Indian Journal of Science and Technology, *Vol 9(2), DOI: 10.17485/ijst/2016/v9i2/81159, January 2016*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Social Ant based Sensitive Item Hiding with Optimal Side Effects for Data Publishing

P. Tamil Selvan* and S. Veni

Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore - 641021, Tamil Nadu, India; tmselvanin@gmail.com, venikarthik04@gmail.com

Abstract

Background/Objectives: This paper proposes an Optimized Social Ant Based Sensitive Item Hiding (OSA-SIH) technique and expands the scope of quality privacy preservation for distributed data mining with optimal side effects on the original dataset. **Methods/Statistical Analysis:** in OSA-SIH technique, initially sensitive items for the given distributed dataset are evaluated using the social ant based relative item set distribution. Based on the evaluated dataset, optimal hiding of sensitive item is arrived with social ant based relative item set distribution even for larger item sets, ensuring time for optimal hiding. Next, sensitive item hiding is performed through multiplicative and transformational data perturbation. This data perturbation is based on socially cohesive relational rate between sensitive and non sensitive item sets, ensuring privacy preservation accuracy. The side effects on the modified dataset are checked for several users' requested item set distribution. **Findings:** The experimental results demonstrated that proposed technique out performed than the existing state of the art works in terms of privacy preservation accuracy, rate of side effects on the modified dataset, and time for optimal hiding. **Improvement/Application:** Experiments revealed that the proposed OSA-SIH technique is able to reduce the rate of side effects on modified dataset as compared to the state-of-the-art works.

Keywords: Perturbation, Privacy Preserving Data Mining, Social Ant, Sensitive Item Hiding, Transformational Data Perturbation, Multiplicative Data Perturbation

1. Introduction

One of the widely used approaches by data miners is data perturbation for Privacy Preserving Data Mining (PPDM). Multilevel Trust in Privacy Preserving Data Mining (MT-PPDM)[1] used the concept of aggregated data without the possibility of accessing the information by the third parties, ensuring trust level. Reducing Side Effects in Privacy Preserving Data Mining (RSE-PPDM)[2] used hiding missing artificial utility algorithm to minimize the number of deleted transactions and number of side effects.

In[3], homomorphic matching technique was used for privacy preservation and improved the privacy level. Secrecy views and null based virtual updates were used in[4] for achieving data privacy with the objective of reducing the computation cost. Direct and indirect discrimination was performed in[5] using legitimate classification rules while preserving data quality. Though a number of side effects and privacy level was improved but at the cost of accuracy. The privacy preservation accuracy is improved in OSA-SIH technique using multiplicative and transformational data perturbation approach.

With the introduction of cloud environment, the data outsourcing and computing services is receiving greater attention than never before. In[6], a privacy preserving mining scheme was developed to introduce scalability and achieved privacy in a large scale. Secure mining of association rules was performed in[7] using Fast Distributed Mining (FDM) algorithm resulting in

**Author for correspondence*

reduced communication and computation cost. Privacy policy on content sharing sites was introduced in[8] with the objective of improving the prediction accuracy through Adaptive Privacy Policy Predictive (APPP) system.

In[9], privacy preserving and content protection was performed to address security issues using Oblivious Transfer and Private Information Retrieval (PIR). To address access control mechanism in[10], Privacy Protection Mechanism (PPM) was introduced using k-anonymity and l-diversity. Though security and privacy was ensured in all the above methods, but rate of side effects remained unsolved. The rate of side effects is minimized in OSA-SIH technique by applying correlation-based privacy preserving.

Publishing of micro-data is one of the most important issues in privacy preserving. In[11], a new method to preserve privacy for data publishing called slicing was introduced for better data utility. Another method called m-privacy[12] for data publishing was introduced to ensure anonymity. In[13], access control for cloud based on privacy preserving was introduced using group key management scheme. Another method used cryptographic techniques[14] to solve the issues related to confidentiality and security through fine grained attribute based access control policies.

In[15], distributed mining of association rules was performed using cryptographic techniques that resulting in reducing the overhead. Anonymous publication of sensitive transactional data[16] was performed using approximate nearest neighbor addressing the issues related to data utility and execution time. Privacy preservation for Online Analytic Processing (OLAP)[17] was addressed through mixed aggregate functions with aiming at providing effective privacy protection.

In[18], privacy preservation was applied in health data collection using identity-based encryption protocol resulting in the improvement of privacy and correctness of the protocol. In[19], efficient clustering were applied with the objective of improving the computational performance and at the same time reducing the computational cost through Fractional Calculus. Hierarchical K-Means Clustering[20] was applied on horizontally partitioned data with aiming at improving the communication cost.

