International Journal of Web Based Communities

ISSN online: 1741-8216 - ISSN print: 1477-8394
https://www.inderscience.com/ijwbc

Privacy protection of multiple sensitive attribute data for users on e-commerce social media platforms

Na Wang, Ji Zhang, Feng Gao

DOI: 10.1504/IJWBC.2024.10061792

Article History:

Received:	25 May 2023
Last revised:	10 July 2023
Accepted:	10 October 2023
Published online:	04 November 2024

Privacy protection of multiple sensitive attribute data for users on e-commerce social media platforms

Na Wang

School of Management,
Changchun University of Architecture and Civil Engineering,
Chang'chun, 130607, China
Email: nawang@mls.sinanet.com

Ji Zhang*

School of Electrical Engineering,
Changchun University of Architecture and Civil Engineering,
Chang'chun, 130607, China
Email: jizhang33@mls.sinanet.com
*Corresponding author

Feng Gao

School of Art,
Changchun University of Architecture and Civil Engineering,
Chang'chun, 130607, China
Email: Fengg@36haojie.com

Abstract: Protecting the privacy of multi-sensitive attribute data is of great significance for safeguarding the interests of individuals, enterprises, and countries, as well as promoting technological development. Therefore, this paper proposes a privacy protection method of multiple sensitive attribute data for users on e-commerce social media platforms. An improved artificial bee colony algorithm is used to improve the KHM algorithm and mine multi-sensitive attribute data. Based on a personalised anonymity model, the mined multi-sensitive attribute data is quantified in a graded manner according to the sensitivity value and user privacy requirements of each attribute data. Low-sensitive attribute data is protected by equivalent expression privacy protection, while high-sensitive attribute data is encrypted and hidden by an improved fully homomorphic encryption method, thus achieving data privacy protection. The experimental results show that the probability of successful abnormal extraction of data by hackers using this method is less than 0.05, which improves privacy security.

Keywords: e-commerce; multiple sensitive; attribute data; privacy protection; KHM algorithm; social media.

Reference to this paper should be made as follows: Wang, N., Zhang, J. and Gao, F. (2024) 'Privacy protection of multiple sensitive attribute data for users on e-commerce social media platforms', *Int. J. Web Based Communities*, Vol. 20, Nos. 3/4, pp.278–297.

Biographical notes: Na Wang graduated from Changchun University of Technology in 2010. She worked in Changchun University of Architecture and Civil Engineering. Her main research fields mainly include civil laws include enterprise management, regional economic management, e-commerce and so on. Currently, she works in the School of Management in Changchun Institute of Architecture as an Associate Professor.

Ji Zhang graduated from Changchun University of Technology in 2014. He worked in Changchun University of Architecture and Civil Engineering. His main research fields mainly include social work, enterprise management, e-commerce, digital economy, etc. Currently, he works in the School of Electrical Engineering in Changchun Institute of Architecture as an Assistant.

Feng Gao graduated from Jilin College of the Arts in 2004. He worked at Changchun University of Architecture and Civil Engineering. His main research fields mainly include data analysis, computer, computer net, environmental design, database, etc. Currently, he works at the School of Art in Changchun Institute of Architecture as a Lecturer.

1 Introduction

E-commerce social media platforms belong to a derivative model of electronic commodity platforms. This platform can provide users with product display, sharing, and purchasing services through social interaction, thereby promoting customers to trade goods on this platform (Chen et al., 2022). In recent years, the development process of traditional e-commerce has been hindered, entering a bottleneck period, and the growth rate of users has gradually slowed down. In order to obtain more customers, the cost between enterprises has also gradually increased (Liao et al., 2020). At this point, many businesses and users enjoy completing business activities on social media platforms. This type of e-commerce model is called 'social e-commerce' (Zennyo, 2020). According to relevant data, in 2018, social e-commerce in China became the core commodity trading mode in the online shopping market. In the context of the widespread application of mobile intelligent terminals, the development speed of social e-commerce has significantly improved. Various e-commerce social media platforms have also emerged (Liu et al., 2020). Although such e-commerce social media platforms promote product transactions, while providing users with social services and product purchase services, the privacy protection of user sensitive attribute data also needs to be highly valued (Zeng et al., 2022).

User privacy belongs to the personal information that users cannot casually tell others in their daily lives. In addition to their ID card information and bank card information, their shopping records, activity trajectories, social information, and other data belong to privacy data (Xiang and Zhang, 2020). In e-commerce social media platforms, user sensitive attribute data privacy is mainly divided into user basic information, trajectory information, social information, and data mining information. Basic information mainly includes privacy information such as the user's name, gender, contact phone number, etc.; track information refers to all real-time geographic data, historical location information, etc. related to the user; social information represents the social chain information of users on e-commerce social media platforms; data mining information

represents the perceived information of e-commerce social media platforms on user preferences (Zhao, 2021).

In e-commerce social media platforms, the factors that lead to privacy risks for users are mainly divided into user factors and platform factors. Many e-commerce social media platforms require users to complete real name registration when logging in to improve customer stickiness. This type of registration information includes the user's actual name, ID card, phone number, and other privacy information (Aivazpour and Rao, 2020). And many e-commerce social media platforms also require access to user contact books, location information, and other information, which requires social account binding and phone number binding. Such behaviour can lead to the leakage of user privacy information (Wang et al., 2021a). In addition, the activity trajectories and daily lives shared by users on e-commerce social media platforms also involve their positioning and photo privacy information. These factors are all user factors. There are three main reasons for user privacy issues caused by platform factors: lack of platform technology, big data analysis, and the need to improve industry self-discipline. Many e-commerce social media platforms have technical vulnerabilities and do not have privacy clauses, which can lead to information leakage when hacker attacks occur. And now, big data analysis technology is becoming increasingly intelligent and advanced. Many hackers use big data analysis technology to mine private data published by users on e-commerce social media platforms, such as text, video, voice, etc. After assembling such private data, it can be leaked without the user's knowledge. When operating on e-commerce social media platforms, the industry self-discipline of social networks is poor, and failure to adopt effective privacy protection technologies can also lead to user privacy leakage issues on e-commerce social media platforms (Sun et al., 2022).

