CHAPTER 4

OPTIMIZED SOCIAL ANT BASED SENSITIVE ITEM HIDING FOR THE PRIVACY PRESERVATION OF DATA PUBLISHING

4.1 INTRODUCTION

Privacy preserving of data publishing is attained great attention for transmitting the data while preserving the individual privacy. Privacy Preserving Data Mining (PPDM) is essential for developing the multiple databases including the personal or sensitive data without discourse the information. The PPDM preserves the secrete information from larger dataset while improving the sensitive item hiding for the data publishing. Then the PPDM is significant to decrease the privacy threats through hiding the sensitive information while mining the required data from the databases. Though, the privacy preserving data mining becomes more challenging problem in resolving the effects of privatizing the user’s data. Therefore, there is a requirement for new model to improve the privacy preservation and to protect data against the disclosure.

Existing Hiding-Missing-Artificial Utility algorithm is used to design the lattice-based approach for hiding the sensitive information during the item set deletion. The transaction with enhanced ratio of sensitive to non sensitive for selecting the entire deletion and reduced side effects using a scalable transaction deletion. However, the Hiding-Missing-Artificial Utility removes the estimation of every large item sets and unable to handle the highest frequency in sensitive rules regarding with the present sensitive transaction. On the other hand, the perturbation-based PPDM to multilevel trust technique enables the flexibility and provides a perturbed copies of data for different trust levels.
Although the trust level of new request is lower and unfeasible to maintain the trust levels for the data to be extracted.

An Optimized Social Ant Based Sensitive Item Hiding (OSA-SIH) technique is proposed for the distributed data mining to achieve a better privacy preservation with optimal side effects on the original dataset. Initially, the sensitive items in distributed datasets are identified with the help of social ant based relative item set distribution. During the evaluated dataset, the optimal hiding of sensitive item is obtained based on social ant based relative item set distribution to larger item sets which ensures the time for optimal hiding. Next, the sensitive item hiding is attained through the multiplicative and transformational data perturbation process in an optimized manner. Thus the data perturbation is mainly depends upon socially cohesive relational rate between the sensitive and non sensitive item sets that ensures the privacy preservation accuracy. Finally, the side effects on the modified dataset investigate different users’ requested item set distribution within the minimum amount of time using the OSA-SIH technique. Therefore, the performance of proposed OSA-SIH technique provides a higher level of privacy preservation accuracy and also reduces the optimal hiding time on high dimensional dataset while hiding the sensitive item sets in an efficient manner.

4.2 OPTIMIZED SOCIAL ANT BASED SENSITIVE ITEM HIDING (OSA-SIH) FRAMEWORK

An efficient framework called Optimized Social Ant Based Sensitive Item Hiding (OSA-SIH) technique is established to maximize the privacy preservation accuracy for data publishing. Besides, the proposed OSA-SIH technique decreases the rate of side effects on the modified dataset at relatively lesser time interval. This is obtained based on user
operational conditions-based sensitive items, social ant-based relative item set distribution and Ant-based based Orthogonal Multiplicative and Transformational algorithm. Then, the proposed OSA-SIH technique gives the detailed description which is presented in the forthcoming sections.

Tong Yi and Minyong Shi (2015) intended the privacy protection method utilizes the association rules among sensitive attributes and data publication for many sensitive attributes. Privacy protection method provides better protection for privacy and also preserves significant relationships for improving the system performance. Though, the hiding the available sensitive attributes from the database is the major issue. In order to overcome such limitations, the OSA-SIH technique is proposed to hide the sensitive items for the privacy protection using user operational conditions-based sensitive items when compared to the privacy protection technique.

