Privacy Preservation in Fuzzy Association Rules Using Rough Computing and DSR

A. Geetha Mary, D. P. Acharjya, N. Ch. S. N. Iyengar

School of Computing Science and Engineering, VIT University, Vellore-632014, Tamilnadu, India
Emails: geethamary@vit.ac.in   dpacharjya@gmail.com   nchsniyengar48@gmail.com

Abstract: *In the present age of Internet, data is accumulated at a dramatic pace. The accumulated huge data has no relevance, unless it provides certain useful information pertaining to the interest of the organization. But the real challenge lies in hiding sensitive information in order to provide privacy. Therefore, attribute reduction becomes an important aspect for handling such huge database by eliminating superfluous or redundant data to enable a sensitive rule hiding in an efficient manner before it is disclosed to the public. In this paper we propose a privacy preserving model to hide sensitive fuzzy association rules. In our model we use two processes, named a pre-process and post-process to mine fuzzified association rules and to hide sensitive rules. Experimental results demonstrate the viability of the proposed research.*

Keywords: *Almost indiscernibility, fuzzy proximity relation, fuzzy approximation space, ordering rules.*

1. Introduction

In the modern era of computing, data is collected at a dramatic pace. However, the real challenge lies in converting huge data into knowledge. Therefore, it is essential for a new generation of computational theories and tools to assist humanity in extracting knowledge from the rapidly growing voluminous data. Thus knowledge discovery in databases is the field that has emerged as an important and active area of research, since the availability of data is high, whereas the availability of

knowledge is low. Instead of generating knowledge from the experts, data mining, a process of extracting knowledge from the data is proved to be more efficient. Association rule mining is one of the most widely used techniques in data mining and knowledge discovery. However, due to the presence of sensitive information in those data, much attention has been paid to information privacy. A multitude of research works has been carried out in this direction by many researchers, but all of them deal only with qualitative data. Also, the efficiency of the system degrades with the increment of the redundancy of the attributes. All this leads to the concept of Privacy Preserving Data Mining (PPDM).

PPDM is the current research interest of scientists ever since the awareness of privacy has arisen among people. From health care to banking, customer details and all kinds of personal data are highly vulnerable by public disclosure. But this data has to be analyzed to retrieve knowledge from it. PPDM paves the way to balance between research and disclosure. Privacy can be applied at two levels; either when the data enters the system, called input privacy or at pattern representation, called output privacy. Input privacy can be achieved by imposing techniques like k-anonymity [1, 2], l-diversity [3, 4] and t-closeness [5], or by distorting the data before given as an input to the data mining system using data perturbation methods [6, 7]. Output privacy techniques are generally selected according to the data mining techniques. It is because the patterns from different data mining techniques are different and all of them have to be handled appropriately [8, 9]. Data available in many real life situations, such as economics, medical science, social science, etc. are not always crisp and contains uncertainties. The different theories that deal with uncertainties are fuzzy sets [10], rough sets [11], intuitionistic fuzzy sets [12], a rough set on fuzzy approximation spaces [13], a rough set on intuitionistic fuzzy approximation space [14-16], etc. Keeping in mind both the uncertainties and privacy, in this paper, we use two processes, such as a pre-process and post-process to mine fuzzified association rules and to hide sensitive rules.

The data, from which the rules are generated, are usually presented in the form of an information system which consists of attributes and objects. It is observed that the attribute values are either symbolic or quantitative. But many of the attributes may not be relevant while studying the information system. Therefore, the attribute reduction becomes an important aspect for handling large databases efficiently, by eliminating superfluous or redundant data. Though privacy, preserving fuzzy association rules hides the quantitative data, as discussed by M. G u p t a and R. C. J o s h i [17], it fails to eliminate the superfluous attribute that has no resemblance with the information system. Besides, it is observed that in the information system, the quantitative attribute values are almost similar, hence it is very difficult to eliminate these attributes from the information system. Therefore, efforts have been made to convert the quantitative attribute values to symbolic values by using a fuzzy proximity relation. In addition, we have introduced ordering rules to get the ordered symbolic information system. Further on, we use rough set techniques to eliminate the superfluous attributes. Finally, the fuzzy association rules can be obtained from the ordered information system and sensitive rules can be hidden using DSR techniques.

In the pre-process we use a rough set on fuzzy approximations with ordering rules and rough set data reduction to reduce the dataset dimensionality and fuzzification to mine fuzzified association rules. In the post-process, we Decrease the Support of the Right hand side (DSR) to hide sensitive fuzzy association rules. The rest of the paper is organized as follows: Section 2 presents the foundations of rough computing. Section 3 provides the basic idea of fuzzy association rule mining. The proposed model is presented in Section 4, followed by an empirical study in Section 5. Experimental results are discussed in Section 6, which is further followed by a conclusion in Section 7.

2. Foundations of rough computing

Real time data contains uncertainties. Therefore, paying no attention to it often leads to a failure of the system. Traditional statistical tools overlook these uncertainties and, therefore, lack in accuracy. Computational intelligence techniques, such as artificial neural networks [18], evolutionary algorithms [19], fuzzy sets [10], rough sets [11], etc. consider these uncertainties and deal specifically with them. The fuzzy set of Z a d e h [10] is the first successful method that captures impreciseness in information. On the other hand, the rough set of P a w l a k [11] captures indiscernibility among objects to model imperfect knowledge. The basic definition of rough sets is based on the approximation of a set by a pair of sets known as lower and upper approximation.

Let U be the universe of a finite non empty set of objects. Let $R \subseteq U \times U$ be an equivalence relation on U. Let X be a subset of U. The equivalence relation R partitions the set U into disjoint classes and we denote it as U/R. Therefore, the target set X can be described by lower and upper approximation. We denote $\underline{R}X$ and $\overline{R}X$ as R-lower and R-upper approximations of X respectively, given as:

(1) $$\underline{R}X = \cup \{Y \in U/R : Y \subseteq X\},$$

(2) $$\overline{R}X = \cup \{Y \in U/R : Y \cap X \neq \phi\}.$$

The boundary region of the set X, $BN_R(X)$, are the objects in X that can be distinguished neither as a member nor as a non-member of X, employing relation R. We denote it by $BN_R(X) = \overline{R}X - \underline{R}X$. A set X is said to be definable if $\underline{R}X = \overline{R}X$. Similarly, it is said to be rough if $\underline{R}X \neq \overline{R}X$ or equivalently $BN_R(X) \neq \phi$.

2.1. Rough set on fuzzy approximation space

Indiscernibility relation, an equivalence relation, is the basic viewpoint of rough sets. But in many real life situations, the equivalence relation does not depict the need accurately. Therefore, it is necessary to make the relations less stringent by excluding one or more requirements of the equivalence relation. A fuzzy proximity relation is more generalized than an equivalence relation defined over the universe U and is more suitable due to its non-transitivity property. Thus, fuzzy proximity

54

relation portrays the present-day examples in a more efficient and appropriate way [13, 20, 21].