In summary, in the application of e-commerce social media platforms, how to protect user privacy and security under various factors is currently a key research issue in the field of information security. According to existing research materials, Zhang et al. (2022) protect user privacy information by deleting identifier attributes in data tables. However, if attackers identify quasi identifiers and background knowledge, they can still obtain user privacy information, which cannot fully guarantee that privacy data will not be extracted by hackers. Wang et al. (2021b) use homomorphic encryption to protect data privacy, but in e-commerce social media platforms, user sensitive data has multiple sensitive attributes, and this method is only applicable to the privacy protection of data with single sensitive attribute. For multi-sensitive attribute data, this method has limitations and a high probability of privacy data leakage. Li et al. (2022) study a blockchain regulatory privacy protection scheme based on group signature and attribute encryption. This method can ensure the security of user sensitive data in the blockchain privacy protection framework through group signature and attribute encryption. However, this method has low operational efficiency, complex process, and needs to be optimised.

Based on the importance of privacy protection of multi-sensitive attribute data of users of e-commerce social media platform, and the problems of previous research methods, this paper studies a new privacy protection method of multiple sensitive attribute data for users on e-commerce social media platforms. Therefore, this method has the characteristics of small intra class distance, large inter class distance, good privacy protection effect of multi-sensitive attribute data, short encryption time, low loss ratio of original data, and low probability of successful data extraction by hackers. The main Technology roadmap of this method is as follows:

1 Using the improved artificial bee colony algorithm to improve the KHM algorithm, after determining the initial clustering centre of multi-sensitive attribute data clustering, mining multi-sensitive attribute data through clustering.

2 Based on the personalised anonymous model, the mined multi-sensitive attribute data will be quantified in a hierarchical manner, and the sensitive attribute values and user privacy needs of each attribute data will be quantified. The low sensitive attribute data will be subject to equivalent privacy protection, and the high sensitive attribute data will be encrypted and hidden by the improved full homomorphic encryption method, so as to achieve data privacy protection.

3 Taking the intra class distance and inter class distance, algorithm fitness, privacy protection effect of multi-sensitive attribute data, encryption time, loss ratio of original data, and the probability of success of data being extracted by hackers from the user's multi-sensitive attribute data mining results of e-commerce social media platform as experimental indicators, the method's practical application effect was tested in depth from multiple aspects.

2 Privacy protection methods of multiple sensitive attribute data for users on e-commerce social media platforms

2.1 User multi-sensitive attribute data mining based on improved KHM algorithm

In recent years, data mining technology has been widely used in the internal information mining problem of hidden data. Cluster analysis is a mining mode of data analysis that can complete data classification mining based on the similarity between data (Wang et al., 2020). This article uses the K-harmonic mean clustering algorithm (referred to as KHM algorithm) to mine user sensitive attribute data. This algorithm can effectively overcome the shortcomings of the K-mean algorithm, but its initial clustering centre cannot be blindly set. Therefore, this article improves it by using an improved artificial bee swarm algorithm to ensure the application performance of the algorithm through optimisation settings. The objectives of using the improved KHM algorithm for user multi-sensitive attribute data mining are as follows:

1 Improving the accuracy of data mining: improving the KHM algorithm is committed to improving the accuracy of mining multi-sensitive attribute data. By optimising or improving algorithms, more accurate and reliable results can be obtained when mining multiple sensitive attributes.

2 Processing multidimensional sensitive attribute data: multiple sensitive attributes of users often have strong correlations and interactive relationships. The improved KHM algorithm aims to effectively handle this multidimensional sensitive attribute data. By introducing appropriate models or indicators, users' preferences, behaviour patterns, and potential association rules can be better mined.

3 Privacy protection and data masking: privacy protection is one of the most important considerations when mining data with multiple sensitive attributes. The improved

KHM algorithm aims at privacy protection and improves data masking to protect users' sensitive information.

4 Improving computational efficiency: improving the KHM algorithm is also committed to optimising the computational process to improve the computational efficiency of the algorithm. This can mine and analyse large-scale, high-dimensional, and multi-sensitive attribute data in a shorter time, improving work efficiency.

The characteristics of using the improved KHM algorithm for user multi-sensitive attribute data mining are as follows:

1 Multidimensional attribute mining: the improved KHM algorithm is particularly suitable for multi-dimensional sensitive attribute data mining. It can simultaneously mine and analyse the relationships between multiple sensitive attributes, providing more accurate and comprehensive user behaviour patterns and preference information.

2 Privacy protection and data encryption: improving the KHM algorithm to focus on user privacy protection and data encryption when conducting data mining. By processing and encrypting sensitive attribute data, sensitive information is protected to prevent unauthorised access and abuse.

3 Efficient computational performance: improved KHM algorithm utilises hash mapping and index-based technology to improve the computational performance and efficiency of the algorithm. This enables the algorithm to handle large-scale multidimensional datasets and perform efficient data mining and analysis in a relatively short period of time.

4 Scalability and flexibility: the improved KHM algorithm has high scalability and flexibility, and can be applied to datasets of different scales and complexities. It can adjust parameters and optimise algorithms according to specific situations to meet the needs of different application fields.

2.1.1 KHM algorithm

In e-commerce social media platforms, the user dataset that needs to be clustered is $Y = \{y_1, \ldots, y_m\}$, where m represents the number of types of user data on e-commerce social media platforms.

When users of e-commerce social media platform cluster their multi-sensitive attribute cluster analysis, the cluster centre set and cluster centre of multi-sensitive attribute data are $D = \{d, \ldots, d_k\}$ and k respectively. The j e-commerce social media platform user data point y_i belongs to the i sensitive attribute data with a membership level set to $n(d_i|y_i)$. The impact of the j e-commerce social media platform user data point y_i on the clustering effect during the iteration process is described by the weight function $\varpi(y_i)$.

The clustering process of the KHM algorithm is as follows:

1 In e-commerce social media platforms, set the initial clustering centre KHM(Y, D) for user sensitive attribute data that needs to be clustered.

2 The objective function value for initial clustering of multi-sensitive attribute data:

$$KHM(Y, D) = \sum_{j=1}^{m} \frac{k}{\sum_{j=1}^{k} \frac{1}{\left\| y_j - d_j \right\|^k}} \tag{1}$$

When the intra class distance is small and the inter class distance is large, it indicates that the data mining accuracy is high (Fadaei et al., 2022). Therefore, the setting of the initial clustering centre for user sensitive attribute data needs to meet the criteria of 'minimum intra class distance and maximum inter class clustering'.