A novel model[21] was designed to describe intrusiveness in the context of personalized digital marketing campaigns. An attribute segregation and perturbation frame work[22] was developed for Multi-Trust Level scenario. However, perturbation of high ranked attributes does not have much effect on the utility of the datasets. The effect of security policies, security awareness and individual characteristics of medical institution employees on security effectiveness are addressed in[23]. A highly scalable parallel BUG approach[24] was designed using Map reduces on cloud to address scalability problem of large-scale data anonymization by Bottom-Up Generalization (BUG) but performing a generalization maybe changes the anonymity of the data set and privacy gain of each candidate will be affected. Another method[25] was developed to discover the hidden useful information from the process data with which results in improvement of quality of the product.

In this paper, an Optimized Social Ant Based Sensitive Item Hiding (OSA-SIH) technique is proposed for distributed data mining to obtain quality privacy preservation with optimal side effects on the original dataset. This is performed using user operational conditions-based sensitive items, social ant-based relative item set distribution and Ant-based based Orthogonal Multiplicative and Transformational algorithm.

Experimental results showed that the AOMT algorithm has good performance in privacy preservation accuracy, optimal time hiding and ensuring high quality privacy preservation of the data items of corresponding user's privileges. Besides, the proposed algorithm can thus generate minimal side effects on the modified dataset compared to the state-of-the-art works for hiding sensitive item sets.

2. Optimized Social Ant based Sensitive Item Hiding

Some applications require protection against the disclosure of private, confidential, or secure data. In this section, an efficient technique called Optimized Social Ant Based Sensitive Item Hiding (OSA-SIH) for data publishing is designed with the objective of improving the privacy preservation accuracy and minimizing the rate of side effects on the modified dataset at relatively lesser amount of time. The elaborate design of OSA-SIH technique is given below.

2.1 Design of user Operational Conditions-based Sensitive Items

The first step in the design of Optimized Social Ant Based Sensitive Item Hiding (OSA-SIH) technique is the effective construction of user operational

conditions-based sensitive items. In this section, the problem of sensitive item hiding for privacy preservation in distributed dataset is evaluated. It is performed based on the user operational conditions, by proposing a system to measure the global frequent item sets for distributed data item being shared.

This is done by designing an algorithm that hides sensitive frequent items from the global frequent items. The optimal hiding of sensitive item is arrived with social ant based relative item set distribution in the corresponding original dataset even for larger item sets. Figure 1 shows the block diagram of user operational conditions-based sensitive items. The block diagram includes two main components, where the support and confidence values are evaluated for sensitive item hiding with aiming at reducing the time for sensitive hiding.

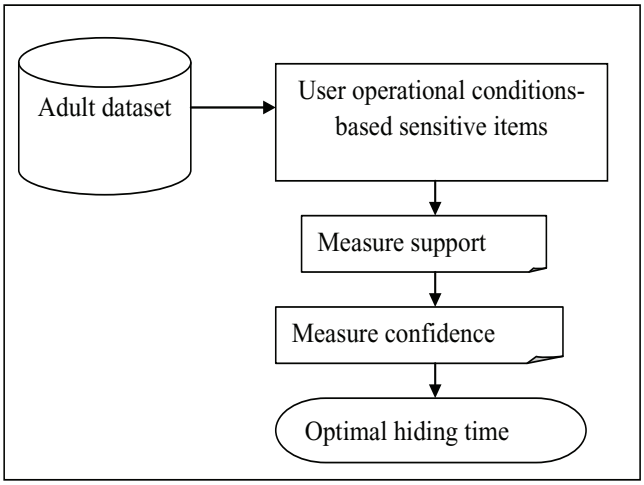

Figure 1. Block diagram of user operational conditions-based sensitive items.