4.2.1 User Operational Conditions-Based Sensitive Items

The proposed OSA-SIH technique is utilized based on the performance of designing the user operational conditions-based sensitive items with better efficient. User operational condition approach is implemented with the objective of hiding the sensitive item from global frequent items for distributed dataset is being shared during the privacy preservation. Then the block diagram of user operational conditions-based sensitive items using proposed OSA-SIH technique is represented in the figure 4.1.
Tamir Tassa (2014) planned a secure mining approach of association rules for improving the present leading protocol on the basis of privacy preservation and efficiency using horizontally distributed databases. The performance of mining approach is maximized while protecting the data items from other larger data set. However, the time taken to hide the globally sensitive item sets is significantly improved. Therefore, the proposed OSA-SIH technique is introduced to hide the sensitive item with the minimal time when compared to the mining approach.

The block diagram of user operational conditions-based sensitive items using proposed OSA-SIH technique involves two values, where the support and confidence values are analyzed and hence minimizing the time for sensitive item hiding. Let us consider the dataset ‘D’, where ‘I = I₁, I₂...Iₙ‘indicates the items consists of ‘n’ number of transaction includes the set of items in such a way that ‘T ∈ I’. After that, the association rule is written as given below.
\[ P \rightarrow Q, \text{ where } P \in I & Q \in I \] (4.1)

From the equation (4.1), where ‘P’ and ‘Q’ represents the antecedent and consequent of rule respectively. Followed by this, the relative strength of an item is mainly depends on either strong or weak nature that is calculated with two factors such as support and confidence of the item. At first, the support value analyze the sensitive item hiding is mathematically represented as follows.

\[ S(P \rightarrow Q) = S(P \cup Q) = \left( \frac{(P \cup Q)}{n} \right) \] (4.2)

From the equation (4.2), the support ‘S’ indicates the proportion of transactions that includes both ‘P’ and ‘Q’ with ‘n’ number of transactions involved during the sensitive item hiding. After that, the confidence value analyze the sensitive item hiding is mathematically expressed as given below.

\[ C(P \rightarrow Q) = \left( \frac{(P \cup Q)}{P} \right) = \left( \frac{S(P \cup Q)}{S(P)} \right) \] (4.3)

From the equation (4.3), the confidence ‘C’ refers the percentage for transaction which includes both ‘P’ and ‘Q’ using user operational condition. When the support and confidence is higher than the user, selected Support Threshold Value (STV) and Confidence Threshold Value (CTV), then the rule is being optimized. Hence the ant-based relative item set distribution method cannot retrieved every item, but only the small element that satisfies the ‘STV’ and ‘CTV’ are efficiently retrieved which resulting in reducing the optimal hiding time.
Chun-Wei Lin et al. (2014) described Hiding-Missing-Artificial Utility (HMAU) algorithm utilize lattice-based approach for hiding the sensitive information using the item set deletion. A transaction with a enhanced ratio of sensitive to non sensitive chooses entire deletion and reduced the side effects for the scalable transaction deletion with lesser time interval. But the HMAU algorithm removes the estimation of large item sets and unable to handle the highest frequency in the sensitive rules related to current sensitive transaction. To address this problem, the proposed OSA-SIH technique is employed to obtain the hiding sensitive data item form larger item set using a social ant-based relative item set distribution when compared to HMAU algorithm.

4.2.2 Social Ant-Based Relative Item Set Distribution

After performing the user operational conditions-based sensitive items, the social ant-based relative item set distribution is designed using the proposed OSA-SIH technique. For attaining the efficient hiding of the sensitive item, the social ant based relative item set distribution is described during the equivalent original dataset for larger item sets. Hiding of sensitive item takes place multiplicative and transformational data perturbation method that is mainly depends upon socially cohesive relational rate among both the sensitive and non sensitive item sets of original dataset to provide the Modified Dataset ‘\( MD \)’ efficiently. Figure 4.2 illustrates the block diagram of Ant-based Orthogonal Multiplicative and Transformational Data Perturbation using a social ant-based relative item set distribution is given below.
Figure 4.2 Block diagram of Ant-based Orthogonal Multiplicative and Transformational Data Perturbation

As shown in the figure 4.2, the block diagram of Ant-based Orthogonal Multiplicative and Transformational Data Perturbation using proposed OSA-SIH technique. The fundamental idea of an ant principle is that the random wandering nature and efficient detection of food come back to colony while laying the pheromone trails downwards. Alternatively, if other ants discover those paths, the ants again cannot pass through random manner, but it blindly follows the trail generated with previous ants. In a same way, while the items in the transaction take place repeatedly, then it is represented as sensitive item. As
a result, changing the item in random manner with the aid of probability functions and therefore the frequent sensitive items are hidden in an optimized way.