3 For the j e-commerce social media platform user data point y_i, the membership degree of operation y_i at each cluster centre point d_i:

$$n\left(d_j \left| y_j \right.\right) = \frac{\left\| y_j - d_j \right\|^{k+2}}{\sum_{j=1}^{k} \left\| y_j - d_j \right\|^{k+2}} \tag{2}$$

4 Calculate the weights of each data point:

$$\varpi\left(y_j\right) = \frac{\sum_{j=1}^{k} \left\| y_j - d_j \right\|^{k+2}}{\left(\sum_{j=1}^{k} \left\| y_j - d_j \right\|^{-k} \right)^2} \tag{3}$$

5 Combining membership and weight, calculate the clustering centres of all data points to complete multi-sensitive attribute data mining:

$$Y' = \frac{\sum_{j=1}^{m} n\left(d_j \left| y_j \right.\right) \varpi\left(y_j\right) y_j}{\sum_{j=1}^{m} n\left(d_j \left| y_j \right.\right) \varpi\left(y_j\right)} \tag{4}$$

6 Repeat steps 2 to 5 multiple times, and KHM(Y, D) can end without significant changes. Output the multi-sensitive attribute data mining result Y' at this time.

2.1.2 Implementation of data mining based on improved KHM algorithm

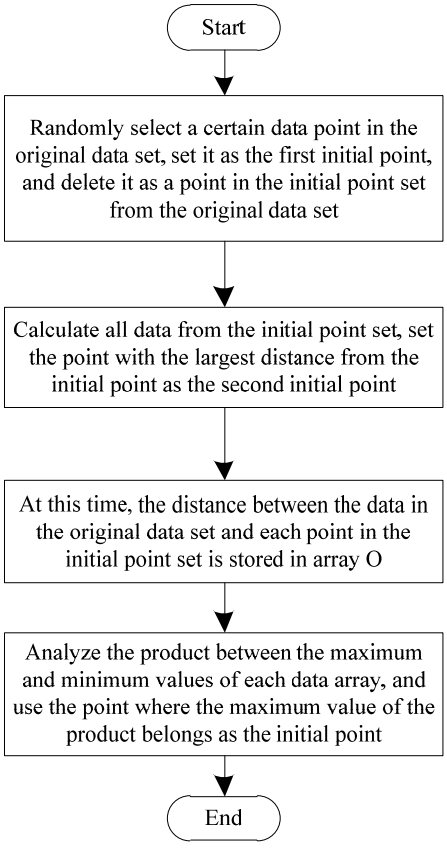
Considering the shortcomings of initialisation, fitness function and location update methods in the original artificial bee colony algorithm, which are easy to fall into local optimisation, this paper improves the original artificial bee colony algorithm. The improved artificial bee colony algorithm applies the maximum minimum distance product method to the task of initialising the bee colony to solve the problem of initialisation randomness, and then uses the new fitness function and the position update method of the global guidance factor to complete the iterative optimisation of the initial cluster centre of user multi-sensitive data (Jaleel et al., 2021).

2.1.2.1 Maximum minimum distance product method

The initialisation of artificial bee populations is crucial in the entire process of optimising the initial clustering centre of user sensitive attribute data, as it directly affects the overall convergence efficiency and solution quality of the algorithm. Therefore, this article uses

the maximum minimum distance product method to set the initial bee colony as the feasible solution set for the initial clustering centre of user sensitive attribute data, with a number of m. It is necessary to set the initial number of points in a feasible solution as k, and the set of k initial points as L. Multiply each data saved in L with each data in the original user dataset Y, and save the product result in array O. Figure 1 is the operational flowchart of the maximum minimum distance product method.

Figure 1 Operation flowchart of the maximum minimum distance product method



2.1.2.2 Fitness function

This function is mainly used to guide the direction of bee colony evolution, and has a direct impact on the evolution mode, iteration number, and initial clustering centre solution quality of the bee colony (Zhou et al., 2022). Different fitness function will also lead to different initial clustering central solutions (Zheng et al., 2022). Therefore, combined with the research content of this paper, a new fitness function is designed:

$$fitness_j = N_j / I_j \tag{5}$$

In the formula, N_j and I_j are the number of data points belonging to the j user's multi-sensitive attribute data, and the distance and value from the intra class target of the j user's multi-sensitive attribute data to the cluster centre point d_j.

2.1.2.3 Location update

The location update results have a direct impact on the accuracy and efficiency of artificial bee individuals in searching for new honey sources. In the position update method of the original artificial bee colony algorithm, although individuals also have the ability to search for honey sources, their ability is insufficient, and the iterative process has randomness, resulting in a high probability of falling into a local optimal solution. Therefore, this article applies the global factor ($c_{best} - c_{ji}$) to the position update problem. Then:

$$U_{ji} = c_{ji} + s_{ji}\left(c_{ni} - c_{hi}\right) + \beta\left(c_{best} - c_{ji}\right) \tag{6}$$

In the formula, U_{ji} represents the update result of the individual position of the artificial bee, while h, n, and i belong to random numbers, with a value range of 1-1. c_{best} represents the honey source with the most significant food richness. c_{ji}, c_{hi} and c_{ni} respectively represent different honey sources; β is the influencing factor.

In the original artificial bee colony algorithm, the difference in position before and after iteration was not analysed, and each leading bee can only obtain its own optimal position and current position, without the ability to search for a global optimal solution (Wang et al., 2022). Based on the analysis from the perspective of swarm intelligence evolution, in the retrieval process, each bee individual can get the experience benefit of peer individuals in the group. Therefore, the use of global factor ($c_{best} - c_{ji}$) can improve the directional and purposeful retrieval of bee individuals. Using influence factor β in front of global factor ($c_{best} - c_{ji}$) can serve as a constraint for the initial clustering centre optimisation amplitude. Based on the analysis of factor composition, if there is a significant difference between the current position and the optimal position, the update step size will also dynamically change.