Let U be a universe. We define a fuzzy relation on U as a fuzzy subset of $(U \times U)$. A fuzzy relation R on U is a fuzzy proximity relation if $\mu_R(x, x) = 1$ for all $x \in U$ and $\mu_R(x, y) = \mu_R(y, x)$ for $x, y \in U$. Let R be a fuzzy proximity relation on U. Then for a given $\alpha \in [0, 1]$ we say that two elements x and y are α-similar with respect to R if $\mu_R(x, y) \geq \alpha$ and we write $x R_\alpha y$ or $(x, y) \in R_\alpha$. Two elements x and y in U are said to be α-identical denoted by $x R(\alpha) y$, if either x is α-similar to y or x is transitively α-similar to y, that is, there exists a sequence $u_1, u_2, u_3, ..., u_n$ in U, such that $x R_\alpha u_1$, $u_1 R_\alpha u_2$, $u_2 R_\alpha u_3$, $..., u_n R_\alpha y$. If x and y are α-identical with respect to the fuzzy proximity relation R, then we write $x R(\alpha) y$, where the relation $R(\alpha)$ for each fixed $\alpha \in [0, 1]$ is an equivalence relation on U. The pair (U, R) is called a fuzzy approximation space. The rough set of X, in the generated approximation space $(U, R(\alpha))$ is denoted by $(\underline{X}_\alpha, \overline{X}_\alpha)$ and is defined with respect to R_α^*, the family of equivalence classes of $R(\alpha)$. The α lower approximation of X, $\underline{X}_\alpha$ and α upper approximation of X, $\overline{X}_\alpha$ are defined as follows:

(3) $$\underline{X}_\alpha = \cup \{ Y : Y \in R_\alpha^* \text{ and } Y \subseteq X \},$$

(4) $$\overline{X}_\alpha = \cup \{ Y : Y \in R_\alpha^* \text{ and } Y \cap X \neq \phi \}.$$

X is said to be α-discernible if and only if $\underline{X}_\alpha = \overline{X}_\alpha$ and X is said to be α-rough if $\underline{X}_\alpha \neq \overline{X}_\alpha$. It is also noticed that $R(\alpha)$ is not an exact indiscernibility relation defined by Pawlak, rather it can be considered as an almost indiscernibility relation on U. The almost indiscernibility relation $R(\alpha)$ reduces to Pawlak's exact indiscernibility relation when $\alpha = 1$ and thus it generalizes Pawlak's indiscernibility relation. The family of all equivalence classes of $R(\alpha)$, i.e., the partition generated for $\alpha \in [0, 1]$ is denoted by $U / R(\alpha)$. If $(x, y) \in R(\alpha)$, then we say that x and y are α-indiscernible. These are the basic building blocks of a rough set on fuzzy approximation space.

2.2. Ordered information system

Classifying the objects in an information system is the basic objective of inductive learning and data mining. It is certainly true for rough computing based approaches. However, in many real life situations, we may not face with a simple classification. Ordering of objects is one such problem. An information system is defined as a quadruple $I = (U, A, V_a, f_a)$ where U is a finite nonempty set of objects called the universe, A is a finite nonempty set of attributes, V_a is a nonempty set of values for $a \in A$, $f_a : U \to V_a$ is an information function.

Consider the sample information system given in Table 1. Here we have $U = \{x_1, x_2, x_3, x_4, x_5\}$; $A = \{$CGPA, Programme (p), Year of joining (yoj), Year of passing (yop)$\}$ and $V_{\text{CGPA}} = \{6.9, 7.8, 8.2, 8.6, 9.3\}$.

Table 1. Sample information system

Object	CGPA	Programme	Year of joining	Year of passing
x_1	8.6	B Tech.	2004	2008
x_2	7.8	M Tech.	2006	2008
x_3	8.2	M Tech.	2005	2007
x_4	9.3	B Tech.	2005	2009
x_5	6.9	B Tech.	2003	2007

An ordered information system is defined as $\text{OIS} = \{I, \{\prec_a : a \in A\}\}$, where I is a standard information system and $\prec_a$ is an order relation on attribute $a \in A$. An ordering of values of a particular attribute induces an ordering of objects:

(5) $$x_i \prec_{\{a\}} x_j \iff f_a(x_i) \prec_a f_a(x_j),$$

where $\prec_{\{a\}}$ denotes an order relation on U induced by the attribute a. An object x_i is ranked ahead of object x_j, if and only if the value of x_i on the attribute a is ranked ahead of the value of x_j on the attribute a. For example, the information system given in Table 1 becomes an order information system after applying the following ordering relations:

$$\prec_{\text{CGPA}}: \quad 9.3 \prec 8.6 \prec 8.2 \prec 7.8 \prec 6.9,$$
$$\prec_p: \quad \text{M Tech.} \prec \text{B Tech.},$$
$$\prec_{\text{yoj}}: \quad 2006 \prec 2005 \prec 2004 \prec 2003,$$
$$\prec_{\text{yop}}: \quad 2009 \prec 2008 \prec 2007.$$

For a subset of attributes $B \subseteq A$, an object x_i is ranked ahead of object x_j if and only if x_i is ranked ahead of x_j according to all attributes in B, i.e.,

$$x_i \prec_B x_j \iff f_a(x_i) \prec_a f_a(x_j) \quad \forall a \in B,$$
$$\iff \bigwedge_{a \in B} f_a(x_i) \prec_a f_a(x_j) \iff \bigcap_{a \in B} \{\prec_a\}.$$

The above definition is a straightforward generalization of the standard definition of equivalence relations in rough set theory, where the equality relation is used [11, 22].

3. Fuzzy association rule mining

The universe can be considered as a large collection of objects. There is some information associated with each object. To find knowledge about the universe we need to elicit some information about these objects. Therefore, it is essential to know the relationship between the attributes and these values to gain knowledge.

The fuzzy association rule is a type of relationship among the attribute values of the objects of the universe.

Let $S = (U, A)$ be an information system. Let C and D be sets of attribute values of the objects of the universe U. Confidentiality of $C \Rightarrow D$ is called a fuzzy association rule if it satisfies certain criteria. We define the criteria in terms of two factors called support of $C \Rightarrow D$ and confidentiality, as follows:

$$(6) \qquad \text{Sup}\,(C \Rightarrow D) = \frac{\Sigma_{(x \in U)}(C \cap D)(x)}{\text{Total number of objects in the universe}},$$

$$(7) \qquad \text{Conf}\,(C \Rightarrow D) = \frac{\text{Sup}\,(C \Rightarrow D)}{\text{Sup}\,(C)}.$$

In this paper we use the notation $\text{Sup}\,(C)$ for support of C and notation $\text{Conf}\,(C \Rightarrow D)$ for confidentiality of $C \Rightarrow D$. If the support of C is greater than or equal to the minimum support, and confidentiality of $C \Rightarrow D$ is greater than or equal to the minimum confidentiality, then $C \Rightarrow D$ is a fuzzy association rule, otherwise it is not an association rule [23, 24]. It means that a fuzzy association rule clearly depends on support and confidentiality. Therefore, an association rule is frequent if its support and confidentiality are greater than the specified level of minimum support and minimum confidentiality. This method works well when the data is qualitative and fails when the data is quantitative. In this paper we reduce the quantitative data to qualitative by using a rough set on the fuzzy approximation spaces and ordering rules.