Let us consider a dataset '*D*' where '$I = I_1, I_2, \ldots I_n$' represents the items, consisting of '*n*' transaction comprises of the set of items in such a way that '$T \in I$'. Then, the association rule is of the form

$$P \rightarrow Q, \text{ where } P \in I \,\&\, Q \in I \qquad (1)$$

Where '*P*' and '*Q*' are said to be the antecedent and consequent of rule respectively. Relative strength of an item with respect to its strong or weak nature is evaluated using two factors namely, support and confidence of the item. The first factor to be measured for sensitive item hiding is the support and is mathematically formulated as given below.

$$S(P \rightarrow Q) = S(P \cup Q) = ((P \cup Q)/n) \qquad (2)$$

From (2), support '*S*' measures the proportion of transactions that includes both '*P*' and '*Q*' respectively, with '*n*' denoting the total number of transactions involved during sensitive item hiding. The second factor to be measured for sensitive item hiding is the confidence formulated as given below.

$$C(P \rightarrow Q) = ((P \cup Q)/P) = (S(P \cup Q)/S(P)) \qquad (3)$$

From (3), the confidence 'C' is the percentage for a transaction that contains 'P' also contains 'Q'. A rule is significant if its support and confidence are higher than the user designated Support Threshold Value (*STV*) and Confidence Threshold Value (*CTV*). As a result, using the ant-based relative item set distribution algorithm not all the items are retrieved, but only a very small member that satisfies the '*STV*' and '*CTV* 'are retrieved. In this way, the time for optimal hiding is significantly reduced.

2.2 Design of Social Ant-based Relative Item Set Distribution

The second step in the design of Optimized Social Ant Based Sensitive Item Hiding (OSA-SIH) technique is the construction of social ant-based relative item set distribution. In order to arrive at optimal hiding of sensitive item and hide sensitive data item, the proposed OSA-SIH technique uses social ant based relative item set distribution in the corresponding original dataset even for larger item sets.

Hiding of sensitive item is done through multiplicative and transformational data perturbation technique based on socially cohesive relational rate between sensitive and non sensitive item sets of the original dataset to generate a modified dataset 'MD'.

Figure 2 shows the block diagram of Ant-based Orthogonal Multiplicative and Transformational Data Perturbation. The basic concept of ant principle is that the random wandering nature and upon successful identification of food return to their colony while laying down pheromone trails. On the other hand, if other ants identify those paths, the ants again do not traverse at random manner, but it blindly follows the trail provided by the earlier ants. In a similar manner, if the items in the transaction is said to occur repeatedly, then it is said to be a sensitive item. By changing the item in a random manner with the aid of probability functions, frequent sensitive items are hidden in an efficient manner.

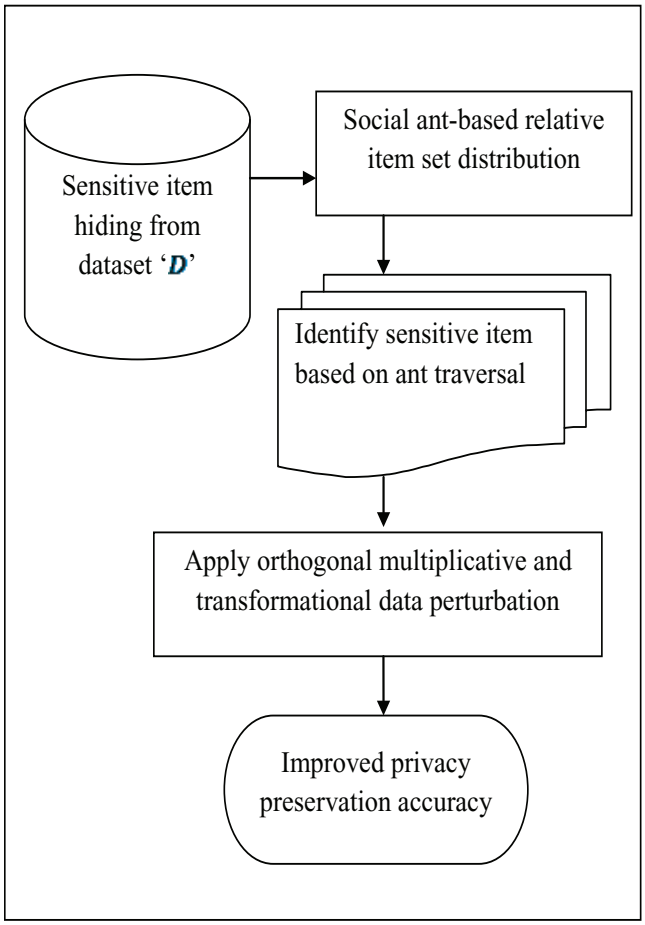

Figure 2. Block diagram of Ant-based Orthogonal Multiplicative and Transformational Data Perturbation.

