data is not reconstructed, this paper proposes a noise random projection data perturbation method that meets differential privacy protection [22].

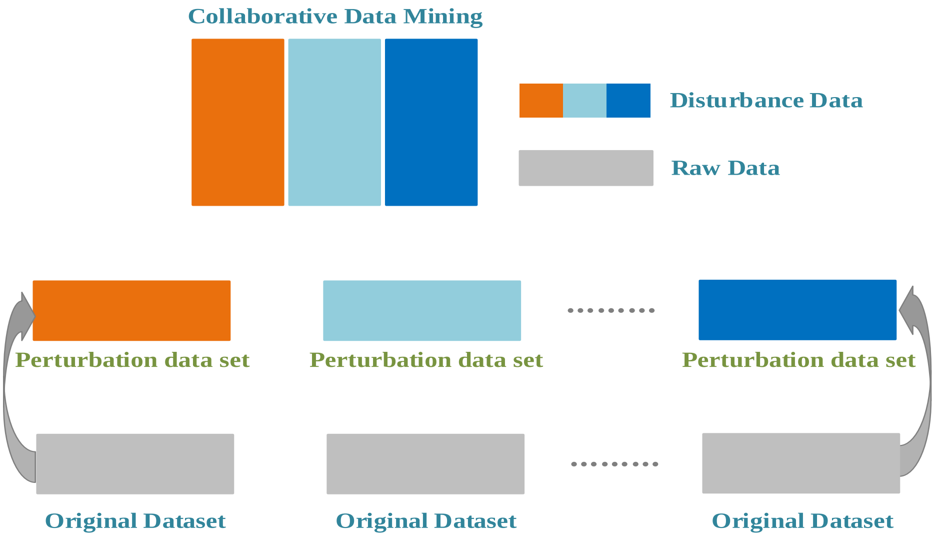


Figure 2 Aiming at the horizontal division of privacy protection data sharing

A key issue in the design of any privacy protection method is that the attacker will use prior knowledge of a certain amount of data to infer privacy protection content. However, when designing a privacy protection algorithm, it is impossible to accurately predict the amount of a priori knowledge that an attacker can grasp, so no strong proof can be given for an unknowable attack. Differential privacy provides a strong, provable and strict mathematical guarantee for privacy protection algorithm protection [23].

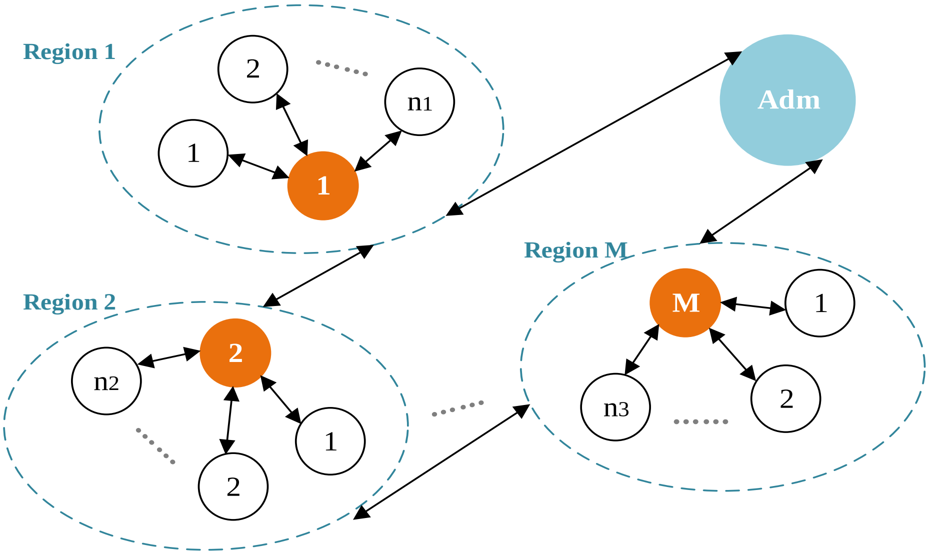


Figure 3 The privacy protection node model for distributed data mining

Aiming at the distributed data flow environment, a system model is constructed in which multiple nodes publish privacy protection data in real time, and integrate data from each node for data mining [24]. Figure 3 shows the privacy protection node model for distributed data mining. As shown in Figure. 3, the aggregation node is responsible for collecting the data of the subordinate terminal nodes and then transmitting them to the management service node Adm. Node Adm needs to classify, cluster, and anomaly detect data. In order to protect the privacy of the original data, the original private data content of each terminal node is not exposed during the data transmission process or reaching the management service node. Suppose that when the aggregation node m broadcasts a data collection instruction to the terminal nodes in its area, n terminal nodes respond [25]. Then next m will collect the data transmitted by the n terminal nodes. Data sent by n terminal nodes can be regarded as n asynchronously updated data streams. Before transmitting data to the aggregation node, the terminal node first disturbs the data.

On the one hand, only the disturbed data is obtained, and the attacker cannot determine the original data. On the other hand, the disturbed data still maintains the measured distance between the original data, and the data mining algorithm can be directly applied to the disturbed data without accessing the original sensitive information.

B. RANDOM INTERFERENCE TECHNOLOGY

The symmetrical difference between the two data sets X, X' are defined as

$$X \ominus X' = \{x | x \in X \cup X' - X \cap X'\} \tag{1}$$

If you count for any two data sets $X, X \in D'$, and any $\hat{D} \subseteq Range(A)$is satisfied

$$Pr[A(X) \in \hat{D}] \leq e^{[X, X']} Pr[A(X') \in \hat{D}] \tag{2}$$

Then the algorithm satisfies differential privacy [26].