Fuzzy association rule mining finds an interesting association among a large set of data items. Many organizations are acquiring interest in mining the fuzzy association rules from their databases. The discovery of the interesting association relationship among a huge amount of data can help in many decision making processes. Besides, the association rule hiding is essential to hide sensitive fuzzy association rules. The association rule hiding technique is broadly classified into two categories, such as a distortion based technique and a blocking based technique. In blocking based technique, the uncertainty is introduced in the input data to hide sensitive rules. In distortion based technique, the input data is distorted, so that the support and confidentiality go below the specified threshold value. In order to achieve this, the association rule hiding technique uses either Increase Support of Left hand side (ISL) or Right hand side DSR [25].

4. Proposed model

In this section, a model for hiding sensitive rules is proposed. The complete working model, which consists of a pre-process and post-process, is shown in Fig. 1. All the data received are not of qualitative format. So the quantitative data have to be converted into discrete data. Therefore, in the pre-process, we process the quantitative data after cleaning the data by using a rough set on the fuzzy approximation space and ordering rules. Further on we use rough set reduction techniques to reduce the number of attributes that do not have influence in the

ordered information system. Based on the classification obtained in the pre-process, fuzzification is used in the post-process to mine the fuzzified association rules. Further, decreasing of the DSR based approach is used to hide sensitive fuzzy association rules from the data. The main advantage of this model is that it works for both qualitative and quantitative data.

The fundamental step of any model is the identification of the right problem. Incorporation of prior knowledge is always associated with the problem definition. However, the potential validity or usefulness of an individual data element or the pattern of a data element may change dramatically from an organization to organization, because of the acquisition of knowledge and reasoning that may be involved in vagueness and incompleteness. It is very difficult for human beings to preserve sensitive rules that are present in the high dimensional data. Therefore, the most important challenge is to hide the sensitive rules that are present in the high dimensional data. To this end, we use a rough set on the fuzzy approximation space with ordering rules and rough set reduction in the pre-process to mine suitable classification. In the pre-process, as shown in Fig. 1, we use a rough set on fuzzy approximation spaces with ordering rules for processing and classifying data after removal of noise and missing data. Based on the classification obtained in the pre-process, we use fuzzification to mine the fuzzified association rules and then hide the sensitive fuzzy association rules by using the DSR approach.
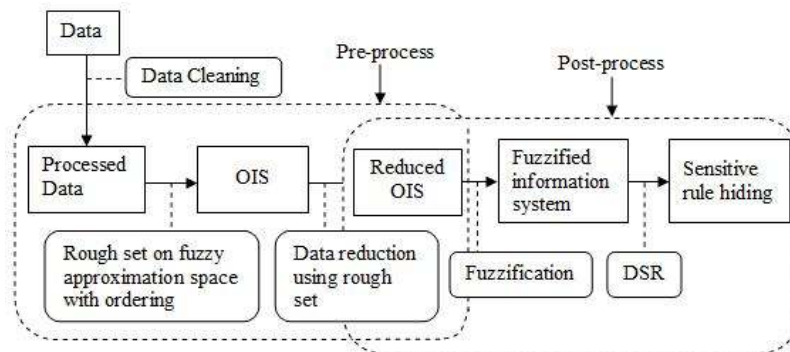
Fig. 1. The proposed sensitive rule hiding model

4.1. Pre-process architecture design

In this section we present our pre-process architecture design that consists of understanding the problem, target data, data cleaning, fuzzy proximity relation, data classification, and ordering rules, as shown in Fig. 2. The problem definition and incorporation of prior knowledge are the fundamental steps of any model. Then structuring the objectives and the associated attributes, a target dataset is created on which the data mining is to be performed. Before further analysis, a sequence of data cleaning tasks, such as removing noise, consistency check, and data completeness are done to ensure that the data is as accurate as possible. Finally for each attribute, we compute the α-equivalence classes based on the almost indiscernibility relation as discussed in Section 2. We define a fuzzy proximity

relation $R(x_i, x_j)$ in order to identify the almost indiscernibility among the objects x_i and x_j, where

(8) $$R(x_i, x_j) = 1 - \frac{\left| V_{x_i} - V_{x_j} \right|}{2(V_{x_i} + V_{x_j})}.$$
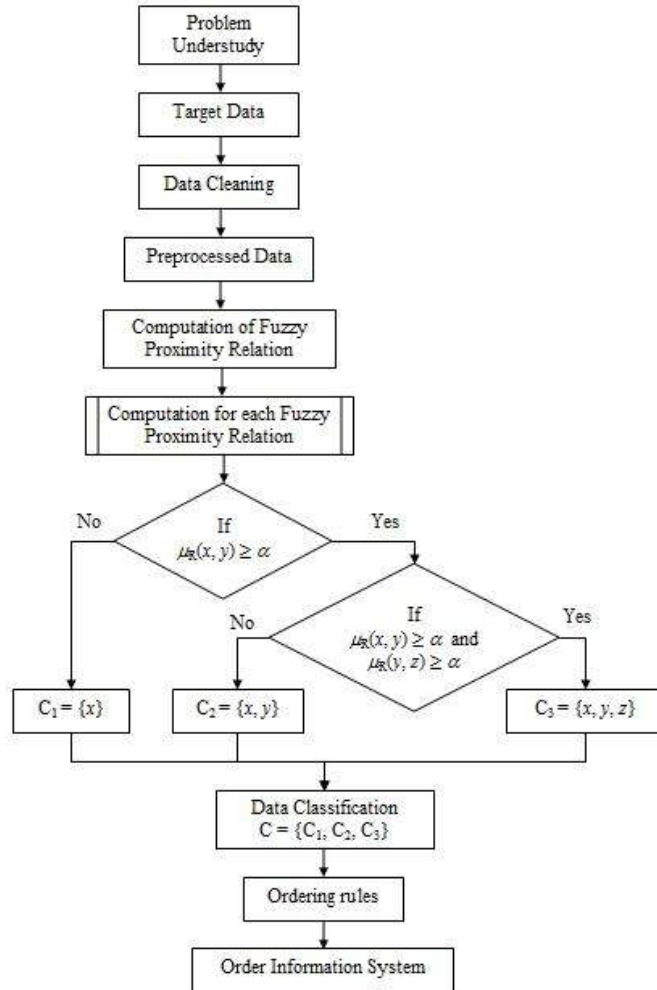
Fig. 2. Pre-process architecture design of proposed model

The membership function has been adjusted in such a manner that their values must lie within [0, 1] and the function must also be symmetric. The fuzzy proximity relation identifies the almost indiscernibility among the objects. This result induces the α-equivalence classes. We obtain qualitative or categorical classes by imposing order relation on this classification. Moreover, the attribute reduction, an important aspect of the rough set theory, is done on the ordered information system. This can minimize the set of attributes and make the object classification satisfy the full set of attributes. In practical applications it can be observed that the reduct attributes

can remove the superfluous attributes with respect to a specific classification generated by attributes $B \subseteq A$ and give the decision maker simple and easy information. If the set of attributes is dependent, using the dependency properties of the attributes, we find all possible minimal subsets of attributes which have the same number of elementary sets without loss of the classification power of the reduced information system [26, 27]. To express the above notions more clearly, we need some auxiliary notations.