Let us set '$\alpha\ (x_a)$' as the pheromone intensity of the 'D' ant (i.e. Dataset) which is at position 'x_i' and be initialized as a constant. Then the probability of the dataset 'D' that hide sensitive item from 'x_a' to 'x_b' is mathematically formulated as given below.

$$p_{ab} = (\alpha\ (x_b)/\alpha\ (x_a)), \text{ where } x_b, x_a \in D \qquad (4)$$

From (4), the position of the dataset provides a solution of the problem for optimal hiding of sensitive item based on social ant based relative item set. Once the optimal sensitive item is obtained, the proposed OSA-SIH performs the task of hiding sensitive item through multiplicative and transformational data perturbation technique. This multiplicative and transformational data perturbation technique is based on sensitive and non sensitive item sets of the original dataset through which a modified dataset is generated.

The objective behind the use of orthogonal multiplicative and transformational data perturbation is to improve the privacy preservation accuracy during the data perturbation process. The orthogonal multiplicative and transformational data perturbation in proposed OSA-SIH technique uses orthogonal transformation.

Let us consider two datasets 'L' and 'M' of size '$i * n$ *matrix*' and '$j * n$ *matrix*' respectively with orthogonal matrix represented as 'O'. Now the mathematical formulation for the orthogonal multiplicative and transformational data perturbation for two datasets 'L' and 'M' is as given below

$$A = LO; \qquad B = MO \qquad (5)$$

$$AA^T = LL^T; \qquad BB^T = MM^T \qquad (6)$$

$$AB^T = L\ OO^T\ B^T = LM^T \qquad (7)$$

From (5), (6) and (7), by applying an orthogonal matrix based on socially cohesive relational rate between sensitive and non sensitive item sets, all the pair distances and similarities from column vectors 'A and B' are preserved in an efficient manner in the perturbed data. At the same time, both the sensitive and non sensitive items and the transformation process are kept secret, whereas only the perturbed data is viewed by the third user. As a result, the privacy preservation accuracy is improved in a significant manner. Figure 3 shows the ant-based orthogonal multiplicative and transformational algorithm.

The Ant-based based Orthogonal Multiplicative and Transformational (AOMT) algorithm includes four main steps. The first step measures the support for sensitive item hiding. The second step evaluates the confidence value for sensitive item hiding. Next, a comparison is made between the support threshold '*STV*' and confidence threshold '*CTV*' with the evaluated confidence '*C*' and support value '*S*'. Followed by this, optimal hiding of sensitive item and orthogonal multiplicative and transformational process is performed. If the values of support '*S*' and confidence '*C*' are less than the support threshold '*STV*' and confidence threshold '*CTV*' respectively, item hiding is performed, otherwise, the same operation is performed with other transactions. In this way, privacy preservation accuracy is ensured in an efficient manner.

2.3 Correlation-based Privacy Preserving

The third step in the design of Optimized Social Ant Based Sensitive Item Hiding (OSA-SIH) technique is privacy

Input: Dataset 'D', Items ' $I = I_1, I_2, \ldots, I_n$', Support Threshold Value (STV), Confidence Threshold Value (CTV)

Output: optimized sensitive item hiding

Step 1: **Begin**
Step 2: **For** each Dataset 'D'
Step 3: **For** each Items 'I'
Step 4: Evaluate support 'S' for sensitive item hiding using ()
Step 5: Evaluate confidence 'C' for sensitive item hiding using ()
Step 6: **If** '$S < STV$' and ' $C < CTV$'
Step 7: Evaluate optimal hiding of sensitive item using ()
Step 8: Evaluate orthogonal multiplicative and transformational process using ()
Step 9: **else**
Step 10: go to step 2
Step 10: **End if**
Step 10: **End for**
Step 11: **End for**
Step 12: **End for**
Step 13: **End**

Figure 3. Ant-based Orthogonal Multiplicative and Transformational algorithm.

preservation through correlation-based approach. The side effects of hiding item sets on the modified dataset are checked for various user requested item set distribution on the privacy preserving distributed data mining, which in turn improve the user trust level.

In order to consider hidden sensitive items and the modified entries (i.e. dataset), the proposed OSA-SIH technique considers the correlation between the items with aiming at minimizing the rate of side effects on the modified dataset. The following lists the two considerations in hiding a sensitive item

Let us consider three transactions 'p', 'q' and 'r', when correlating with other sensitive item, if '$r \otimes q$', then removing the item 'q' is better than removing 'p'as removing the former 'r' affects both items. Next, when correlating with a non sensitive item, if '$q \otimes p$', then inserting 'p' into the transactions that do not contain 'q' is better than deleting 'p' or 'q' from the transactions. As a result, rate of side effects on the modified dataset is reduced in an extensive manner.