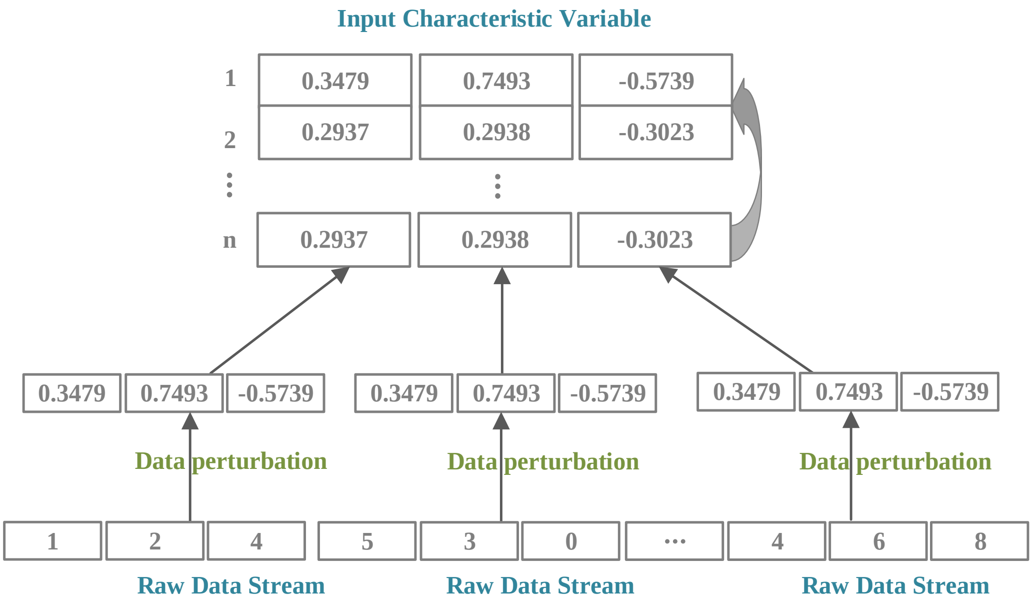


Figure 4 Schematic diagram of the disturbance data released by the terminal

The point cloud models are expressed as P_1 and P_2 respectively, and the corresponding number of point clouds is counted. Calculate the positions of the global and local point clouds P_1 and P_2 respectively projected in the financial distributed data stream [27].

Suppose that you have obtained the point pairs P_{k1} and P_{k2} in the point cloud models P_1 and P_2. By finding an approximate transformation T, the coordinate alignment of all corresponding point sets in P_{k1} and P_{k2} is completed. The transformation T is composed of a translation vector L, a rotation matrix R and a scaling factor s, so the transformation objective function problem of the transformation T problem is solved.

$$E(P_{k1},P_{k2})=\|P_{k1}-(s\cdot P_{k1}\cdot R+L)\| \tag{3}$$

Analytical formula (3) shows that the optimization problem of the function is the Procrustes problem. Use the following steps to gradually solve the corresponding transformation, and use formula (4) to calculate the geometric centers of P_{k1} and P_{k2}.

$$O=\sum_{i=1}^{N} X_i/N \tag{4}$$

In which, O represents the geometric center. X_i represents the coordinate of the ith point. N represents the number of point clouds. After the geometric centers O_1 and O_2 of P_{k1} and P_{k2} are obtained, the translation vector L is calculated using equation (5).

$$L=(O_1\text{-}O_2)+(O\text{-}O_1) \tag{5}$$

The coordinates of P_{k1} and P_{k2} are normalized, without considering the scaling factor, and expressed by equation (6).

$$E(P_{k1},P_{k2})=\|P_{k1}\text{-}P_{k2}\cdot R\| \tag{6}$$

$$R=U\cdot V^T \tag{7}$$

In order to meet the invariance of the internal structure during the conversion process, the rotation matrix R needs to satisfy det (R) = 1, and the minimization problem is solved using Procrustes orthogonal problem theory the optimal rotation matrix R is solved by formula (8).

$$s=tr\left(P_{K1}^T\cdot P_{k2}\cdot R\right)/tr\left(P_{K2}^T\cdot P_{k2}\right) \tag{8}$$

In which, U and V^T are the eigenvector matrix results obtained by singular value decomposition, and the scaling factor is calculated using equation (9).

$$\widetilde{P}_2=s\cdot P_2\cdot R+L \tag{9}$$

Figure 4 is a schematic diagram of disturbance data released by financial risk distributed terminals. First determine the projection dimension, then perturb each updated attribute value, and finally calculate the perturbation data. The figure shows an example of data mining performed by n terminals issuing disturbance data. Each terminal disturbs its original data stream and sends it to the base station. The base station summarizes the disturbance data coming from each terminal and integrates it into an aggregate data set for subsequent data mining work [28].

IV METHODS

A. TEST ENVIRONMENT AND DATA

The experiment selects artificial data and real data separately for testing. The real data contains two data sets. One is the financial risk distributed data set, which contains some financial risk data obtained from a financial platform. Use the data structure commonly used in text data mining to represent the data into a document matrix. In the experiment, the financial risk data set with 200 attributes after feature selection is selected [29]. Another real data set is a financial risk data set published by a financial analysis agency. By selecting the frequency, a sparse integer type data set is generated, and the data record has an attribute dimension of 1000. In addition, the experiment generated an artificial data set with a data attribute dimension of 100 and sparsity from 1 to 100. The experimental environment is: the hardware environment is Intel(R) Core (TM) i7 CPU @2.40GHz, the installed memory is 4GB; the software environment is Windows 10 operating system, and the algorithms are implemented in Python.

B. METHODS

Data sparsity and reconstruction data quality experiments first selected a data record with the most non-zero elements from a financial platform and a financial analysis institution data set [30].