Let $B \subseteq A$ and $a \in B$. We say that the attribute a is dispensable in B, if the following condition (9) holds; otherwise a is indispensable in B,

$$(9) \qquad U / B = U /(B - \{a\}).$$

Set B is independent if all its attributes are indispensable. Reduct B' of B is a subset of attributes B such that the equivalence class, induced by the reduced attribute set B', is the same as the equivalence class structure induced by the attribute set B, i.e., $U / B = U / B'$. The core of the attribute set B is the set of all indispensable attributes of B. The important property connecting the notion of a core and reducts is defined in (10), where Red(B) is the set of all reducts of B,

$$(10) \qquad \text{Core}(B) = \bigcap \text{Red}(B).$$

4.2. Post-process architecture design

The fundamental objective of data mining is to extract hidden patterns from databases. The objective of the post-process is to hide certain sensitive information so that the sensitive information cannot be discovered through a data mining technique. The frequent rule which is having sensitive information and not supposed to be disclosed is termed as a sensitive rule. It indicates that all sensitive rules are frequent rules, whereas the converse is not true.

In order to hide a fuzzy association rule, we either decrease the DSR, or we increase the support of the left hand side of the rule (ISL). In the post-process we hide critical fuzzy association rules from the reduced ordered information system by using DSR method. The data obtained from the ordered information system is then fuzzified by using triangular and trapezoidal membership functions. The triangular membership function is defined below by using three parameters p, q and r, where p is considered as the left end of the triangle, r is considered as the right end of the triangle and q is considered as the peak of the triangle,

$$(11) \qquad \mu = \text{Max}\left(\text{Min}\left(\frac{x - p}{q - p}, \frac{r - x}{r - q} \right),\ 0 \right).$$

The second, trapezoidal membership function is defined below by using two parameters s and t, where s is considered as the left end of the trapezium and t is the left peak of the trapezium,

$$(12) \qquad \mu = \text{Max}\left(\text{Min}\left(\frac{x - s}{t - s}, 1 \right),\ 0 \right).$$

60

The membership functions, defined in equations (11) and (12), are depicted in Fig. 3, where x represents the attribute values in the ordered information system. For example, if the attribute value is 7, then $\mu_A = 0.25$ and $\mu_B = 0.75$.
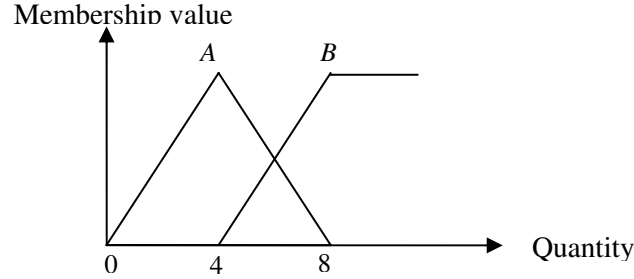

Fig. 3. A triangular membership function

Now we propose a sensitive rule hiding algorithm that hides all the sensitive fuzzified association rules by decreasing the support value of the right hand side from the dataset. We apply the following steps in order to hide the sensitive fuzzified association rules. The following abbreviations are used in the proposed algorithm.

Min_Sup: Minimum Support value;

Min_Conf: Minimum Confidentiality value;

SC: Support Count;

AR: Attribute Region;

SC_AR: Support Count of the Attribute Region;

S: attribute region Set;

SR: Set of sensitive Rules;

$C \Rightarrow D$:   rule generated from the attribute region set S;

$F_{AR}(x_i)$:   fuzzified value of the attribute region for object x_i.

Algorithm

Input: Reduced OIS, Min_Sup, Min_Conf.

Output: A transformed information system from which sensitive fuzzy association rules cannot be mined.

   Step 1. Reduced OIS.

   Step 2. Fuzzification of reduced OIS.

   Step 3. Compute SC for each *AR* in the fuzzified information system.

   Step 4. $S = \phi, \text{SR} = \phi.$

   Step 5. If (SC_AR > Min_Sup).

   Step 6. $S = S \cup \{\text{AR}\}.$

   Step 7. Else exit.

   Step 8. For S, generate all the rules.

   Step 9. Compute the Conf of each rule.

   Step 10. For each rule $C \Rightarrow D$, if (Conf $(C \Rightarrow D)$ > Min_Conf) and is sensitive.

Step 11. $\text{SR} = \text{SR} \cup \{C \Rightarrow D\}$.

Step 12. Else go to Step 27.

Step 13. For each rule $(C \Rightarrow D)$ in SR.

Step 14. For each object $x_i \in U$.

Step 15. If $F_D(x_i) > 0.5$, $F_D(x_i) > F_C(x_i)$ and $F_D(x_i) \neq 1$.

Step 16. $F_D(x_i) = 1 - F_D(x_i)$.

Step 17. Recalculate Conf $(C \Rightarrow D)$.

Step 18. If (Conf $(C \Rightarrow D)$ > Min_Conf.)

Step 19. For each object $x_i \in U$.

Step 20. If $F_D(x_i) = 1$.

Step 21. $F_D(x_i) = 0.5$.

Step 22. Recalculate Conf $(C \Rightarrow D)$.

Step 23. If (Conf $(C \Rightarrow D)$ > Min_Conf.)

Step 24. For each object $x_i \in U$.

Step 25. If $F_D(x_i) > 0.25$.

Step 26. $F_D(x_i) = F_D(x_i) - 0.25$.

Step 27. Output the transformed information system, from which the sensitive rules cannot be mined.

Step 28. End.

5. An empirical study on marketing strategies

In this section we demonstrate the proposed model by considering a real life problem for hiding sensitive information. We consider the case study, in which we evaluate the different cosmetic company's business strategies in a country. In Table 2, given below, we consider a few parameters for business strategies to get maximum sales, their possible range of values and a fuzzy proximity relation as defined in (8), which characterizes the relationship between parameters.