3. Experimental Settings

Optimized Social Ant Based Sensitive Item Hiding (OSA-SIH) technique is developed for data publishing using JAVA platform. The OSA-SIH technique uses the Adult data set from the University of California Irvine data repository that contains information on individuals such as age, level of education and current employment type.

The dataset used in this work has forty nine thousand records and also binomial label that indicates the salary of less or greater than fifty thousand US dollars, referred to as <50K or >50K in this work. The data for experimental purpose has been divided into a training dataset containing thirty two thousand records and a test dataset containing sixteen thousand records.

There are fourteen attributes consisting of seven polynomials, one binomial and six continuous attributes and are used in the OSA-SIH technique to preserve the privacy of certain attributes including salary, relationship and marital status. The employment class attribute denotes the employer type (i.e. self employed or federal) and occupation refers to the employment type (i.e. farming or managerial). The education attribute comprises high school graduate or doctorate. The relationship attribute includes the information related to unmarried or married.

The final nominal attributes are country of residence, gender and race. The continuous attributes are age, hours worked per week, education number, capital gain and loss and a survey weight attribute assigned to an individual based on information such as area of residence and type of employment. The performance of the OSA-SIH technique is evaluated for parameters such as number of transactions, size of transaction, privacy preservation accuracy, rate of side effects on the modified dataset, and time for optimal hiding.

The privacy preservation accuracy measures the ratio of privacy preserved perturbed copies to the total number of perturbed copies taken for experimental evaluation. The mathematical formulation of privacy preservation accuracy is given below.

$$A = ((\text{privacy preserved perturbed copies}) / n) * 100 \quad (8)$$

From (8), the privacy preservation accuracy 'A' is measured with respect to the number of perturbed copies 'n' and is measured in terms of percentage (%). Higher privacy preservation accuracy, the more efficient method is said to be.

The time for optimal hiding is measured based on the total number of transactions and the time required for single transaction. The time for optimal hiding is measured in terms of milliseconds (ms) and is formulated as given below.

$$Time = n * Time\ (item\ hiding\ for\ single\ transaction) \quad (9)$$

From (9), the time for optimal hiding '*Time*' is measured with respect to the total number of transactions '*n*'. Lower the time for optimal hiding more efficient the method is said to be. The rate of side effects measures the difference between the actual size of transaction and the modified dataset generated during privacy preserving. The mathematical formulation for rate of side effects is given below.

$$RoSE = (Size - MD) \quad (10)$$

From (10), the rate of side effects '*RoSE*' is measured on the basis of the size of transaction '*Size*', modified dataset '*MD*' respectively. It is measured in terms of kilobytes (KB).

4. Discussion

The Optimized Social Ant Based Sensitive Item Hiding (OSA-SIH) technique is compared against the existing Multilevel Trust in Privacy Preserving Data Mining (MT-PPDM)[1] and Reducing Side Effects in Privacy Preserving Data Mining (RSE-PPDM)[2]. The experimental results using JAVA are compared and analyzed through table and graph form as given below.

4.1 Impact of Privacy Preservation Accuracy

To support transient performance, in Table 1 we apply an Ant-based Orthogonal Multiplicative and Transformational algorithm and comparison made with two other existing methods namely MT-PPDM and RSE-PPDM. Figure 4 shows that the Optimized Social Ant based Sensitive Item Hiding (OSA-SIH) technique provides higher privacy preservation accuracy when compared to MT-PPDM[1] and RSE-PPDM[2]. The privacy preservation accuracy is increased with the application of social ant-based relative item set distribution with the aid of probability functions, frequent sensitive items are hidden in an efficient manner.

Table 1. Tabulation for privacy preservation accuracy

Age (Number of perturbed copies)	Privacy preservation accuracy (%)		
	OSA-SIH	MT-PPDM	RSE-PPDM
10	69.35	59.48	49.31
20	71.49	65.46	60.40
30	73.55	67.52	62.46
40	70.28	64.25	59.19
50	74.97	68.94	63.88
60	78.32	72.39	67.33
70	82.45	76.42	71.36

Figure 4 shows the privacy preservation accuracy rate efficiency with age taken as the attribute (i.e. number of perturbed copies generated) for distributed data mining. With the application of orthogonal multiplicative and transformational data perturbation, the data perturbation is performed in an efficient manner on the basis of orthogonal matrix with training samples. This in turn helps in improving the privacy preservation accuracy for distributed data mining using OSA-SIH by 8.91% compared to MT-PPDM[1]. Moreover, the OSA-SIH technique by applying multiplicative and transformational perturbation technique that takes multiple instances helps in increasing the privacy preservation accuracy by 16.79% compared to RSE-PPDM[2].