Let us assume \(\alpha (x_a)\) indicates the pheromone intensity of ‘\(D\)’ ant (i.e. Dataset) that locates ‘\(x_i\)’ to be initialized as constant. Then the probability of the dataset ‘\(D\)’ which hides the sensitive item from ‘\(x_a\)’ to ‘\(x_b\)’ is mathematically represented as follows.

\[
p_{ab} = \frac{\alpha (x_b)}{\alpha (x_c)}, \text{where } x_b, x_c \in D
\]  

(4.4)

From the equation (4.4), the \(P_{ab}\) defines the position of the dataset that generates better results due to the optimal hiding of sensitive item using social ant based relative item set. If optimal sensitive item is achieved, the proposed OSA-SIH technique is accomplished with the aim of hiding sensitive item through multiplicative and transformational data perturbation method. Hence the multiplicative and transformational data perturbation method supports both sensitive and non sensitive item sets of the original dataset through which the modified dataset is being provided. In addition, the orthogonal multiplicative and transformational data perturbation is employed to enhance the privacy preservation accuracy using the data perturbation process.

Zahid Pervaiz et al. (2013) designed accuracy-constrained privacy-preserving access control technique for the relational data which includes access control and privacy protection mechanisms. The access control scheme enables the authorized query predicates on the sensitive data. Then the privacy preserving system is efficiently detects the data to achieve privacy requirements and accuracy constraints through the access control method.
However, sometimes the accuracy and access control using the privacy-preserving access control is not sufficient. Therefore, the OSA-SIH technique is established to ensure the better accuracy when compared to accuracy-constrained privacy-preserving access control technique.

Let us consider two datasets ‘L’ and ‘M’ of size ‘i * n matrix’ and ‘j * n matrix’ respectively with the orthogonal matrix is denoted as ‘O’. At present, the mathematical expression for orthogonal multiplicative and transformational data perturbation using two datasets ‘L’ and ‘M’ is formulated as.

\[
A = LO; \quad B = MO \tag{4.5}
\]

\[
AA^T = LL^T; \quad BB^T = MM^T \tag{4.6}
\]

\[
AB^T = L O O^T B^T = LM^T \tag{4.7}
\]

From the equations (4.5), (4.6) and (4.7), with the use of orthogonal matrix based on socially cohesive relational rate with the sensitive and non sensitive item sets, every pair distances and column vectors ‘A and B’ that are more secured during the perturbed data. Simultaneously, both the sensitive and non sensitive items as well as the multiplicative and transformational data perturbation method are kept secret while providing the data is viewed by the third user which resulting in enhance the privacy preservation accuracy. The figure 4.3 illustrates the algorithmic process of ant-based orthogonal multiplicative and transformational algorithm using social ant based relative item set is given below.
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 (4.2)
Step 5: Evaluate confidence ‘\(C\)’ for sensitive item hiding using (4.3)
Step 6: If ‘\(S < STV\)’ and ‘\(C < CTV\)’
Step 7: Evaluate optimal hiding of sensitive item using (4.4)
Step 8: Evaluate orthogonal multiplicative and transformational process using (4.5), (4.6) and (4.7)
Step 9: else
Step 10: go to step 2
Step 11: End if
Step 12: End for
Step 13: End for
Step 14: End for
Step 15: End

Figure 4.3 Ant-based Orthogonal Multiplicative and Transformational algorithm

As shown in the figure 4.3, the Ant-based based Orthogonal Multiplicative and Transformational (AOMT) algorithm involves four steps. Initially, the AOMT algorithm assesses the support value for the sensitive item hiding. Then the AOMT algorithm analyzes the confidence value for hiding the sensitive item. Next, the comparison is made with the Support Threshold Value ‘\(STV\)’ and Confidence Threshold Value ‘\(CTV\)’ through
the estimated confidence ‘C’ and support ‘S’ value. After that, the optimal hiding of sensitive item and orthogonal multiplicative and transformational method is being obtained in an optimized manner. Once the values of support ‘S’ and confidence ‘C’ is smaller than the ‘STV’ and ‘CTV’, the item hiding is achieved or else similar operations is executed with other transactions which in turn ensures the privacy preservation accuracy.