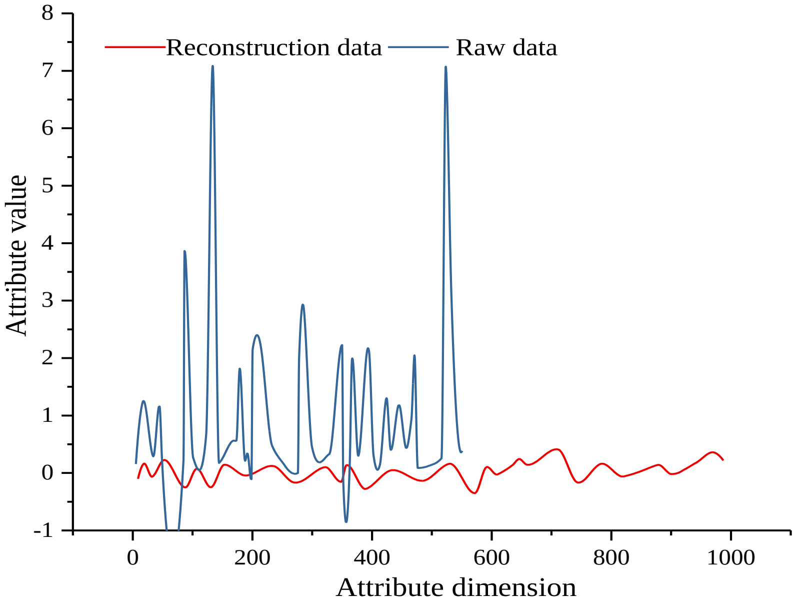


Figure 5 Comparison of original data and reconstructed data in a financial platform data set

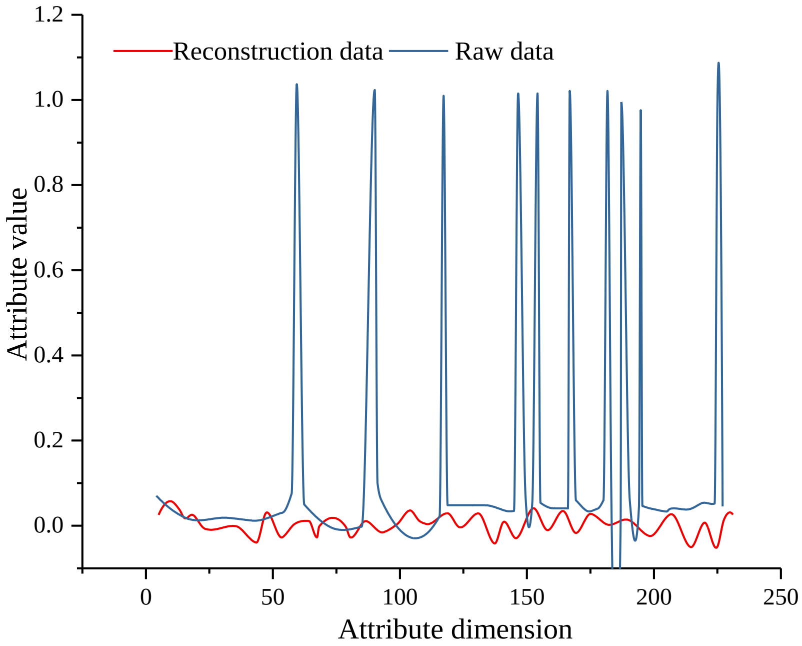
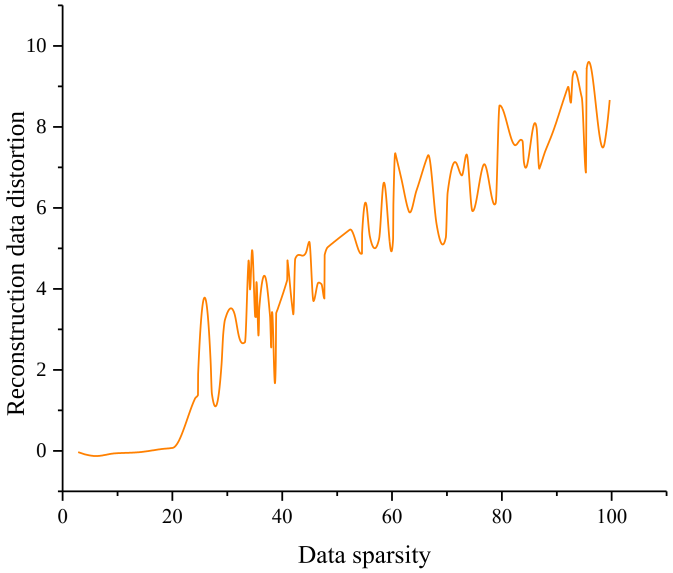


Figure 6 Comparison of original data and reconstructed data in a financial analysis institution's data set

Among them, the data selected from a financial platform contains 10 non-zero items with a data sparsity of 4.37%, and the data selected from a financial analysis institution contains 82 non-zero items with a data sparsity of 8.2%. First, the random projection data perturbation method is used to perturb the data to generate perturbation data with 50% projection ratio, and then the original data is reconstructed from the perturbation data. Figures 5 and 6 show the comparison between the original data and the reconstructed data in the data set of a financial platform and a financial analysis institution. It can be seen that the original data records are accurately reconstructed.

C. RESULTS

The following will specifically analyze the relationship between the sparsity of the original financial data and the quality of the reconstructed data, and the relationship between the projection dimension of the disturbance data and the quality of the reconstructed data. In the test of the sparsity of the original financial data and the quality of the reconstructed data, an artificial data set containing data records of different sparsity was generated. The original record dimension in the data set is 100, the projection disturbance dimension is 50, and the data types include continuous type, Boolean type, and discrete integer type. Among them, the unknown original financial data type is first tested, and the quality of the original data is approximately reconstructed, and then the quality of the original data is accurately reconstructed when the original data type is known. In the experiment, the quality of the original data is evaluated by the degree of distortion of the reconstructed data.