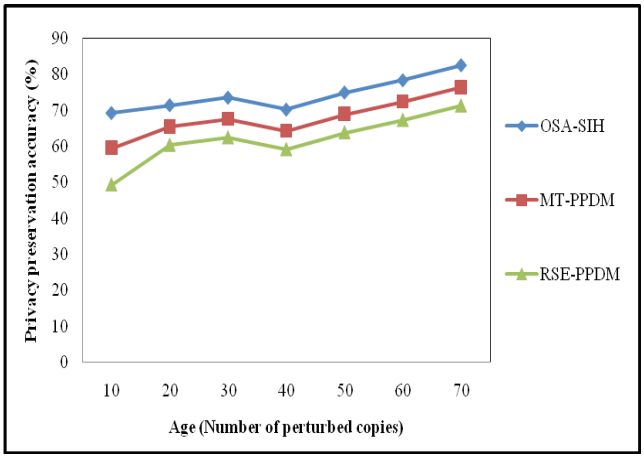

Figure 4. Measure of privacy preservation accuracy.

4.2 Impact of Time for Optimal Hiding

The comparison of time for optimal hiding is presented in Table 2 with respect to the total number of transactions in the range of 5 – 35 collected at different time stamps from the adult dataset records. With increase in the number of

transactions, the time for optimal hiding is also increased though not observed to be linear. This is because of the different types and nature of the transaction, the time for optimal hiding also gets varied.

Table 2. Tabulation for time for optimal hiding

Total number of transactions (n)	Time for optimal hiding (ms)		
	OSA-SIH	MT-PPDM	RSE-PPDM
5	2.22	2.65	3.55
10	3.85	4.15	5.30
15	5.55	5.85	6.90
20	7.93	8.20	9.35
25	5.25	5.55	6.70
30	8.35	8.65	9.75
35	11.21	11.51	12.61

To ascertain the performance of the time for optimal hiding, comparison is made with two other existing methods Multilevel Trust in Privacy Preserving Data Mining (MT-PPDM)[1] and Reducing Side Effects in Privacy Preserving Data Mining (RSE-PPDM)[2].

In Figure 5, the number of transactions is varied between 5 and 35, where the transactions involve the hiding of sensitive item salary and marital status at different time intervals. From the Figure it is illustrative that the time for optimal hiding is less using the proposed OSA-SIH technique when compared to the two other existing methods.

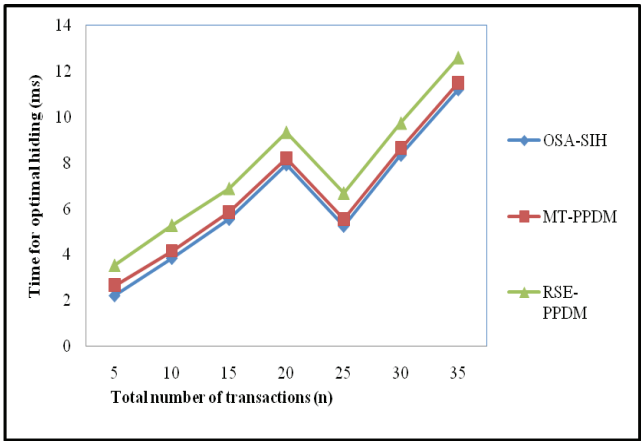

Figure 5. Measure of time for optimal hiding.

The time for optimal hiding is reduced by applying of user operational conditions-based sensitive items in OSA-SIH. With the application of user operational conditions-based sensitive items, where the support and confidence values are evaluated for sensitive item hiding that provides the results with respect to total number of transactions reducing the time for optimal hiding by 6.85% compared to MT-PPDM[1]. Besides, by applying the association rules and comparing with the user designated Support Threshold Value (*STV*) and Confidence Threshold Value (*CTV*) minimizes the time for optimal hiding by 28.09% compared to RSE-PPDM[2].

4.3 Impact of Rate of Side Effects

The rate of side effects to obtain quality privacy preservation for distributed data mining using OSA-SIH, MT-PPDM and RSE-PPDM is elaborated in Table 3. We consider the technique with differing size of transaction in the range of 100KB to 700KB for experimental purpose using JAVA.