Yaping Li et al. (2012) planned Perturbation-based PPDM to Multilevel Trust (MLT-PPDM) method enables flexibility and provide data perturbed for various trust levels. Data owner’s provides enhanced flexibility and diversity gain in joint estimation with minimum error evaluation. Several trust levels are accurately connected with the reconstructed original data. But, the data owner is unable to predict all the possible trust level for the data to be mined. In order to overcome this issue, the proposed OSA-SIH framework is attained to improve the trust level using correlation-based privacy preservation when compared to MLT-PPDM method.

4.2.3 Correlation-based Privacy Preserving

After performing the social ant-based relative item set distribution, the privacy preservation through the correlation-based approach is designed on the basis of proposed OSA-SIH technique. A side effect of hiding the item sets investigates the different user requested item set distribution on modified dataset that enhances the user trust level by using PPDM. Also, the relationship between the hidden sensitive items and modified entries (i.e. dataset) being evaluated which aims to reduces the rate of side effects on the modified dataset using OSA-SIH technique.
Chun-Wei Lin et al. (2014) described a compact prelarge Genetic Algorithm-based Deleted Transactions (cpGA2DT) algorithm for hiding the sensitive itemsets. The flexible fitness function contains three modifiable weights to detect the suitable transactions to be deleted for hiding the sensitive itemsets with the minimum side effects. But, the cpGA2DT algorithm not only reduces the execution time and also improves the side effects during PPDM. Hence, the proposed OSA-SIH technique is designed to reduce the side effects when compared to cpGA2DT algorithm.

Proposed OSA-SIH technique utilizes correlation-based method for reducing the side effects involved during the privacy preservation technique. Let us assume three transactions ‘p’, ‘q’ and ‘r’. If correlating with another sensitive item, if ‘r → q’, then deleting the item ‘q’ is optimized than deleting ‘p’ as deleting the previous ‘r’ affects both items. Followed by this, if correlating with non sensitive item, if ‘q → p’, then inserting ‘p’ into the transactions which cannot involves ‘q’ is optimized than deleting ‘p’ or ‘q’ from the transactions. As a result, the rate of side effects on the modified dataset is decreased which resulting in hiding the sensitive item in a significant manner.

4.3 EXPERIMENTAL SETTINGS

To evaluate the effectiveness of the proposed method called Optimized Social Ant Based Sensitive Item Hiding (OSA-SIH) technique which is implemented with the help of JAVA language. The OSA-SIH technique utilizes the Adult data set from the University of California Irvine data repository which involves the information on individuals namely age, level of education and current employment type. The dataset consists of 49K records and also binomial label which represents the salary of <50K or >50K US dollars. Then the
simulation is conducted with the data is about 32K records for training dataset and 16K records for test dataset.

There are fourteen attributes including the seven polynomials, one binomial and six continuous attributes using the proposed OSA-SIH technique to preserve the privacy of specific attributes such as salary, relationship and marital status. The employment class attribute indicates the employer type (i.e. self employed or federal) and occupation refers to the employment type (i.e. farming or managerial). The education attribute contains high school graduate or doctorate. Then the relationship attribute comprises of the information related to unmarried or married. The concluding nominal attributes are country of residence, gender and race. In addition, the continuous attributes are age, hours worked per week, education number, capital gain and loss and also survey weight attribute allocates an individual based on information like area of residence and kind of employment.