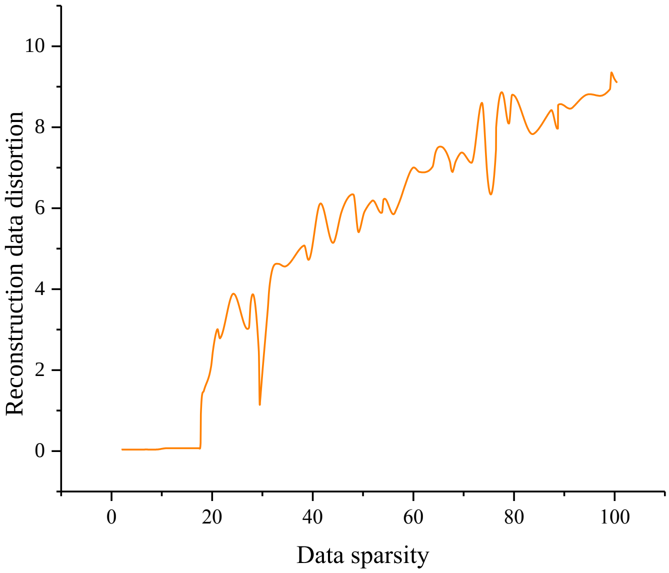
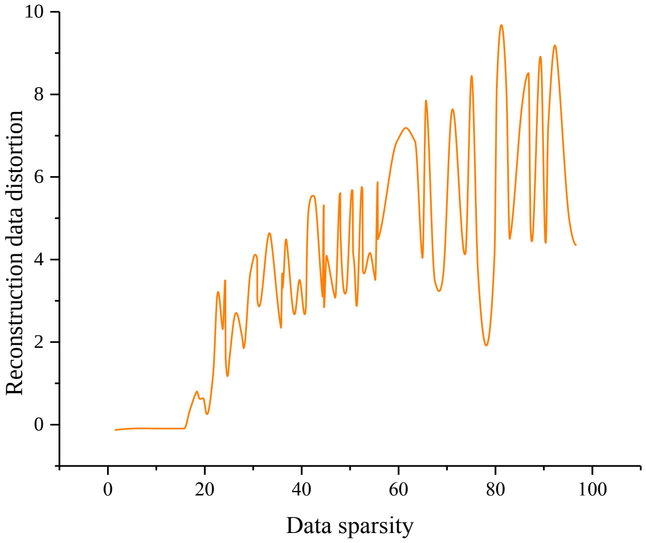


Figure 7 Boolean unrefined reconstruction data distortion **Figure 8** Boolean precision reconstruction data distortion



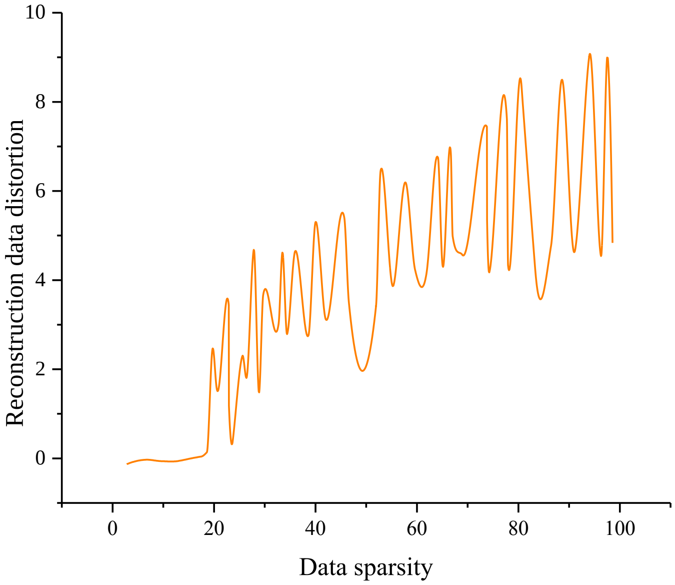


Figure 9 Discrete integer unreconstructed reconstruction data distortion

Figure 10 Discrete integer precision reconstruction data distortion

Figure 7-10 shows the distortion of the reconstructed data under different data sparsity. Figures 7 and 8 show the unrefined reconstruction results and refined results of Boolean data, respectively. Figures 9 and 10 show the unrefined reconstruction results and refined results of discrete integer data. result. As can be seen from Figure 7-10, the distortion of the reconstructed data starts to increase after the data sparsity reaches a certain amount (about 20% in the figure), when the sparsity is within a certain range (below 20% Left and right), the degree of distortion is 0 or approximately 0. This is because, when the projection dimension is constant, as long as the number of non-zero items contained in the data record is less than a certain value, the data reconstruction method based on minimization can find a unique sparse solution. This means that data records with sparseness within this range can be accurately or approximately reconstructed. Comparing Figure.7-10, it can be found that if the original data is known to be Boolean or discrete integer, then the precision can slightly increase the number of accurate reconstruction data and the quality of the reconstruction data. This is because the reconstructed data will have a slight error compared to the original data. When the original data type is known, the reconstruction algorithm can remove this error, making the reconstruction result more accurate. Of course, even if the original data type is unknown, this small error does not affect the accuracy of the reconstruction result.

V DISCUSSION

This research focuses on the dual privacy protection of the original data and rules, as well as the high efficiency and privacy protection required by large-scale distributed computing. Based on the analysis of financial risks, this paper presents a distributed data stream mining processing model for dual privacy protection. While solving the problem of privacy leakage, it can better mine the knowledge contained in the distributed data stream. In addition, this paper proposes a privacy oriented distributed data mining algorithm. According to the given processing model, random interference and double privacy protection technology are used to protect the original sensitive data, while increasing the incremental judgment and transmission of critical closure. Finally, simulation experiments are carried out on the algorithm proposed in this paper. The results show that the algorithm is getting better privacy protection effect. In addition, the distributed data stream mining for privacy protection researched in this paper is still in the theoretical research stage, and should be more closely integrated with the reality to make further improvements.