Table 3. Tabulation for rate of side effects

Size of transaction (KB)	Rate of side effects (KB)		
	OSA-SIH	MT-PPDM	RSE-PPDM
100	68	72	74
200	75	83	87
300	90	98	102
400	103	111	115
500	120	128	132
600	130	138	142
700	145	153	157

In Figure 6, we depict the rate of side effects while generating a modified dataset from an original dataset with size of transaction range from 100 KB to 700 KB for the purpose of experiment. From the figure, the rate of side effects resulted using the proposed OSA-SIH technique is lower when compared to two other existing methods MT-PPDM[1] and RSE-PPDM[2]. Besides it can also be observes that by increasing the size of transaction, the rate of side effects is also increased using all the methods. But comparatively, it is lower using OSA-SIH technique.

Figure 6 as shown measures the rate of side effects generated when original dataset is transformed to modified dataset with the objective of hiding certain items (i.e. attributes) for the purpose of privacy preservation. The rate of side effects of hiding item sets on the modified dataset are verified for various user requested item set distribution on the privacy preserving distributed data

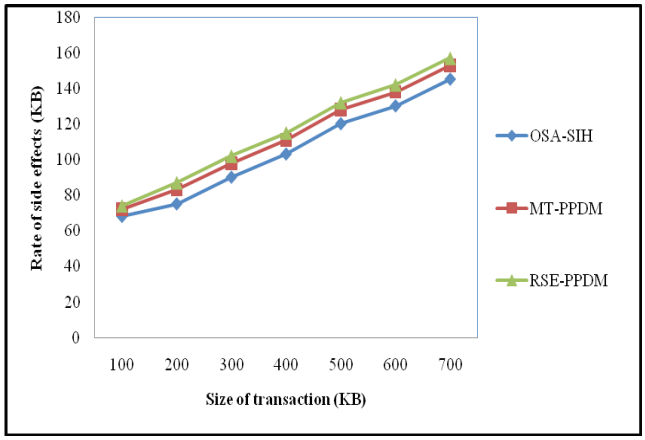

Figure 6. Measure of rate of side effects.

mining, which in turn also improves the user trust level. By applying correlation based privacy preserving, the rate of side effects is minimized using OSA-SIH by 7.36% compared to MT-PPDM[1]. In addition, by following the two considerations in hiding a sensitive item using correlation-based approach, the rate of side effects is reduced by 11.04% compared to RSE-PPDM[2].

5. Conclusion

An Optimized Social Ant Based Sensitive Item Hiding (OSA-SIH) technique with scope of quality privacy preservation for distributed data mining with optimal side effects on original dataset has been designed. The objective of providing such a design is to ensure high quality privacy preservation of the data items of corresponding user's privileges for distributed data and to decrease the time for optimal hiding for various user requested item set distribution. A user operational conditions-based sensitive item are designed as a measure for identifying the support and confidence value and proposed a proposed a system to measure the global frequent item sets for distributed data item being shared based on user query. The proposed social ant-based relative item set distribution provides privacy preservation accuracy for large item sets through multiplicative and transformational data perturbation technique. In addition, Ant-based based Orthogonal Multiplicative and Transformational algorithm with probability function help in improving the privacy preservation accuracy. Experimental evaluation is conducted with the Adult Data Set extracted from UCI repository to provide high quality privacy preservation of data items and measured the performance in terms of privacy preservation accuracy, optimal time hiding and rate of side effects on answering user query requests. Performances results reveal that the proposed OSA-SIH technique provides higher level of privacy preservation accuracy efficiency and also strengthen the optimal time hiding on high dimensional dataset. The proposed OSA-SIH technique provides 12.85% high rate of privacy preservation accuracy and minimizes the time for optimal hiding by 17.47% when compared to state of the art works.