The companies, having high expenditure in marketing, advertisement, distribution, miscellaneous, research and development, are the ideal cases for getting maximum sales. But such a blend of cases is rare in practice. So, a company may not excel in all the parameters in order to get maximum sales. However, out of these parameters, some parameters may have greater influence on others. But the attribute values on these parameters obtained are almost indiscernible and hence, they can be classified by using a rough set on fuzzy approximation space and ordering rules. The companies are judged by the sales output that is produced. The amount of sales is judged by the different parameters of the companies. These parameters form the attribute set for our analysis. Here the marketing expenditure means all expenditure incurred for corporate promotion, which includes event marketing, sales promotion, direct marketing, etc. which comes up to around 6%. The advertising expenditure includes promotional activities using various media

like television, newspaper, Internet, etc., which comes to about 36%. The miscellaneous expenditure is mainly incurred through activities like corporate social responsibility and it leads to maximum of 28%. The distribution costs include expenses on logistic, supply chain, etc. and it comes around 24%. The investment made on new product development and other research activities taken on research and development activities, and it takes around 6%. The last one, the sales which basically deal with the sales that a company can produce after investing the expenditure in different fields, above mentioned. The company can observe the profit by subtracting the value of the total expenditure from the value of the total sales. The data collected from ten different companies is considered to be the representative figure and tabulated below in Table 3. Here we use the notation x_i, $i = 1, 2, 3, \cdots, 10,$ for different companies for the purpose of our study to demonstrate the proposed sensitive rule hiding model. It is to be noted that, in the information table all non-ratio figures shown in Table 3, are ten million INR.

Table 2. Notation representation table

Parameter	Attribute	Possible range	Parameter	Attribute	Possible range
Expenditure on marketing	Mkt	[1-150]	Expenditure on miscellaneous	Misc	[1-700]
Expenditure on advertisement	Advt	[1-900]	Expenditure on research and development	R&D	[1-150]
Expenditure on distribution	Dist	[1-600]	Sales	Sales	[1-12000]

Table 3. Sample information system

Company	Mkt	Advt	Dist	Misc	R&D	Sales
x_1	18.276	162.236	30.236	72.146	9.156	1220.586
x_2	2.076	5.393	6.793	8.290	0.383	215.767
x_3	0.496	1.330	0.433	2.733	0.393	42.593
x_4	0.940	0.060	0.666	5.890	1.243	166.41
x_5	27.333	38.660	16.496	24.343	1.523	561.697
x_6	7.033	866.916	508.676	637.530	38.963	11449.56
x_7	4.323	4.173	1.753	3.176	0.003	60.89
x_8	38.516	40.046	3.126	8.026	0.056	303.57
x_9	0.466	0.460	0.993	3.803	0.053	62.836
x_{10}	0.603	0.036	0.393	0.613	0.016	20.523

5.1. Pre-process of empirical study

In this section we discuss in detail the subsequent steps of the pre-process for the empirical study taken under consideration. A target dataset for analysis, as shown in Table 3, is considered. We have designed fuzzy proximity relations based on the attributes and computed the almost similarity between them. The fuzzy proximity relation identifies the almost indiscernibility among the objects. This result induces

the equivalence classes. The fuzzy proximity relation $R_i, i = 1, 2, 3, 4, 5, 6$, corresponding to the attributes Mkt, Advt, Dist, Misc, R&D and sales are calculated. We present the fuzzy proximity relation to the attribute for the attribute Mkt in Table 4. Keeping in mind the length of the paper, the computation of the other fuzzy proximity relations for the attributes Advt, Dist, Misc, R&D, and Sales are omitted.

Table 4. Fuzzy proximity relation for the attribute marketing

R_1	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
x_1	1.000	0.602	0.526	0.549	0.901	0.778	0.691	0.822	0.525	0.532
x_2	0.602	1.000	0.693	0.812	0.571	0.728	0.824	0.551	0.683	0.725
x_3	0.526	0.693	1.000	0.845	0.518	0.566	0.603	0.513	0.984	0.951
x_4	0.549	0.812	0.845	1.000	0.533	0.618	0.679	0.524	0.831	0.891
x_5	0.901	0.571	0.518	0.533	1.000	0.705	0.637	0.915	0.517	0.522
x_6	0.778	0.728	0.566	0.618	0.705	1.000	0.881	0.654	0.562	0.579
x_7	0.691	0.824	0.603	0.679	0.637	0.881	1.000	0.601	0.597	0.622
x_8	0.822	0.551	0.513	0.524	0.915	0.654	0.601	1.000	0.512	0.515
x_9	0.525	0.683	0.984	0.831	0.517	0.562	0.597	0.512	1.000	0.936
x_{10}	0.532	0.725	0.951	0.891	0.522	0.579	0.622	0.515	0.936	1.000

Now by considering the almost similarity of 90%, i.e., $\alpha \geq 0.90$, it is observed from Table 4 that $R_1(x_1, x_1) = 1$; $R_1(x_1, x_5) = 0.901$; $R_1(x_2, x_2) = 1$; $R_1(x_3, x_3) = 1$; $R_1(x_3, x_9) = 0.984$; $R_1(x_3, x_{10}) = 0.951$; $R_1(x_4, x_4) = 1$; $R_1(x_5, x_5) = 1$; $R_1(x_5, x_8) = 0.915$; $R_1(x_6, x_6) = 1$; $R_1(x_6, x_7) = 0.881$. Thus, the companies x_1, x_5, x_8 are α-identical. Similarly, x_3, x_9, x_{10} are α-identical; x_2 is α-identical; x_4 is α-identical; x_6 is α-identical and x_7 is α-identical. Therefore, we get

$$U/R_1^\alpha = \{\{x_1, x_5, x_8\}, \{x_2\}, \{x_3, x_9, x_{10}\}, \{x_4\}, \{x_6\}, \{x_7\}\}$$

Therefore, the values of the attribute expenditure on marketing are classified into six categories, namely low, average, medium, high, very high, and outstanding. Thus, it can be ordered. Similarly, the different equivalence classes obtained for the attributes Advt, Dist, Misc, R&D, and Sales are given below.

$$U/R_2^\alpha = \{\{x_1\}, \{x_2, x_7\}, \{x_3\}, \{x_4\}, \{x_5, x_8\}, \{x_6\}, \{x_9\}, \{x_{10}\}\},$$
$$U/R_3^\alpha = \{\{x_1\}, \{x_2\}, \{x_3, x_{10}\}, \{x_4, x_9,\}, \{x_5\}\{x_6\}, \{x_7\}, \{x_8\}\},$$
$$U/R_4^\alpha = \{\{x_1\}, \{x_2, x_4, x_8\}, \{x_3, x_7, x_9\}\{x_5\}, \{x_6\}, \{x_{10}\}\},$$
$$U/R_5^\alpha = \{\{x_1\}, \{x_2, x_3\}, \{x_4, x_5\}, \{x_6\}, \{x_7\}, \{x_8, x_9\}, \{x_{10}\}\},$$
$$U/R_6^\alpha = \{\{x_1\}, \{x_2, x_4, x_8\}, \{x_3, x_7, x_9\}, \{x_5\}, \{x_6\}, \{x_{10}\}\}.$$

From the above classification, it is clear that the values of the attribute expenditure on advertisement and distribution are classified into eight categories, namely poor, very low, low, average, medium, high, very high and outstanding. The values of the attribute expenditure on miscellaneous are classified into six

64

categories, namely low, average, medium, high, very high and outstanding. The values of the attribute expenditure on research and development are classified into seven categories, namely very low, low, average, medium, high, very high and outstanding. Finally, the values of the attribute sales are classified into six categories, namely low, average, medium, high, very high and outstanding. The ordered information system of the business strategies of different cosmetic companies of Table 3 is given below in Table 5. On considering the weights of outstanding, very high, high, medium, average, low, very low and poor as 8, 7, 6, 5, 4, 3, 2 and 1 respectively, the OIS is given below in Table 5.