List of abbreviations

(CPU)	Central Processing Unit

Declarations

Availability of data and material

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

declares that he has no conflict of interest.

* Research involving human participants and/or animals

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

* Informed consent

All authors agree to submit this version and claim that no part of this manuscript has been published or submitted elsewhere.

Funding

None

Authors' contributions

Yuhao Zhao wrote the entire article.

Acknowledgements

None

REFERENCES

[1] S. A. O. U. Akyüz, A. M. G. Pinheiro, and T. Ebrahimi, "Privacy protection of tone-mapped HDR images using false colours," *Iet Signal Processing,* 2017.

[2] Ali, Serhan, Koyuncugil, Nermin, and Ozgulbas, "Early Warning System for Financially Distressed Hospitals Via Data Mining Application," *Journal of Medical Systems,* 2011.

[3] D. D. Benetti, r. I. Benetti, r. A. Rivera, r. Sardenberg, and r. O'Hanlon, "Site Selection Criteria for Open Ocean Aquaculture," *Marine Technology Society Journal,* vol. 44, no. 3, pp. P.22-35, 2010.

[4] H. K. Bhuyan and N. K. Kamila, "Privacy preserving sub-feature selection in distributed data mining," *Applied Soft Computing,* vol. 36, p. S1568494615004536, 2015.

[5] L. Bonnafous, U. Lall, and J. Siegel, "An index for drought induced financial risk in the mining industry," *Water Resources Research,* vol. 53, no. 2, pp. 1509-1524, 2017.

[6] Domenico, Talia, Paolo, and Trunfio, "How distributed data mining tasks can thrive as knowledge services," *Communications of the Acm,* 2010.

[7] O. J. Driskell *et al.*, "Inappropriate Requesting of Glycated Hemoglobin (Hb A1c) Is Widespread: Assessment of Prevalence, Impact of National Guidance, and Practice-to-Practice Variability," *Clinical Chemistry,* no. 5, p. 5, 2020.

[8] D. Ergu, G. Kou, Y. Peng, Y. Shi, and Y. Shi, "The analytic hierarchy process: task scheduling and resource allocation in cloud computing environment," *Journal of Supercomputing,* vol. 64, no. 3, pp. 835-848, 2013.

[9] T. Gneiting and R. Ranjan, "Combining predictive distributions," *Electronic Journal of Statistics,* vol. 7, pp. 1747-1782, 2013.

[10] M. R. Guariguata, B. Locatelli, and F. Haupt, "Adapting tropical production forests to global climate change: risk perceptions and actions," *International Forestry Review,* 2012.

[11] J. Isaak and M. J. Hanna, "User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection," *Computer,* vol. 51, no. 8, pp. 56-59, 2018.

[12] M. Jin, Y. Wang, and Y. Zeng, "Application of Data Mining Technology in Financial Risk Analysis," *Wireless Personal Communications,* 2018.

[13] S. Kim, J. Kim, and J. B. Weissman, "A Security-enabled Grid System for MINDS Distributed Data Mining," *Journal of Grid Computing,* vol. 12, no. 3, pp. 521-542, 2014.

[14] A. S. Koyuncugil and N. Ozgulbas, "Financial early warning system model and data mining application for risk detection," *Expert Systems with Applications,* vol. 39, no. 6, pp. p.6238-6253, 2012.

[15] Kugelberg and Elisabeth, "Infection: Double skin protection," *Nature Reviews Immunology,* vol. 15, no. 2, pp. 68-69, 2015.

[16] C. Li, X. Xie, Y. Huang, H. Wang, and C. Niu, "Distributed Data Mining Based on Deep Neural Network for Wireless Sensor Network," *International Journal of Distributed Sensor Networks,* vol. 2015, pp. 1-7, 2015.

[17] J. E. R. Matthew Bohm, "Treatment of eosinophilic esophagitis: overview, current limitations, and future direction," *American Journal of Gastroenterology,* vol. 103, no. 10, p. 2635, 2016.

[18] M. Nakazaki *et al.*, "Double balloon protection during carotid artery stenting for vulnerable carotid stenosis reduces the incidence of new brain lesions," *Acta Neurochirurgica,* vol. 158, no. 7, pp. 1377-1386, 2016.

[19] Qiao *et al.*, "Trajectory Data Mining in Distributed Sensor Networks," *International Journal of Distributed Sensor Networks,* 2015.

[20] S. Ronnqvist and P. Sarlin, "Bank distress in the news: Describing events through deep learning," *Neurocomputing,* vol. 264, no. nov.15, pp. 57-70, 2017.

[21] S. K. Roy, H. Sekhon, and J. F. Devlin, "Perceptions of fair treatment in financial services," *Journal of Endocrinology,* vol. 214, no. 2, pp. 165-175, 2012.

[22] N. Shibuya, D. C. Jupiter, L. J. Ciliberti, V. Vanburen, and J. L. Fontaine, "Characteristics of Adult Flatfoot in the United States," *Journal of Foot & Ankle Surgery,* vol. 49, no. 4, pp. 363-368, 2010.

[23] Sridhar and S., "Improving diagnostic accuracy using agent-based distributed data mining system," *Informatics for health & social care,* 2013.

[24] Y. Wang, J. Shi, and C. W. W. Ng, "Numerical modeling of tunneling effect on buried pipelines," *Canadian Geotechnical Journal,* vol. 48, no. 7, pp. 1125-1137, 2011.

[25] Xavier *et al.*, "A Windowing strategy for Distributed Data Mining optimized through GPUs," *Pattern Recognition Letters,* 2017.

[26] G. Xu, H. Li, S. Liu, M. Wen, and R. Lu, "Efficient and Privacy-Preserving Truth Discovery in Mobile Crowd Sensing Systems," *IEEE Transactions on Vehicular Technology,* vol. 68, no. 4, pp. 3854-3865, 2019.