The performance of the proposed OSA-SIH technique is analyzed with the following metrics including the privacy preservation accuracy, time for optimal hiding and rate of side effects on the modified dataset to be carried out for the optimal hiding of sensitive item which are discussed as follows.

4.4 RESULTS AND DISCUSSION

In this section, the result analysis of the proposed Optimized Social Ant Based Sensitive Item Hiding (OSA-SIH) technique is compared with the existing methods including the Multilevel Trust in Privacy Preserving Data Mining (MLT-PPDM) method developed by Yaping Li et al. (2012) and Hiding-Missing-Artificial Utility (HMAU) algorithm developed by Chun-Wei Lin et al. (2014). The performance of OSA-SIH
framework is carried out with different metrics such as, privacy preservation accuracy, time for optimal hiding and rate of side effects on the modified dataset which are discussed as follows.

4.4.1 Measure of the Privacy Preservation Accuracy

The privacy preservation accuracy is defined as the ratio of privacy preserved perturbed copies to the total number of perturbed copies taken for the experimental purpose, Privacy preservation accuracy is measured in terms of percentage (%). The mathematical formulation of privacy preservation accuracy is mathematically written as given as below.

\[ A = \left( \frac{\text{privacy preserved perturbed copies}}{n} \right) \times 100 \]  

(4.8)

From the equation (4.8), where ‘A’ is represented as privacy preservation accuracy which is evaluated based on the number of perturbed copies ‘n’. When the privacy preservation accuracy gets higher, thus the method is significantly more efficient.

**Table 4.1 Tabulation of the Privacy Preservation Accuracy**

<table>
<thead>
<tr>
<th>Age (Number of perturbed copies)</th>
<th>Proposed OSA-SIH</th>
<th>Existing MLT-PPDM</th>
<th>Existing HMAU</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>70.35</td>
<td>64.48</td>
<td>52.31</td>
</tr>
<tr>
<td>20</td>
<td>74.41</td>
<td>70.48</td>
<td>58.4</td>
</tr>
<tr>
<td>30</td>
<td>75.39</td>
<td>72.54</td>
<td>60.54</td>
</tr>
<tr>
<td>40</td>
<td>73.28</td>
<td>71.26</td>
<td>59.91</td>
</tr>
<tr>
<td>50</td>
<td>76.97</td>
<td>73.96</td>
<td>63.88</td>
</tr>
<tr>
<td>60</td>
<td>80.32</td>
<td>77.43</td>
<td>67.33</td>
</tr>
</tbody>
</table>
The table 4.1 illustrates the privacy preservation accuracy for the proposed OSA-SIH technique and existing system namely MLT-PPDM method by Yaping Li et al. (2012) and HMAU algorithm by Chun-Wei Lin et al. (2014). Number of perturbed copies is taken as input which is varied from the range of 10 to 100 for conducting test. For increasing the number of age, privacy preservation accuracy is also increased using all the techniques are shown in the figure 4.4. However, the proposed OSA-SIH framework provides maximum privacy preservation accuracy when compared to existing method.

<table>
<thead>
<tr>
<th></th>
<th>Proposed OSA-SIH</th>
<th>Existing MLT-PPDM</th>
<th>Existing HMAU</th>
</tr>
</thead>
<tbody>
<tr>
<td>70</td>
<td>84.36</td>
<td>80.44</td>
<td>70.36</td>
</tr>
<tr>
<td>80</td>
<td>87.14</td>
<td>83.17</td>
<td>72.3</td>
</tr>
<tr>
<td>90</td>
<td>89.53</td>
<td>86.11</td>
<td>74.22</td>
</tr>
<tr>
<td>100</td>
<td>91.26</td>
<td>87.84</td>
<td>75.83</td>
</tr>
</tbody>
</table>

**Figure 4.4 Measure of the Privacy Preservation Accuracy**
Figure 4.4 demonstrates the privacy preservation accuracy for the proposed OSA-SIH technique and compared with the existing methods including MLT-PPDM method by Yaping Li et al. (2012) and HMAU algorithm by Chun-Wei Lin et al. (2014). As shown in the figure 4.4, the proposed OSA-SIH framework significantly improves the privacy preservation accuracy when compared to the existing methods. This is improved based on the application of orthogonal multiplicative and transformational data perturbation, the data perturbation is measured efficiently in terms of orthogonal matrix with the training samples. Hence, the privacy preservation accuracy of the proposed OSA-SIH technique is improved by 5% when compared to the existing MLT-PPDM method by Yaping Li et al. (2012) and 23% when compared to the existing HMAU algorithm by Chun-Wei Lin et al. (2014) respectively.