Table 5. Ordered information system

Company	Mkt	Advt	Dist	Misc	R&D	Sales
x_1	Outstanding (8)	Very high (7)	Very high (7)	Very high (7)	Very high (7)	Very high (7)
x_2	Medium (5)	Medium (5)	Medium (5)	Medium (5)	Medium (5)	Medium (5)
x_3	Low (3)	Average (4)	Poor (1)	Average (4)	Medium (5)	Average (4)
x_4	Average (4)	Very low (2)	Very low (2)	Medium (5)	High (6)	Medium (5)
x_5	Outstanding (8)	High (6)	High (6)	High (6)	High (6)	High (6)
x_6	Very high (7)	Outstanding (8)	Outstanding (8)	Outstanding (8)	Outstanding (8)	Outstanding (8)
x_7	High (6)	Medium (5)	Low (3)	Average (4)	Very Low (2)	Average (4)
x_8	Outstanding (8)	High (6)	Average (4)	Medium (5)	Average (4)	Medium (5)
x_9	Low (3)	Low (3)	Very low (2)	Average (4)	Average (4)	Average (4)
x_{10}	Low (3)	Poor (1)	Poor (1)	Low (3)	Low (3)	Low (3)

$\prec_{Mkt}$: Outstanding $\prec$ Very high $\prec$ high $\prec$ Medium $\prec$ Average $\prec$ Low

$\prec_{Adv}$: Outstanding $\prec$ Very high $\prec$ High $\prec$ Medium $\prec$ Average $\prec$ Low $\prec$ Very low $\prec$ Poor

$\prec_{Dist}$: Outstanding $\prec$ Very high $\prec$ High $\prec$ Medium $\prec$ Average $\prec$ Low $\prec$ Very low $\prec$ Poor

$\prec_{Misc}$: Outstanding $\prec$ Very high $\prec$ high $\prec$ Medium $\prec$ Average $\prec$ Low

$\prec_{R\&D}$: Outstanding $\prec$ Very high $\prec$ High $\prec$ Medium $\prec$ Average $\prec$ Low $\prec$ Very low

$\prec_{Sales}$: Outstanding $\prec$ Very high $\prec$ High $\prec$ Medium $\prec$ Average $\prec$ Low

5.2. Post-process of empirical study

In this section we discuss the subsequent steps of the post-process for the empirical study taken into consideration. The objective of this process is to hide fuzzy sensitive rules that are generated through fuzzification from the ordered information system. The fuzzified information system, which is generated from the ordered information system using triangular membership function (11) and trapezoidal membership (12), is given in Table 6.

Table 6. Fuzzified information system

Comp.	Mkt		Advt		Dist		Misc		R&D		Sales	
	Mkt_A	Mkt_B	$Advt_A$	$Advt_B$	$Dist_A$	$Dist_B$	$Misc_A$	$Misc_B$	$R\&D_A$	$R\&D_B$	$Sales_A$	$Sales_B$
x_1	0	1	0.25	0.75	0.25	0.75	0.25	0.75	0.25	0.75	0.25	0.75
x_2	0.75	0.25	0.75	0.25	0.75	0.25	0.75	0.25	0.75	0.25	0.75	0.25
x_3	0.75	0	1	0	0.25	0	1	0	0.75	0.25	1	0
x_4	1	0	0.5	0	0.5	0	0.75	0.25	0.5	0.5	0.75	0.25
x_5	0	1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
x_6	0.25	0.75	0	1	0	1	0	1	0	1	0	1
x_7	0.5	0.5	0.75	0.25	0.75	0	1	0	0.5	0	1	0
x_8	0	1	0.5	0.5	1	0	0.75	0.25	1	0	0.75	0.25
x_9	0.75	0	0.75	0	0.5	0	1	0	1	0	1	0
x_{10}	0.75	0	0.25	0	0.25	0	0.75	0	0.75	0	0.75	0
S	4.75	4.5	5.25	3.25	4.75	2.5	6.75	3	6	3.25	6.75	3

Considering the minimum support 3.5 and minimum confidentiality value 80%, the regions Mkt_A, Mkt_B, $Advt_A$, $Dist_A$, $Misc_A$, $R\&D_A$ and $Sales_A$ have their support values greater than the minimum support. These attribute regions are considered in forming the rules and finding the corresponding confidentiality values. The two frequent item sets generated are $\{Mkt_A, Advt_A\}$, $\{Mkt_A, Dist_A\}$, $\{Mkt_A, Misc_A\}$, $\{Mkt_A, R\&D_A\}$, $\{Mkt_A, Sales_A\}$, $\{Mkt_B, Advt_A\}$, $\{Mkt_B, Dist_A\}$, $\{Mkt_B, Misc_A\}$, $\{Mkt_B, R\&D_A\}$, $\{Mkt_B, Sales_A\}$, $\{Advt_A, Dist_A\}$, $\{Advt_A, Misc_A\}$, $\{Advt_A, R\&D_A\}$, $\{Advt_A, Sales_A\}$, $\{Dist_A, Misc_A\}$, $\{Dist_A, R\&D_A\}$, $\{Dist_A, Sales_A\}$, $\{Misc_A, Sales_A\}$, $\{Misc_A, R\&D_A\}$, $\{R\&D_A, Sales_A\}$. These two frequent item sets generate many fuzzy association rules. Some of them are $Mkt_A \Rightarrow Advt_A$, $Mkt_A \Rightarrow Misc_A$, $Dist_A \Rightarrow Mkt_A$, $Misc_A \Rightarrow Mkt_A$, $Advt_A \Rightarrow Dist_A$, $Advt_A \Rightarrow Misc_A$, $Dist_A \Rightarrow Advt_A$, $Advt_A \Rightarrow R\&D_A$ and $Dist_A \Rightarrow Misc_A$. To illustrate the post-process, we consider the fuzzy association rule $Dist_A \Rightarrow Misc_A$. In order to compute the confidentiality of the fuzzy association rule, we present the fuzzy values of $Dist_A$, $Misc_A$ and the support of $Dist_A \Rightarrow Misc_A$ in the following Table 7. From the table it is clear that:

Support count ($Dist_A \Rightarrow Misc_A$) = 4.5;

Confidentiality $(Dist_A \Rightarrow Misc_A) = \dfrac{Sup\,(Dist_A \Rightarrow Misc_A)}{Sup\,(Dist_A)} = \dfrac{4.5}{4.75} = 94.7\%.$

This indicates that the fuzzy association rule $Dist_A \Rightarrow Misc_A$ is a sensitive strong fuzzy association rule and it has to be hidden.