[27] B. W. Yap, H. O. Seng, and N. H. M. Husain, "Using data mining to improve assessment of credit worthiness via credit scoring models," *Expert Systems with Applications*, vol. 38, no. 10, pp. p.13274-13283, 2011.

[28] L. Yue *et al.*, "High performance hollow carbon@SnO2@graphene composite based on internal-external double protection strategy for lithium ion battery," *Electrochimica Acta*, vol. 220, pp. 222-230, 2016.

[29] Zhu and Hongfeng, "A provable privacy-protection system for multi-server environment," *Nonlinear Dynamics*, 2015.

[30] H. Zhu, "A provable privacy-protection system for multi-server environment," *Nonlinear Dynamics*, vol. 82, no. 1, pp. 835-849, 2015.

Figure 1:

Title: Schematic diagram of financial risk concept

Legend: Figure 1 show Schematic diagram of financial risk concept

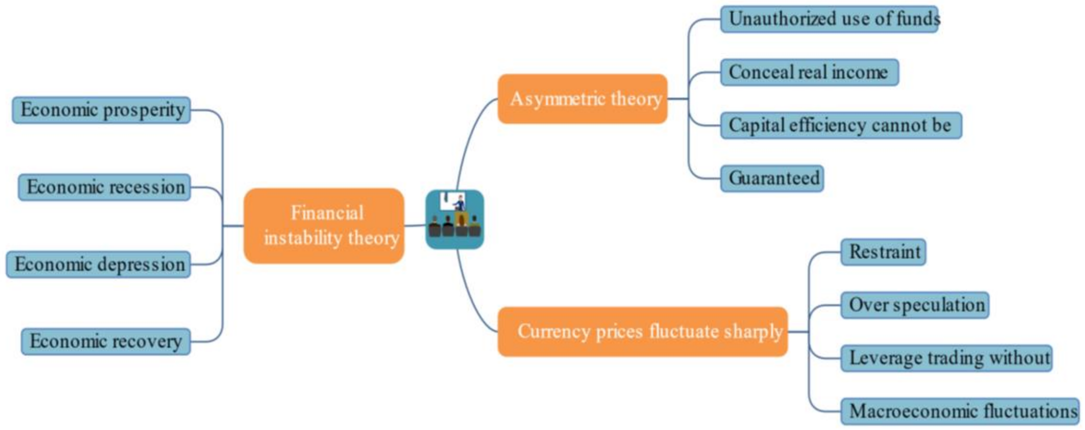


Figure 2:

Title: Aiming at the horizontal division of privacy protection data sharing

Legend: Figure 2 show Aiming at the horizontal division of privacy protection data sharing

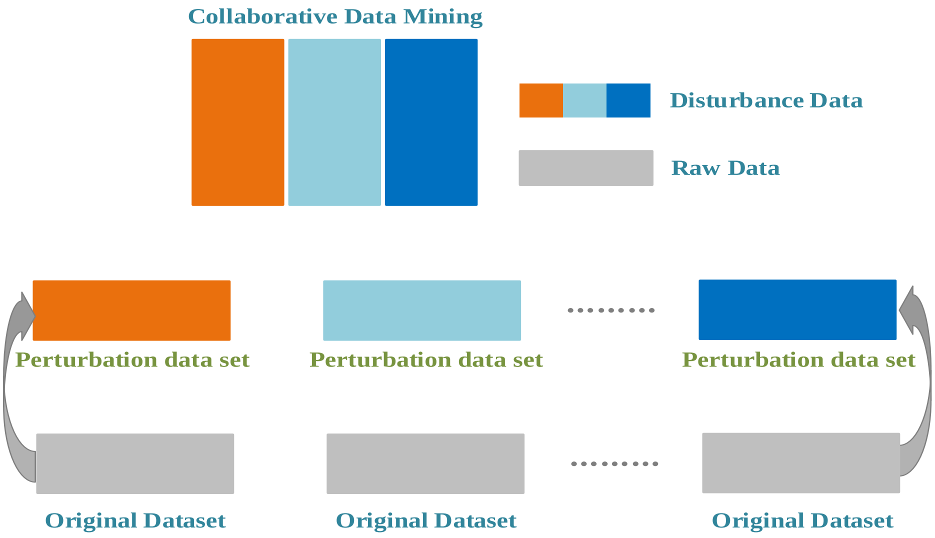


Figure 3:

Title: The privacy protection node model for distributed data mining

Legend: Figure 3 show The privacy protection node model for distributed data mining

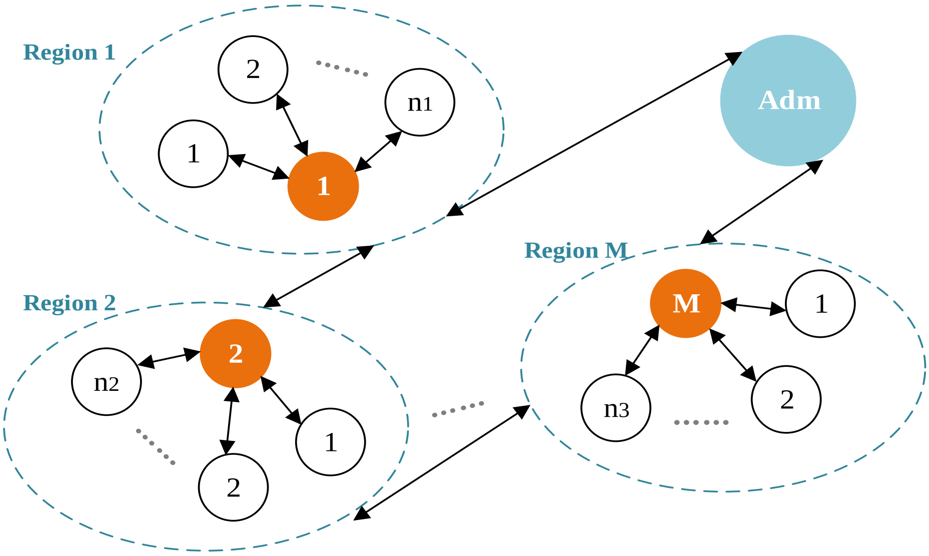


Figure 4:

Title: Schematic diagram of the disturbance data released by the terminal

Legend: Figure 4 show Schematic diagram of the disturbance data released by the terminal

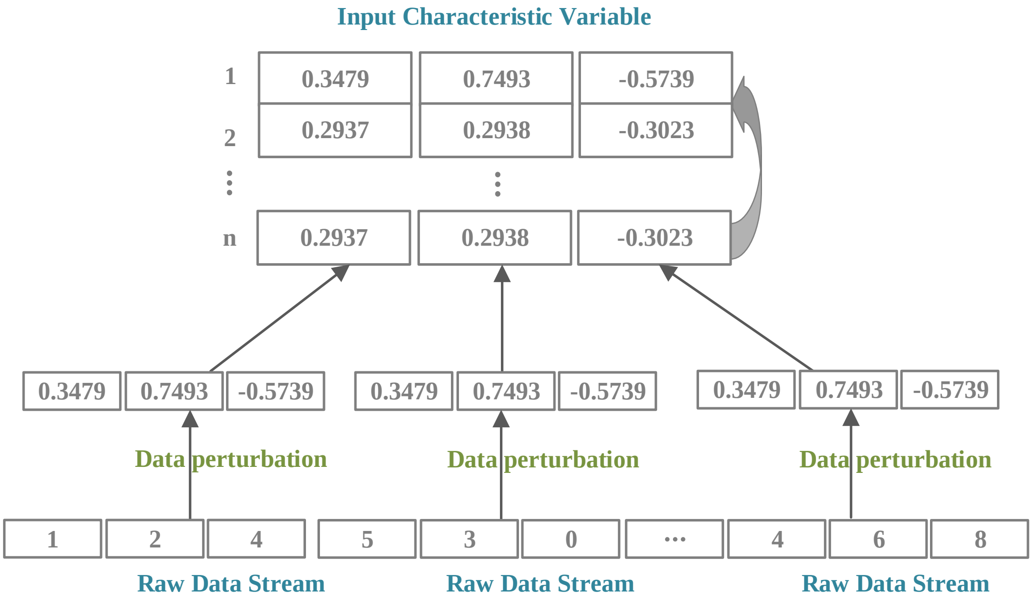


Figure 5:

Title: Comparison of original data and reconstructed data in a financial platform data set

Legend: Figure 5 show Comparison of original data and reconstructed data in a financial platform data set

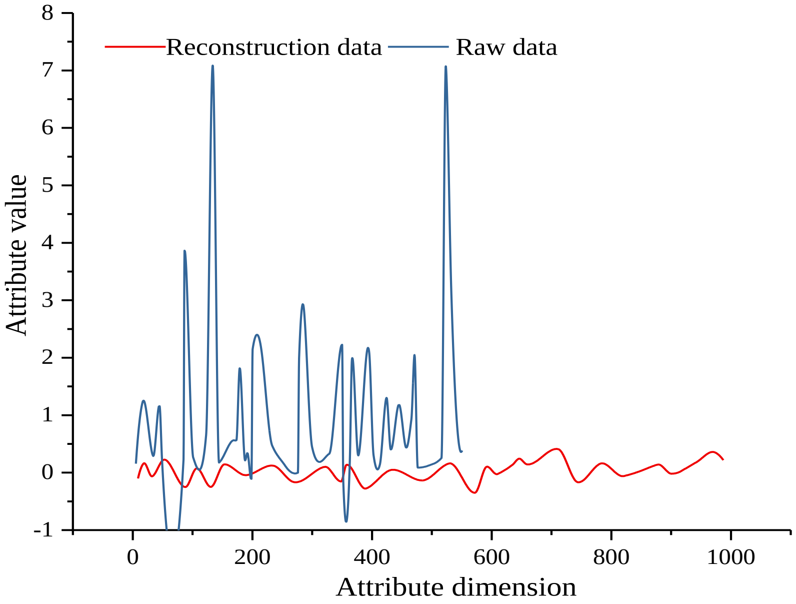


Figure 6:

Title: Comparison of original data and reconstructed data in a financial analysis institution's data set

Legend: Figure 6 show Comparison of original data and reconstructed data in a financial analysis institution's data set

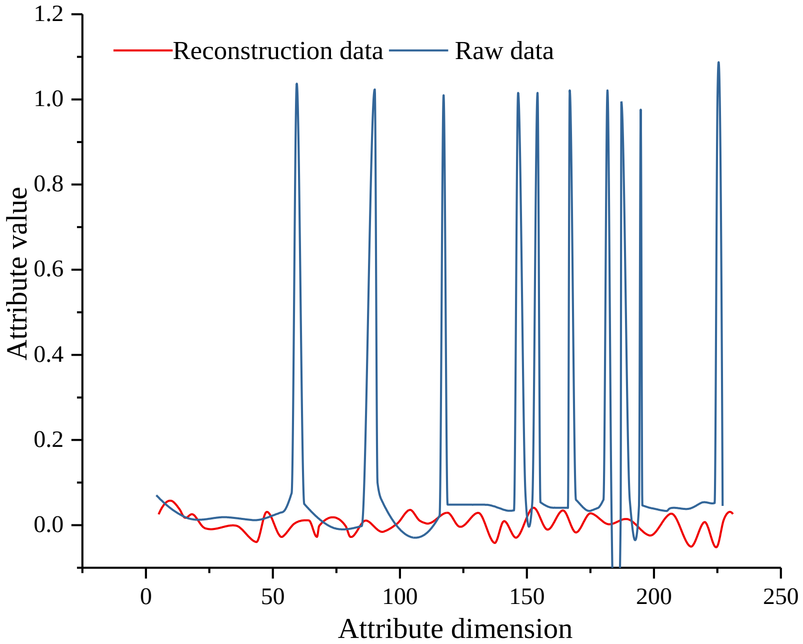


Figure 7:

Title: Boolean unrefined reconstruction data distortion

Legend: Figure 7 show Boolean unrefined reconstruction data distortion

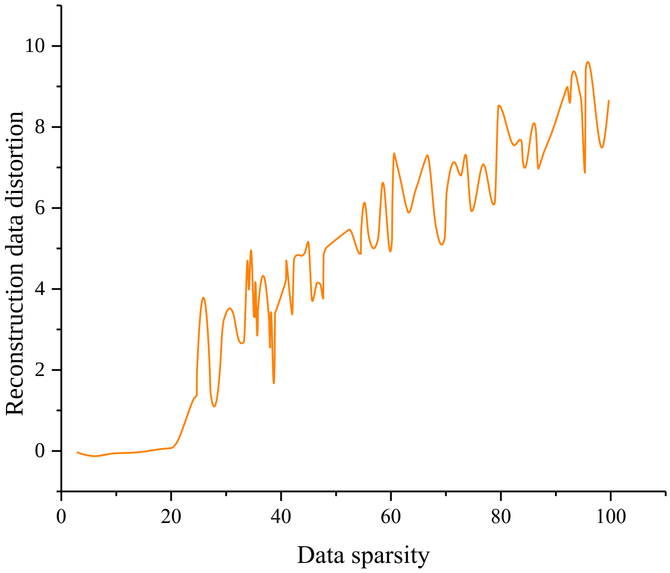


Figure 8:

Title: Boolean precision reconstruction data distortion

Legend: Figure 8 show Boolean precision reconstruction data distortion

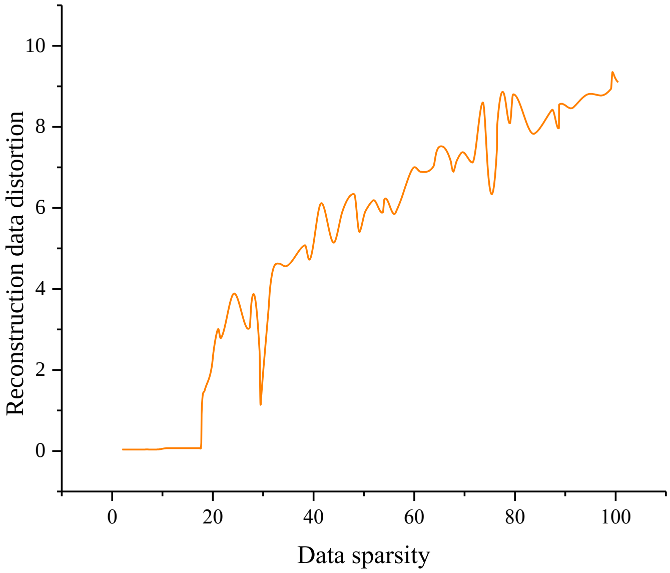


Figure 9:

Title: Discrete integer unreconstructed reconstruction data distortion

Legend: Figure 9 show Discrete integer unreconstructed reconstruction data distortion

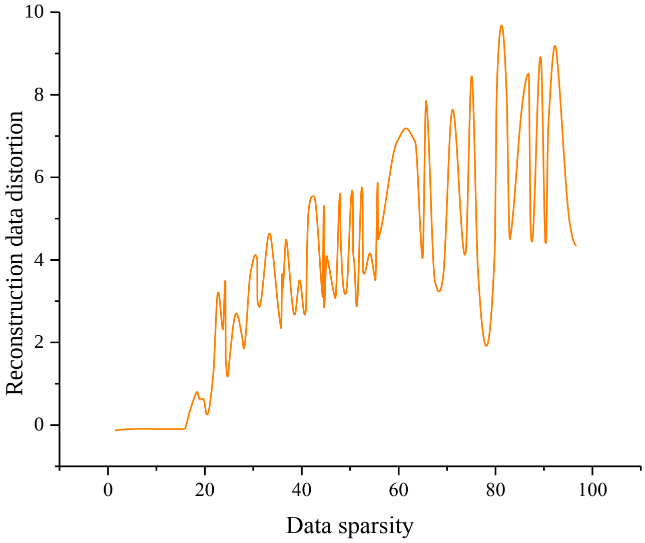
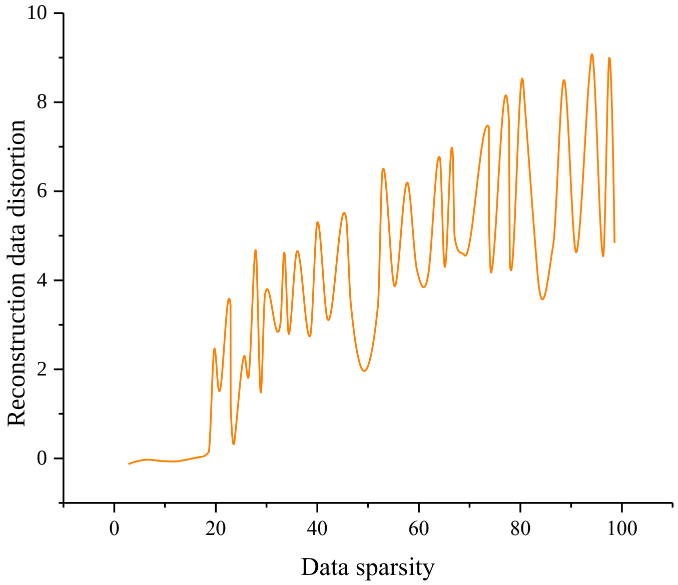


Figure 10:

Title: Discrete integer precision reconstruction data distortion

Legend: Figure 10 show Discrete integer precision reconstruction data distortion


Reconstruction data distortion
Data sparsity