The main reasons for improving the KHM algorithm include addressing the need for multiple sensitive attributes, improving privacy protection capabilities, improving the effectiveness and accuracy of the algorithm, and reducing computational complexity. By improving the relevant aspects of the algorithm, the KHM algorithm can be more adaptable and meet the data mining needs in various practical scenarios. The improved artificial bee algorithm will be used in the initial clustering setting stage of the KHM algorithm, and the specific steps are as follows:

1 Algorithm initialisation

When mining user sensitive attribute data on e-commerce social media platforms, the initial number of clustering centres in the KHM algorithm is used to set artificial bee individuals as feasible solutions for the initial clustering centre optimisation setting. Artificial bee individuals are divided into leading bees, following bees, and reconnaissance bees, where the number of leading bees is consistent with the number of following bees; set the maximum number of iterations to W, the current number of iterations is w, the initial value of iterations is 1, and the number of cluster types is k. Initialise the bee colony using the maximum minimum distance product method.

2 Cluster the initial bee colony once, calculate the fitness of each artificial bee individual, and arrange the size according to the size of fitness. Set half of the bee individuals with larger fitness as the leading bees, and the remaining half of the artificial bees as the following bees.

3 Leading bees to use formula (6) to perform neighbourhood search and obtain new honey source locations.

4 The following bees follow their own goals with probability Q_i and roulette principle. If Q_i is large, it means that the leading bee j has a large fitness, and its probability as a following bee is also large. If the follower bee has selected the leader bee, it will directly start the neighbourhood search. In the search process, the greedy selection principle is used, and the most prominent position of fitness is taken as the optimal position for the initial cluster centre.

5 When all the following bees complete the search task, set the obtained location as the initial clustering centre of the user degree multiple sensitive attribute data, perform iterative clustering, and update the bee colony with a new clustering centre in the user degree multiple sensitive attribute dataset after clustering.

6 If the number of iterations of the leader bee meets the maximum value, it will be converted into a reconnaissance bee and any new position will be generated to replace the original position.

7 If the current number of iterations exceeds the maximum, the iteration will be ended, and the bee with the largest fitness will be output, and the initial cluster centre of the user degree multiple sensitive attribute data represented by it will be taken as the optimal solution. On the contrary, skip to step 2.

2.2 A privacy protection method for multi-sensitive attribute data based on personalised anonymous model

Personalised anonymity model is a privacy protection method, which aims to data anonymisation sensitive attribute data to protect individual privacy. It is based on the data anonymisation technology, which meets the data analysis needs while minimising the risk of user identity information disclosure. The personalised anonymous model has the following advantages for privacy protection of multi-sensitive attribute data:

1 Targeted privacy protection: personalised anonymous models can provide flexible privacy protection based on different sensitive attributes and privacy needs in the dataset. It allows you to select different anonymity levels for each sensitive attribute according to the different impact of each sensitive attribute or the signature requirements for individual users.

2 Customised anonymity level control: through a personalised anonymity model, specific details of the anonymity level can be controlled, such as sharing an anonymous bucket among several users, the number of users contained in each anonymous bucket, and so on. These control options enable a better balance between the economic benefits of data and privacy protection.

3 Effectively combining multiple sensitive attributes: personalised anonymous models can handle the correlation between different attributes and provide more precise

privacy protection by combining multiple sensitive attributes. The improved anonymous algorithm can simultaneously handle multiple sensitive attributes, reduce the risk of privacy leakage, and maintain the analytical availability of data while protecting privacy.

4 Accuracy maintenance and data analysis capabilities: personalised anonymous models aim to achieve privacy protection while maximising the accuracy and usefulness of data. It protects user privacy by appropriately perturbing and aggregating sensitive information, while maintaining the basic statistical features and data analysis capabilities of the dataset.

5 Improve privacy protection flexibility: personalised anonymous models allow people to customise privacy protection mechanisms based on different datasets, tasks, and policies. It provides users with an integrated privacy technology and customised approach for multiple sensitive attributes, making privacy protection more flexible and feasible.

For the multi-sensitive attribute data mined in Section 2.1, because there are multiple sensitive attribute values in each data tuple, it is necessary to quantify the sensitivity of each sensitive attribute value. In addition, considering the differences in privacy needs of users for multiple sensitive attributes, it is necessary to analyse users' privacy needs when quantifying sensitive attribute values.

Quantify sensitive attribute values and user privacy needs in a hierarchical manner. The higher the sensitivity, the greater the corresponding attribute values and privacy needs.

Based on the user's privacy protection level requirements, set their demand sensitivity in the demand level table. If the level is l_p and the user's demand sensitivity is u_q, then:

$$l_p = \frac{n - u_q}{n - 1} \tag{7}$$

Among them, $u_q \in [0, 1]$, $l_p \in [0, 1]$. n is the number of sensitive attributes.

Set a data table with n sensitive attributes as Y', and the values of each sensitive attribute B_j are b_i, then the comprehensive sensitivity of this sensitive attribute value is:

$$u_l = \frac{\theta * u_{q1} + \vartheta * u_{q2} + \mu * u_{q3}}{\theta + \vartheta + \mu} \tag{8}$$

Among them, θ, ϑ and μ all belong to parameters between 0 and 1, representing the bias level of frequency sensitivity u_{q1}, grading sensitivity u_{q2}, and user demand sensitivity u_{q3}.

Set the n groups of Y' as $F = \{F_1, F_2, \ldots, F_n\}$. If the comprehensive sensitivity difference of a certain column of sensitive attribute values between two random tuples in each group is greater than the threshold, then this data table can be personalised and anonymous. In a certain group, it is necessary to control the difference in the comprehensive sensitivity of each tuple. For tuples in the same group, it is necessary to ensure that there are differences in the sensitivity of their sensitive information and their sensitivity levels, in order to prevent a certain group from privacy leakage due to the presence of high sensitive attribute values and being attacked by hackers. Then, in e-commerce social media platforms, the privacy protection method of multi-dimensional bucket processing is used to disrupt the sensitive attributes of original user information

and protect user privacy security. For highly sensitive data that cannot be processed equivalently, an improved all homomorphic encryption method is used to complete the hiding process.

The specific steps for this method are:

1 Calculate the comprehensive sensitivity of each sensitive attribute value in each sensitive attribute data tuple.

2 Design a multidimensional bucket structure, treating sensitive attribute tuples as multidimensional vectors, with each sensitive attribute being the dimensions of the multidimensional vector. Therefore, each sensitive attribute data tuple can be mapped to a multidimensional bucket.

3 Repeat grouping. Set all buckets to the unshielded mode, extract the non-empty buckets of the unshielded mode respectively, and incorporate them into the equivalence class group. Use the same quasi identifier value for the data of the equivalence class group, and then the low sensitive data can be hidden in a group of records. Then, in the remaining buckets, block buckets with dimensions similar to oneself. This operation circulates once to obtain an equivalence class group. Multiple loop operations can be stopped when the remaining tuples cannot form a complete group. Perform covert processing on the highly sensitive data in the remaining tuples, which cannot be publicly transmitted.