Table 7. Computation of the support and support count of $\text{Dist}_A \Rightarrow \text{Misc}_A$

Company	Dist_A	Misc_A	Support of ($\text{Dist}_A \Rightarrow \text{Misc}_A$)
x_1	0.25	0.25	0.25
x_2	0.75	0.75	0.75
x_3	0.25	1	0.25
x_4	0.5	0.75	0.5
x_5	0.5	0.5	0.5
x_6	0	0	0
x_7	0.75	1	0.75
x_8	1	0.75	0.75
x_9	0.5	1	0.5
x_{10}	0.25	0.75	0.25
SC	4.75	6.75	4.5

By employing the post-process algorithm discussed in Section 4.2, the fuzzified values of Misc_A are altered. The modified computation of the support and support count of the related sensitive fuzzy association rule are given in Table 8.

Table 8. Modified computation of the support and support count of $\text{Dist}_A \Rightarrow \text{Misc}_A$

Company	Dist_A	Misc_A	Support of ($\text{Dist}_A \Rightarrow \text{Misc}_A$)
x_1	0.25	0.25	0.25
x_2	0.75	0.5	0.5
x_3	0.25	0.25	0.25
x_4	0.5	0.25	0.25
x_5	0.5	0.25	0.25
x_6	0	0	0
x_7	0.75	0.25	0.25
x_8	1	0.5	0.5
x_9	0.5	0.25	0.25
x_{10}	0.25	0.25	0.25
SC	4.75	2.75	2.75

$$\text{Confidentiality } (\text{Dist}_A \Rightarrow \text{Misc}_A) = \frac{\text{Sup} (\text{Dist}_A \Rightarrow \text{Misc}_A)}{\text{Sup} (\text{Dist}_A)} = \frac{2.75}{4.75} = 57.8\%.$$

Since the confidentiality of the rule is found to be 89.4%, we observe that the sensitive rule is hidden successfully. The modified fuzzified information system, in which the sensitive rule $\text{Dist}_A \Rightarrow \text{Misc}_A$ is hidden, is presented in Table 9. After defuzzification, the ordered information system, in which the sensitive rule $\text{Dist}_A \Rightarrow \text{Misc}_A$ is hidden, is presented in Table 10.

Table 9. Modified fuzzified information system

Company	Mkt		Advt		Dist		Misc		R&D		Sales	
	Mkt_A	Mkt_B	$Advt_A$	$Advt_B$	$Dist_A$	$Dist_B$	$Misc_A$	$Misc_B$	$R\&D_A$	$R\&D_B$	$Sales_A$	$Sales_B$
x_1	0	1	0.25	0.75	0.25	0.75	0.25	0.75	0.25	0.75	0.25	0.75
x_2	0.75	0.25	0.75	0.25	0.75	0.25	0.5	0.25	0.75	0.25	0.75	0.25
x_3	0.75	0	1	0	0.25	0	0.25	0	0.75	0.25	1	0
x_4	1	0	0.5	0	0.5	0	0.25	0.25	0.5	0.5	0.75	0.25
x_5	0	1	0.5	0.5	0.5	0.5	0.25	0.5	0.5	0.5	0.5	0.5
x_6	0.25	0.75	0	1	0	1	0	1	0	1	0	1
x_7	0.5	0.5	0.75	0.25	0.75	0	0.25	0	0.5	0	1	0
x_8	0	1	0.5	0.5	1	0	0.5	0.25	1	0	0.75	0.25
x_9	0.75	0	0.75	0	0.5	0	0.25	0	1	0	1	0
x_{10}	0.75	0	0.25	0	0.25	0	0.25	0	0.75	0	0.75	0
S	4.75	4.5	5.25	3.25	4.75	2.5	2.75	3	6	3.25	6.75	3

Table 10. Ordered information system, in which the sensitive rule $Dist_A \Rightarrow Misc_A$ is hidden

Company	Mkt	Advt	Dist	Misc	R&D	Sales
x_1	Outstanding (8)	Very high (7)	Very high (7)	Very high (7)	Very high (7)	Very high (7)
x_2	Medium (5)	Medium (5)	Medium (5)	Medium (5)	Medium (5)	Medium (5)
x_3	Low (3)	Average (4)	Poor (1)	Poor (1)	Medium (5)	Average (4)
x_4	Average (4)	Very low (2)	Very low (2)	Medium (5)	High (6)	Medium (5)
x_5	Outstanding (8)	High (6)	High (6)	High 6)	High (6)	High (6)
x_6	Very high (7)	Outstanding (8)	Outstanding (8)	Outstanding (8)	Outstanding (8)	Outstanding (8)
x_7	High (6)	Medium (5)	Low (3)	Poor (1)	Very Low (2)	Average (4)
x_8	Outstanding (8)	High (6)	Average (4)	Medium (5)	Average (4)	Medium (5)
x_9	Low (3)	Low (3)	Very low (2)	Poor (1)	Average (4)	Average (4)
x_{10}	Low (3)	Poor (1)	Poor (1)	Poor (1)	Low (3)	Low (3)

6. Experimental results

To explore the characteristics and performance of the proposed model, experimental analysis has been conducted. By varying the measuring values, the system was proved to be stable. The experiments are conducted on a computer with the following configuration: Intel Pentium Processor, 1GB RAM and Windows XP operating system. We have run the model on a database of an information system of business strategies, collected from Prowess-CMIE, with all of its attributes being

numerical. Prowess maintains the database of financial performances of Indian companies.
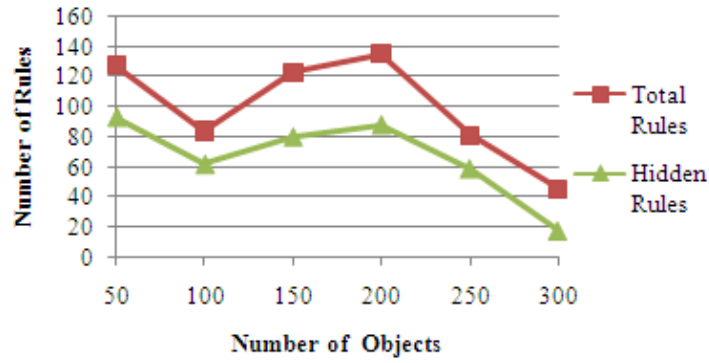
Fig. 4. Number of objects versus number of rules generated
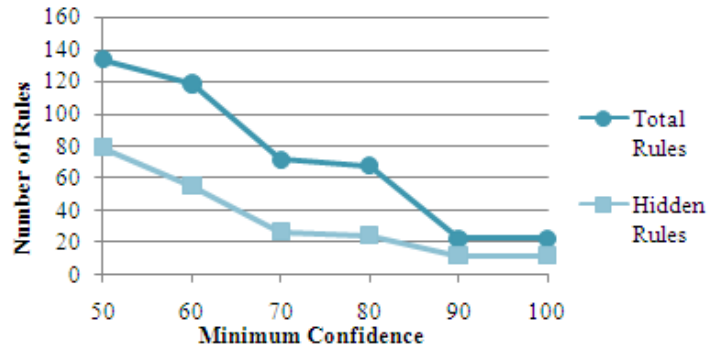
Fig. 5. Rules generated according to various minimum confidentiality
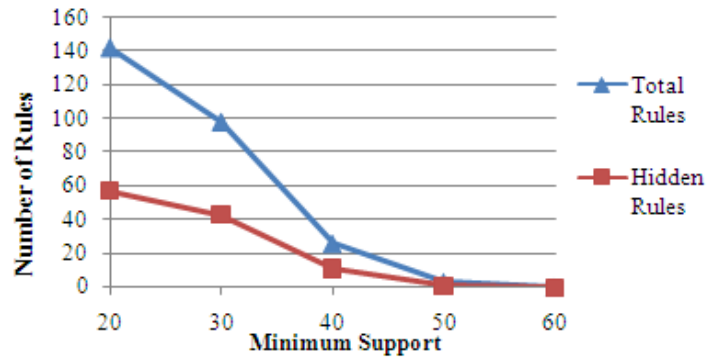
Fig. 6. Rules generated by considering various minimum support