6. References

1. Li Y, Chen M, Li Q, Zhang W. Enabling multilevel trust in privacy preserving data mining. IEEE Transactions on Knowledge and Data Engineering. 2012 Sep; 24(9):1598–612.
2. Lin CW, Hong TP, Hsu HC. Reducing side effects of hiding sensitive item sets in privacy preserving data mining. The Scientific World Journal. 2014; 2014:12.
3. Karapiperis D, Verykios VS. An LSH-based blocking approach with a homomorphic matching technique for privacy-preserving record linkage. IEEE Transactions on Knowledge and Data Engineering. 2015 Apr; 27(4):909–21.
4. Bertossi L, Li L. Achieving data privacy through secrecy views and null-based virtual updates. IEEE Transactions on Knowledge and Data Engineering. 2013 May; 25(5):987–1000.
5. Hajian S, Ferrer JD. A methodology for direct and indirect discrimination prevention in data mining. IEEE Transactions on Knowledge and Data Engineering. 2013 Jul; 25(7):1445–59.
6. Giannotti F, Lakshmanan VS, Monreale A, Pedreschi D, Wang HW. Privacy-preserving mining of association rules from outsourced transaction databases. IEEE Systems Journal. 2013 Sep; 7(3):385–95.
7. Tassa T. Secure mining of association rules in horizontally distributed databases. IEEE Transactions on Knowledge and Data Engineering. 2014 Apr; 26(4):970–83.
8. Squicciarini AC, Lin D, Sundareswaran S, Wede J. Privacy policy inference of user-uploaded images on content sharing sites. IEEE Transactions on Knowledge and Data Engineering. 2015 Jan 1; 27(1):193–206.
9. Paulet R, Md Kaosar G, Yi X, Bertino E. Privacy-preserving and content-protecting location based queries. IEEE Transactions on Knowledge and Data Engineering. 2014 May; 26(5):1200–10.
10. Pervaiz Z, Walid G, Ghafoor A, Prabhu N. Accuracy-constrained privacy-preserving access control mechanism for relational data. IEEE Transactions on Knowledge and Data Engineering. 2014 Apr; 26(4):795–807.

11. Li T, Li N, Zhang J, Molloy I. Slicing: A new approach to privacy preserving data publishing. IEEE Transactions on Knowledge and Data Engineering. 2012 Mar; 24(3):561–74.
12. Goryczka S, Xiong L, Fung BCM. m-Privacy for Collaborative Data Publishing. 7th International Conference on Collaborative Computing: Networking, Applications and Work sharing (Collaborate.Com). 2011 Oct 15-18. p. 1–10.
13. Nabeel M, BertiIssue E. Privacy preserving delegated access control in public clouds. IEEE Transactions on Knowledge and Data Engineering. 2014 Sep; 26(9):2268–80.
14. Nabeel M, Bertino E. Privacy-preserving fine-grained access control in public clouds. IEEE Computer Society Technical Committee on Data Engineering. 2012 Dec; 35(4):1–10.
15. Glu MK, Clifton C. Privacy-preserving Distributed Mining of Association Rules on Horizontally Partitioned Data. 12th International Conference on Hybrid Intelligent Systems (HIS). 2012. p. 2–13.
16. Ghinita G, Kalnis P, Tao Y. Anonymous publication of sensitive transactional data. IEEE Transactions on Knowledge and Data Engineering. 2014 Sep; 23(2):161–74.
17. Zhang N, Zhao W. Privacy-preserving OLAP: An information-theoretic approach. IEEE Transactions on Knowledge and Data Engineering. 2011 Jan; 23(1):122–38.
18. Guan S, Zhang Y, Ji Y. Privacy-preserving health data collection for preschool children. Computational and Mathematical Methods in Medicine. 2013 Sep; 2013:5.
19. Bhaladhare PR, Jinwala DC. A clustering approach for the *l*-diversity model in privacy preserving data mining using fractional calculus-bacterial foraging optimization algorithm. Advances in Computer Engineering. 2014 Sep; 2014:12.
20. Xue A, Jiang D, Ju S, Chen W, Ma H. Privacy-preserving hierarchical-k-means clustering on horizontally partitioned data. International Journal of Distributed Sensor Networks. 2009; 5(1):81.
21. Mamlouk L, Segard O. Big data and intrusiveness: Marketing issues. Indian Journal of Science and Technology. 2015 Feb; 8(S4):189–93.
22. Priyadarsini1 RP, Valarmathi ML, Sivakumari S. Attribute segregation based on feature ranking framework for privacy preserving data mining. Indian Journal of Science and Technology**.** 2015 Aug; 8(17): 59772.
23. Kim S, Jeoung K. Effects of security policies, security awareness of hospital employee to patients' personal information protection. Indian Journal of Science and Technology**.** 2015 Sep; 8(21):IPL0244.
24. Irudayasamy A, Arockiam L. Parallel bottom-up generalization approach for data anonymization using map reduce for security of data in public cloud. Indian Journal of Science and Technology. 2015 Sep; 8(22):IPL0251.
25. Shankar R, Sundararajan M. Manufacturing quality improvement with data mining outlier approach against conventional quality measurements. Indian Journal of Science and Technology. 2015 Jul; 8(15):IPL037.