The hiding processing is mainly completed by the improved all homomorphic encryption method. This method is divided into three stages: key design, encryption, and decryption.

Firstly, construct two hidden key vectors τ and ξ with dimension m, and together form a set of real number pairs:

$$\lambda(m) = \left[(\tau_1, \xi_1), \ldots, (\tau_m, \xi_m) \right] \tag{9}$$

Then, set the highly sensitive data tuple to be encrypted as F', and then introduce random noise set η when encrypting:

$$\eta = \left[(\varepsilon_1, \eta_1), \ldots, (\varepsilon_{m-1}, \eta_{m-1}) \right] \tag{10}$$

Among them, ε_{m-1} is a random number. The ciphertext of highly sensitive data after introducing noise is F_η', and F_η' has m sub-ciphertexts. The operation method for sub-ciphertexts is:

$$f_j' = \begin{cases} \tau_j * \left(\xi_j * F' + \eta_j + \right) + \eta_j, & 1 \le j \le m-1 \\ \tau_m * \xi_m * \sum_{j=1}^{m-1} \left(\eta_j + \dfrac{\eta_j}{\tau_j} \right), & j = m \end{cases} \tag{11}$$

Introduce the mapping function ρ to convert the j sub-ciphertext of highly sensitive data ciphertext F_η' into the j sub-ciphertext of highly sensitive data unordered ciphertext F_η'', and set it to f_{jd}'. Under ρ mapping, the mapping results between ciphertexts are independent, and for different ciphertexts, the order of sub ciphertexts is independent, and the distribution is random.

Use the mapped highly sensitive data ciphertext as the encryption result of the data. When decrypting, if the sub ciphertext of F_η' is set to f_{jd}', the plaintext value of the decrypted highly sensitive data is:

$$F_\eta' = \sum_{j=1}^{m-1} \frac{f_{jd}'}{\tau_j * \xi_j} - \frac{f_j'}{\tau_m * \xi_m} \tag{12}$$

3 Experimental analysis

3.1 Experimental scheme

3.1.1 Experimental data

To test the effectiveness of this method in protecting the privacy of user sensitive attribute data in e-commerce social media platforms, the privacy protection data of eight users shown in Table 1 were used as the experimental subjects of this method. In Table 1, the sensitive attribute data types of the 8 users are mainly divided into four categories: gender, age, purchasing behaviour preference information, and postal code information of the shipping address. The experimental data is relatively comprehensive and reasonable.

Table 1 User information of e-commerce social media platforms

Code	*Gender*	*Age*	*ZIP code*	*Purchase preferences*
1	Female	20	230000	Books
2	Female	30	100000	Electrical equipment
3	Female	40	400000	Clothing
4	Female	50	350000	Home Furnishing
5	Male	20	230000	Books
6	Male	30	100000	Electrical equipment
7	Male	40	400000	Clothing
8	Male	50	350000	Home Furnishing

3.1.2 Evaluation indicators

Taking the intra class distance and inter class distance, algorithm fitness, privacy protection effect of multi-sensitive attribute data, encryption time, loss ratio of original data, and the probability of success of data being extracted by hackers from the user's multi-sensitive attribute data mining results of e-commerce social media platform as experimental indicators, the method's practical application effect was tested in depth from multiple aspects.

From a quantitative perspective, when analysing the mining effect of the method in this article on user multi-sensitive attribute data in e-commerce social media platforms, the mining effect can be analysed through intra class distance and inter class distance of multi-sensitive attribute data. When the intra class distance is small and the inter class

distance is large, it indicates that the data mining accuracy is high. The calculation method for intra class distance $In1$ and inter class distance $In2$ is:

$$In1 = \sum_{j=1}^{k} \sum_{d_j} \left\| Y - d_j \right\| \tag{13}$$

$$In2 = \min_{1 \le d_j} \left\| y_1 - y_2 \right\| \tag{14}$$

Among them, y_1 and y_2 represent different sensitive data.

In the evolution process of algorithms, the evaluation and optimisation of fitness functions can affect individual selection, crossover, and mutation operations. By defining the fitness function, the algorithm can select individuals with higher fitness in each generation and inherit their superiority, enabling the algorithm to continuously find better solutions.

The better the original feature concealment effect of multi-sensitive attribute data, the better the privacy protection effect of multi-sensitive attribute data.

Multi-sensitive attribute data encryption time refers to the amount of time required to encrypt multi-sensitive attribute data. It is one of the key indicators for measuring encryption speed and efficiency. The shorter the encryption time, the faster and more efficient the encryption process.

The loss ratio of original data is a measure of the degree of loss of original data information in the process of data encryption or processing. It indicates the impact of encryption or processing operations on data accuracy and integrity.

The probability of successful extraction of data by hackers refers to the probability of data being illegally obtained by hackers or unauthorised visitors. This indicator evaluates the security of data and measures the degree of threat to data from unauthorised access or hacker attacks.

3.2 Experimental result

After mining the multi-sensitive attribute data of users on e-commerce social media platforms using this method, the intra class distance and inter class distance of the multi-sensitive attribute data are shown in Figures 2 and 3.

Figure 2 Intraclass distance of multi-sensitive attribute data

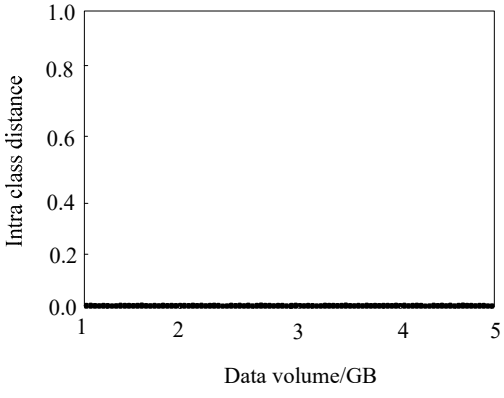
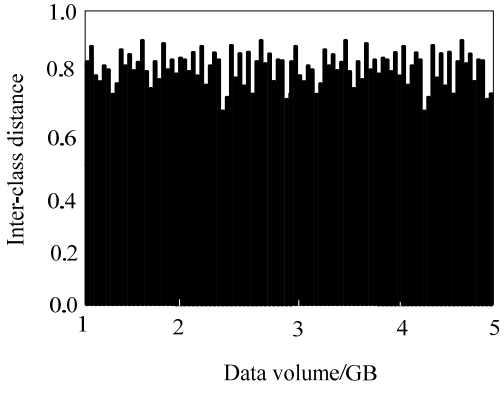