4.4.2 Measure of Time for the Optimal Hiding

The time for optimal hiding is measured based on the total number of transactions and the time taken to require single transaction using the proposed OSA-SIH framework. The time for optimal hiding is measured in terms of milliseconds (ms) and is formulated as given below.

\[
\text{Time} = n \times \text{Time (item hiding for single transaction)} \quad \text{(4.9)}
\]

From the equation (4.9), the time for optimal hiding ‘Time’ is analyzed with respect to the total number of transactions ‘n’. Lower the time for optimal hiding, thus the technique is said to be more efficient.
Table 4.2 Tabulation of Time for the Optimal Hiding

<table>
<thead>
<tr>
<th>Total number of transactions (n)</th>
<th>Time for Optimal Hiding (ms)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Proposed OSA-SIH</td>
<td>Existing MLT-PPDM</td>
<td>Existing HMAU</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>4.22</td>
<td>6.62</td>
<td>10.55</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>5.85</td>
<td>7.56</td>
<td>12.3</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>7.55</td>
<td>9.32</td>
<td>13.9</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>9.11</td>
<td>10.52</td>
<td>15.35</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>8.62</td>
<td>10.79</td>
<td>15.7</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>11.46</td>
<td>13.32</td>
<td>17.75</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>13.14</td>
<td>15.31</td>
<td>19.61</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>15.33</td>
<td>17.36</td>
<td>21.36</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>16.34</td>
<td>18.53</td>
<td>22.53</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>17.56</td>
<td>19.3</td>
<td>23.8</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.2 demonstrates the time for the optimal hiding based on the total number of transaction using the proposed OSA-SIH technique and existing system namely MLT-PPDM method by Yaping Li et al. (2012) and HMAU algorithm by Chun-Wei Lin et al. (2014). Total number of transaction is varied from the range of 5 to 50 for the experimental evaluation. For all techniques, table 4.2 shows that the time for optimal hiding is increased with respect to increasing the total number of transactions. However, the proposed OSA-SIH technique significantly measured by means of reducing the optimal hiding time when compared to the existing method.
Figure 4.5 Measure of Time for Optimal Hiding

Figure 4.5 describes the time for optimal hiding using the proposed OSA-SIH technique and comparison is made with the existing system including MLT-PPDM method by Yaping Li et al. (2012) and HMAU algorithm by Chun-Wei Lin et al. (2014). As shown in the figure 4.5, it is clear that the proposed OSA-SIH technique comparatively reduces the time for optimal hiding when compared to the existing methods. During the user operational conditions-based sensitive items, the support and confidence values are performed for sensitive item hiding which generates the better results with respect to total number of transactions minimize the time for optimal hiding. Therefore, the time for optimal hiding is reduced using the proposed OSA-SIH technique by 17% when compared to the existing MLT-PPDM method by Yaping Li et al. (2012) and 40% when compared to the existing HMAU algorithm by Chun-Wei Lin et al. (2014) respectively.
4.4.3 Measure of Rate of Side Effects

The rate of side effects is defined as the difference between the actual size of transaction and the modified dataset being generated through the privacy preservation. The mathematical formulation for the rate of side effects is given as below.

\[
RoSE = (Size - MD)
\]  

(4.10)

From the equation (4.10), the Rate of Side Effects ‘\(RoSE\)’ is measured on the basis of size of transaction ‘\(Size\)’ and Modified Dataset ‘\(MD\)’ respectively. Rate of side effects is measured in terms of kilobytes (KB). If the rate of side effects gets reduced, thus the method is said to be more effective.