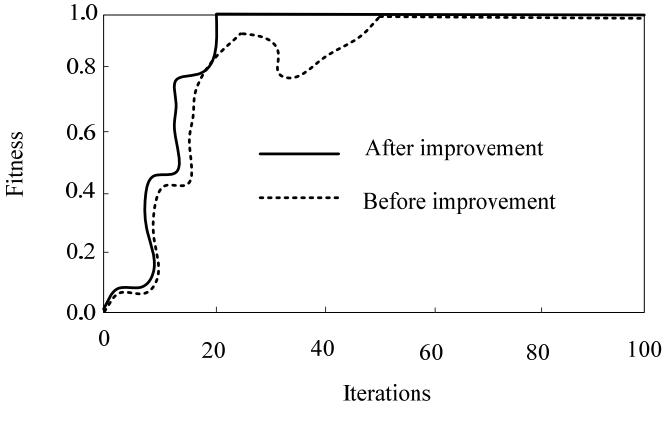
Figure 3 Inter class distance of multi-sensitive attribute data



As shown in Figures 2 and 3, after mining user multi-sensitive attribute data on e-commerce social media platforms using our method, the intra class distance of the multi-sensitive attribute data is relatively small, less than 0.1, and the inter class distance is relatively large, greater than 0.6. This indicates that our method has higher data mining accuracy for user multi-sensitive attribute data.

Figure 4 shows the fitness change of the initial clustering centre optimisation setting of the KHM algorithm before and after the improved artificial bee colony algorithm is used in the data mining link of this method.

Figure 4 Change of fitness before and after using improved artificial bee colony algorithm



As shown in Figure 4, before using the improved artificial bee colony algorithm, the KHM algorithm requires 48 iterations to obtain the initial clustering centre optimal solution of the KHM algorithm, which is slow and has poor iteration performance. After using the improved artificial bee colony algorithm, the fitness of the initial clustering centre optimisation setting of the KHM algorithm quickly converges to the optimal solution. Only 20 iterations can obtain the optimal solution of the initial clustering centre of the KHM algorithm, thereby ensuring the mining efficiency of the KHM algorithm for multi-sensitive attribute data.

3.3 Analysis of privacy protection effectiveness for multiple sensitive attribute data

After mining multiple sensitive attribute data for users on e-commerce social media platforms, test the privacy protection effect of this method, Zhang et al.'s (2022) method, Wang et al.'s (2021b) method, and Li et al.'s (2022) method on multiple sensitive attribute data in Table 1. The renderings are shown in Tables 2, 3, 4 and 5.

Table 2 The privacy protection effect of this method on user sensitive attribute data

Code	*Gender*	*Age*	*Zip code*	*Purchase preferences*
1	*	**	******	**
2	*	**	******	**
3	*	**	******	**
4	*	**	******	**
5	*	**	******	**
6	*	**	******	**
7	*	**	******	**
8	*	**	******	**

Table 3 The privacy protection effect of Zhang et al.'s (2022) method on user multi-sensitive attribute data

Code	*Gender*	*Age*	*Zip code*	*Purchase preferences*
1	*	20	230000	Books
2	*	30	100000	Electrical equipment
3	*	40	400000	Clothing
4	*	50	350000	Home furnishing
5	*	20	230000	Books
6	*	30	100000	Electrical equipment
7	*	40	400000	Clothing
8	*	50	350000	Home furnishing

Table 4 The privacy protection effect of Wang et al.'s (2021b) method on user multi-sensitive attribute data

Code	*Gender*	*Age*	*Zip code*	*Purchase preferences*
1	female	*	230000	Books
2	female	*	100000	Electrical equipment
3	female	*	400000	Clothing
4	female	*	350000	Home furnishing
5	male	*	230000	Books
6	male	*	100000	Electrical equipment
7	male	*	400000	Clothing
8	male	*	350000	Home furnishing

Table 5 The privacy protection effect of Li et al.'s (2022) method on user sensitive attribute data

Code	*Gender*	*Age*	*Zip code*	*Purchase preferences*
1	female	20	230000	*
2	female	30	100000	*
3	female	40	400000	*
4	female	50	350000	*
5	male	20	230000	*
6	male	30	100000	*
7	male	40	400000	*
8	male	50	350000	*

Comparing Tables 2, 3, 4 and 5, it can be seen that this method, Zhang et al.'s (2022) method, Wang et al.'s (2021b) method, and Li et al.'s (2022) method, can achieve encryption protection for multiple sensitive attribute data in Table 1, and all original features of the multiple sensitive attribute data are hidden. However, the Zhang et al.'s (2022) method, Wang et al.'s (2021b) method, and Li et al.'s (2022) method can only achieve single sensitive attribute data encryption protection. In contrast, this method is more suitable for user privacy protection issues on e-commerce and social media platforms.

In the privacy protection process of multiple sensitive attribute data in Table 1 using the methods in this article, Zhang et al.'s (2022) method, Wang et al.'s (2021b) method, and Li et al.'s (2022) method, the encryption time changes of the four methods are shown in Table 6.

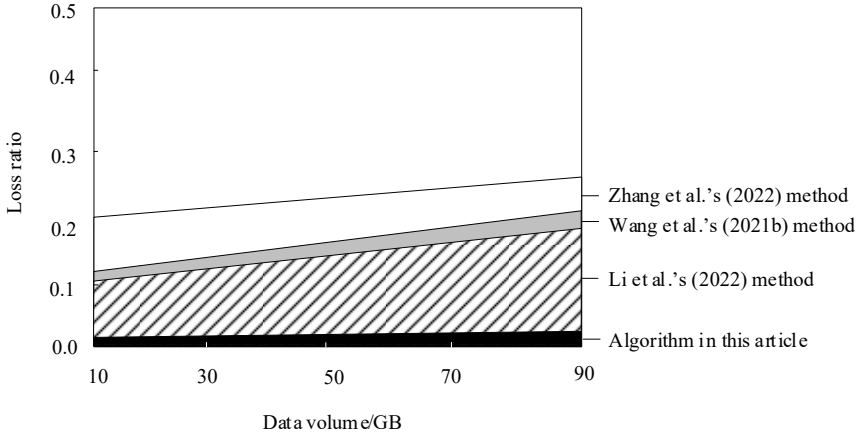
Table 6 Changes in encryption time for four methods

Code	*Encryption time /ms*			
	Proposed method	*Zhang et al.'s (2022) method*	*Wang et al.'s (2021b) method*	*Li et al.'s (2022) method*
1	34	102	392	293
2	23	112	389	233
3	33	102	409	297
4	43	104	412	283
5	25	109	390	212
6	23	108	397	232
7	44	109	387	234
8	32	98	395	234

According to the analysis of Table 6, it can be seen that the maximum encryption time of the method in this paper, Zhang et al.'s (2022) method, Wang et al.'s (2021b) method, and Li et al.'s (2022) method for privacy protection of multi-sensitive attribute data in Table 1 is 44 ms. The maximum encryption time of the Zhang et al.'s (2022) method is 112 ms, the maximum encryption time of the Wang et al.'s (2021b) method is 412 ms, the maximum encryption time of the Li et al.'s (2022) method is 297 ms. By comparison, it can be seen that the encryption speed of the method in this paper is the fastest.

Testing four methods for privacy protection of multi-sensitive attribute data shows changes in the proportion of loss of raw data, as shown in Figure 5.

Figure 5 Change in the proportion of data loss for multiple sensitive attributes



Analysing Figure 5, it can be seen that among the four methods for privacy protection of multi-sensitive attribute data, the method used in this paper has the smallest loss ratio of only 0.02, with Zhang et al.'s (2022) method, Wang et al.'s (2021b) method and Li et al.'s (2022) method gradually increasing with the increase of data volume, and the loss ratio is greater than 0.1. In contrast, after the use of the method in this article, the privacy protection effect of multi-sensitive attribute data is the best, and it will not seriously affect the original features of multi-sensitive attribute data.

After testing the method used in this article, Zhang et al.'s (2022) method, Wang et al.'s (2021b) method and Li et al.'s (2022) method, the privacy leakage status of the e-commerce social media platform is mainly reflected by the success probability of hacker anomaly extraction. The results are shown in Figures 6, 7, 8 and 9.

Figure 6 The privacy leakage status of e-commerce social media platforms after the use of this method

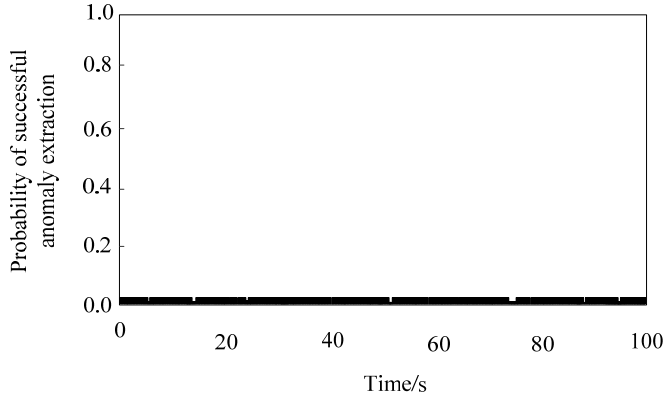


Figure 7 The privacy leakage status of e-commerce social media platforms after the use of Zhang et al.'s (2022) method

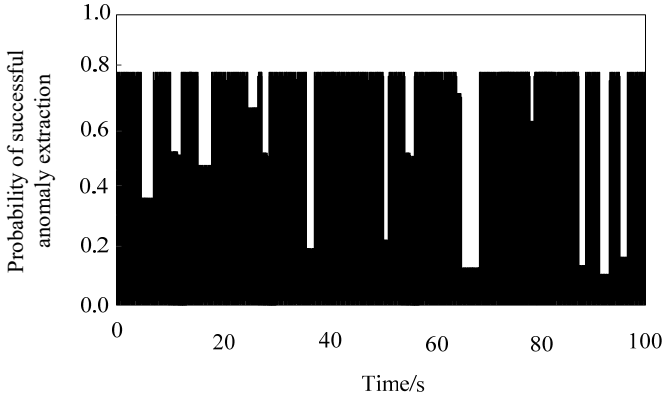


Figure 8 The privacy leakage status of e-commerce social media platforms after the use of Wang et al.'s (2021b) method

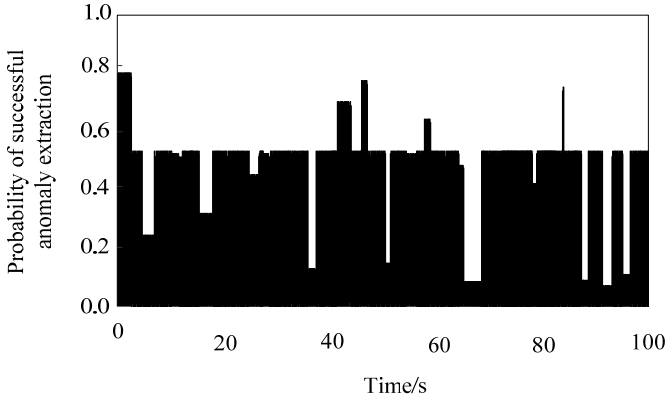
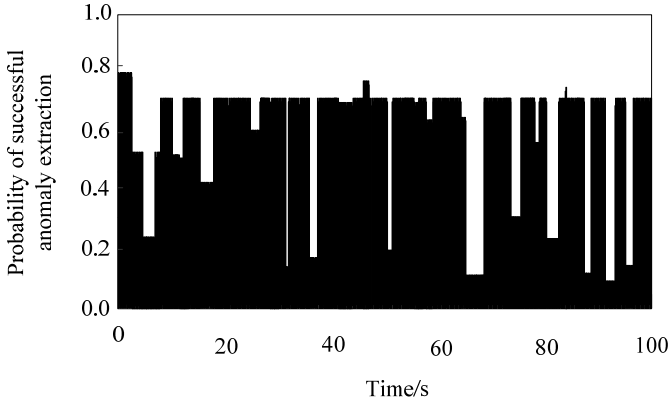


Figure 9 The privacy leakage status of e-commerce social media platforms after the use of Li et al.'s (2022) method



Compared with Figures 6, 7, 8, and 9, after the use of the method in this article, the probability of successful extraction of privacy data by hackers on the e-commerce social media platform is less than 0.05. The probability of successful extraction of data by hackers on the Zhang et al.'s (2022) method is about 0.79, the probability of successful extraction of data by hackers on the Wang et al.'s (2021b) method is about 0.52, and the probability of successful extraction of data by hackers on the Li et al.'s (2022) method is about 0.71, the method proposed in this article has a better effect on protecting the privacy of user sensitive attribute data.