<table>
<thead>
<tr>
<th>Size of Transaction (KB)</th>
<th>Rate of Side Effects (KB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Proposed OSA-SIH</td>
</tr>
<tr>
<td>100</td>
<td>45</td>
</tr>
<tr>
<td>200</td>
<td>55</td>
</tr>
<tr>
<td>300</td>
<td>70</td>
</tr>
<tr>
<td>400</td>
<td>83</td>
</tr>
<tr>
<td>500</td>
<td>100</td>
</tr>
<tr>
<td>600</td>
<td>113</td>
</tr>
<tr>
<td>700</td>
<td>121</td>
</tr>
<tr>
<td>800</td>
<td>127</td>
</tr>
<tr>
<td>900</td>
<td>130</td>
</tr>
<tr>
<td>1000</td>
<td>133</td>
</tr>
</tbody>
</table>

Table 4.3 Tabulation of the Rate of Side Effects
Table 4.3 shows the rate of side effects with respect to the size of transaction using the proposed OSA-SIH technique and existing system namely MLT-PPDM method by Yaping Li et al. (2012) and HMAU algorithm by Chun-Wei Lin et al. (2014). Size of transaction is varied from the range of 100 to 1000 for the experimental purpose. As shown in the table 4.3, by increasing the size of transaction, the rate of side effects is also increased using all the techniques. However, the proposed OSA-SIH technique provides better performance in terms of improving the rate of side effects when compared to existing method.

Figure 4.6 illustrates the rate of side effects for the proposed OSA-SIH technique which is compared with the existing system including MLT-PPDM method by Yaping Li et al. (2012) and HMAU algorithm by Chun-Wei Lin et al. (2014). From the figure 4.6, the proposed OSA-SIH technique effectively reduces the rate of side effects when compared to the existing methods.
The rate at which the side effects of hiding the item sets on the modified dataset are demonstrated for several users requested the item set distribution during the privacy preserving distributed data mining. This in turn also helps to improve the user trust level in an efficient manner. Hence, the rate of side effects using the proposed OSA-SIH technique is reduced by 27% when compared to the existing MLT-PPDM method by Yaping Li et al. (2012) and 36% when compared to the existing HMAU algorithm by Chun-Wei Lin et al. (2014) respectively.

4.5 SUMMARY

An efficient framework called Optimized Social Ant Based Sensitive Item Hiding (OSA-SIH) introduced a better privacy preservation for the distributed data mining with the optimal side effects during original dataset. With the aim of ensuring the improved privacy preservation of data items of corresponding the user’s privileges for the distributed data and reducing the optimal hiding time for various user requested item set distribution, the OSA-SIH technique is designed. Initially, the user operational conditions-based sensitive item discovers the support and confidence value from global frequent item sets for the distributed datasets. Next, the social ant-based relative item set distribution provides a privacy preservation accuracy for the large item sets through multiplicative and transformational data perturbation technique. Finally, Ant-based Orthogonal Multiplicative and Transformational algorithm is designed with the aid of probability function for improving the privacy preservation accuracy. Then the performance of OSA-SIH technique is tested with the metrics such as time for optimal time hiding, rate of side effects and privacy preservation accuracy.
The simulation results illustrate the proposed OSA-SIH technique provides a better performance with the improvements of privacy preservation accuracy by 14% and reduction of optimal time hiding by 29% when compared to the state-of-the-art works. However, the privacy preservation is not sufficient. In order to increase the efficiency of privacy preserving association rule mining further with the constraint minimization, the Swarm optimization and Iterative Privacy Rule Preservation (SIPRP) method is designed. SIPRP technique utilizes the association rules to offers privacy preserving distribution database using support and confidence threshold. Then, the sensitive rules are efficiently hiding and preserving the better confidential privacy rules with the help of particle swarm optimization. Also, the SIPRP method attains the sensitive item sets for providing the specific sensitive rules to be hided with the minimum effect which is discussed in detail in the forthcoming chapter.