4 Conclusions

After conducting research on the privacy protection of user multi-sensitive attribute data in e-commerce social media platforms, the article proposes a privacy protection method for user multi-sensitive attribute data on e-commerce social media platforms. This method first uses clustering to mine user multi-sensitive attribute data on e-commerce social media platforms, narrow the scope of privacy protection processing, and improve the efficiency of privacy protection for user multi-sensitive attribute data on e-commerce social media platforms. In the process of protecting the privacy and security of user sensitive attribute data on e-commerce social media platforms, a personalised anonymous model is used. This model can utilise multi-dimensional bucket technology, combined with the comprehensive sensitivity of user sensitive attribute data, to target and conceal low sensitivity and high sensitivity data, greatly improving the effectiveness of data privacy protection. In the experiment, the intra class distance of the multi-sensitive attribute data in this method is less than 0.1, and the inter class distance is greater than 0.6. All original features of the multi-sensitive attribute data are hidden, and the maximum encryption time is 44 ms. The loss ratio of the multi-sensitive attribute data is only 0.02, and the probability of successful extraction of privacy data by hackers is less than 0.05. Compared with similar methods, this method has more advantages in protecting the privacy of user sensitive attribute data on e-commerce social media platforms.

Acknowledgements

This work was supported by Scientific Research Project Foundation of Jilin Provincial Education Department under grant no.JJKH20231490SK.

References

Aivazpour, Z. and Rao, V. (2020) 'Information disclosure and privacy paradox: the role of impulsivity', *The Data Base for Advances in Information Systems*, Vol. 51, No. 1, pp.14–36.

Chen, X., Wu, Y. and Zhong, R. (2022) 'Optimal marketing channel and strategy in social commerce', *RAIRO – Operations Research*, Vol. 56, No. 3, pp.1203–1221.

Fadaei, S., Pooya, A. and Soleymanifard, O. (2022) 'Taxonomy of production systems with combining K-means and evolutionary algorithms', *Journal of Advanced Manufacturing Systems*, Vol. 21, No. 3, pp.515–536.

Jaleel, E.A., Anzar, S.M. and Beegum, T.R. (2021) 'System identification and control of heat integrated distillation column using artificial bee colony based support vector regression', *Chemical Engineering Communications*, Vol. 23, No. 4, pp.1–20.

Li, L., Du, H.N. and Li, T. (2022) 'Blockchain supervisable privacy protection scheme based on group signature and attribute encryption', *Computer Engineering*, Vol. 48, No. 6, pp.132–138.

Liao, Y.K., Chang, C. and Truong, G. (2020) 'Investigating B-to-B social media implementation: e-marketing orientation and media richness perspective', *Journal of Electronic Commerce in Organizations*, Vol. 18, No. 1, pp.18–35.

Liu, G., Fei, S. and Yan, Z. (2020) 'An empirical study on response to online customer reviews and e-commerce sales: from the mobile information system perspective', *Mobile Information Systems*, Vol. 83, No. 1, pp.1–12.

Sun, Y., Zhang, F. and Feng, Y. (2022) 'Do individuals disclose or withhold information following the same logic: a configurational perspective of information disclosure in social media', *Aslib Journal of Information Management*, Vol. 74, No. 4, pp.710–726.

Wang, L., Xu, X. and Zhao, X. (2021a) 'A randomized block policy gradient algorithm with differential privacy in content centric networks', *International Journal of Distributed Sensor Networks*, Vol. 17, No. 12, pp.1385–1394.

Wang, X., Yang, Y. and Chen, M. (2020) 'AGNES-SMOTE: an oversampling algorithm based on hierarchical clustering and improved SMOTE', *Scientific Programming*, Vol. 32, No. 2, pp.1–9.

Wang, Y., Liang, X. and Hei, X. (2021b) 'Deep learning data privacy protection based on homomorphic encryption in AIoT', *Mobile Information Systems*, Vol. 15, No. 2, pp.1–11.

Wang, Y.L., Fan, W., Yan, Z.Y. and Zhang, L. (2022) 'Effect of back protection process on EH oil pipe joint performance', *Ordnance Material Science and Engineering*, Vol. 45, No. 2, pp.119–123.

Xiang, D. and Zhang, Z. (2020) 'Cross-border e-commerce personalized recommendation based on fuzzy association specifications combined with complex preference model', *Mathematical Problems in Engineering*, Vol. 20, No. 4, pp.1–9.

Zeng, Y., Song, S. and Peng, W. (2022) 'Optimal add-on items recommendation service strength strategy for e-commerce platform with full-reduction-promotion', *RAIRO – Operations Research*, Vol. 56, No. 2, pp.1031–1049.

Zennyo, Y. (2020) 'Strategic contracting and hybrid use of agency and wholesale contracts in e-commerce platforms', *European Journal of Operational Research*, Vol. 281, No. 1, pp.231–239.

Zhang, S.B., Yuan, L.J., Mao, X.J. and Zhu, G.M. (2022) 'Privacy protection method for K-modes clustering data with local differential privacy', *Acta Electronica Sinica*, Vol. 50, No. 9, pp.2181–2188.

Zhao, H. (2021) 'A cross-border e-commerce approach based on blockchain technology', *Mobile Information Systems*, Vol. 21, No. 4, pp.1–10.

Zheng, J., Yang, X.Y., Yu, J. and Li, X. (2022) 'Semi supervised recommendation of mobile terminal big data based on artificial bee colony', *Computer Simulation*, Vol. 39, No. 7, pp.497–501.

Zhou, X.Y., Hu, J.C., Wu, Y.L., Zhong, M.S. and Wang, M.W. (2022) 'A multi-strategy artificial bee colony algorithm based on fitness grouping', *Pattern Recognition and Artificial Intelligence*, Vol. 35, No. 8, pp.688–700.