Fig. 4 depicts the relationship of the total number of rules and hidden rules according to the number of objects. The rules generated by setting up a minimum support of 30 and minimum confidence of 80 illustrate that if the number of objects is increased, the number of rules will not amplify and also that the hidden rules are proportional to the total rules generated. When the value of the minimum support or

minimum confidence is escalated, the rules generated are reduced. Fig. 5 illustrates the model operation over an information system of 300 objects and minimum support of 20. Fig. 6 shows the functions of the system by considering minimum confidence of 75 over an information system of 100 objects.

From the experimental results, it can be easily seen that our algorithm hides more rules for different values of minimum support and minimum confidence. Unlike the previous approaches, which hide the association rules in the quantitative database [17], our approach reduces the attributes in the pre-process and hides the sensitive rules efficiently in the post-process.

7. Conclusion

In this paper we propose a privacy preserving data mining model by hiding the fuzzy association rules. The proposed model handles the task of hiding the fuzzy association rules in quantitative datasets after reducing the superfluous attributes from the information system. For this purpose it uses two processes, named as pre-process and post-process. In the pre-process, a rough set on fuzzy approximation space and ordering rules are employed for the reduction of the attributes that have no influence in the information system. In the post-process, the concept of a fuzzy set is applied to mine fuzzy association rules and to hide sensitive rules. Numerical experiments have been conducted on Prowess-CMIE database to demonstrate the viability of the proposed research. The results of the proposed approach are consistent and, hence, encouraging.

R e f e r e n c e s

1. S w e e n e y, L. Computational Disclosure Control: A Primer on Data Privacy Protection. Ph.D. Thesis. Massachusetts Institute of Technology, 2001.
2. S w e e n e y, L. Achieving k-Anonymity Privacy Protection Using Generalization and Suppression. – International Journal on Uncertainty, Fuzziness and Knowledge-Based Systems, Vol. **10**, 2002, No 5, 571-588.
3. M a c h a n a v a j j h a l a, A., D. K i f e r, J. G e h r k e, M. V e n k i t a s u b r a m a n i a m. L-Diversity: Privacy Beyond K-Anonymity. – ACM Trans. Knowl. Discov. Data, Vol. **1**, 2007, No 1, 1-52.
4. T i a n, H., W. Z h a n g. Extending l-Diversity for Better Data Anonymization. – In: 6th International Conference on Information Technology: New Generations, ITNG 2009, IEEE Xplore, 2009, 461-466.
5. L i, N., T. L i, S. V e n k a t a s u b r a m a n i a n. t-Closeness: Privacy Beyond k-Anonymity and l-Diversity. – In: 23rd International Conference on Data Engineering, IEEE Xplore, 2007, 106-115.
6. L i u, L., M. K a n t a r c i o g l u, B.T h u r a i s i n g h a m. The Applicability of the Perturbation Based Privacy Preserving Data Mining for Real-World Data. – Data and Knowledge Engineering, Vol. **65**, 2008, No 1, 5-21.
7. L i, X., B. S u m i t S a r k a r. A Tree-Based Data Perturbation Approach for Privacy-Preserving Data Mining. – Knowledge and Data Engineering, Vol. **18**, 2006, No 9, 1278-1283.
8. W u, X., C h u, C h a o-H s i e n C h u, Y. W a n g, F. L i u, D. Y u e. Privacy Preserving Data Mining Research: Current Status and Key Issues. – In: Yong Shi, G. D. Van Albada, J. Dongarra, M. A. Peter, Eds. LNCS 4489. Springer, 2007, 762-772.

9. A g g a r w a l, C. C., P h i l i p S. Y u. A General Survey of Privacy Preserving Data Mining Models and Algorithms. – In: C. C. Aggarwal, Philip S. Yu., Eds. Privacy-Preserving Data Mining. Boston, Kluwer Academic Publishers, 2008, 11-52.

10. Z a d e h, L. A. Fuzzy Sets. – Information and Control, Vol. **8**, 1965, 338-353.

11. P a w l a k, Z. Rough Sets. – International Journal of Computer Information Science, Vol. **11**, 1982, 341-356.

12. A t a n a s o v, K. T. Intuitionistic Fuzzy Sets. – Fuzzy Sets and Systems, Vol. **20**, 1986, 87-96.

13. A c h a r j y a, D. P., B. K. T r i p a t h y. Rough Sets on Fuzzy Approximation Spaces and Applications to Distributed Knowledge Systems. – International Journal of Artificial Intelligence and Soft Computing, Vol. **1**, 2008, No 1, 1-14.

14. A c h a r j y a, D. P., B. K. T r i p a t h y. Rough Sets on Intuitionistic Fuzzy Approximation Spaces and Knowledge Representation. – International Journal of Artificial Intelligence and Computational Research, Vol. **1**, 2009, No 1, 29-36.

15. T r i p a t h y, B. K., D. P. A c h a r j y a. Association Rule Granulation Using Rough Sets on Intuitionistic Fuzzy Approximation Spaces and Granular Computing. – Annals. Computer Science Series, Vol. **9**, 2011, No 1, 125-144.

16. A c h a r j y a, D. P., L. E z h i l a r s i. A Knowledge Mining Model for Ranking Institutions Using Rough Computing with Ordering Rules and Formal Concept Analysis. – International Journal of Computer Science Issues, Vol. **8**, 2011, No 2, 417-425.

17. M a n o j G u p t a, R. C. J o s h i. Privacy Preserving Fuzzy Association Rules Hiding in Quantitative Data. – International Journal of Computer Theory and Engineering, Vol. **1**, October 2009, No 4, 382-388.

18. A h n, B. S., S. S. C h o, C. Y. K i m. The Integrated Methodology of Rough Set Theory and Artificial Neural Network for Business Failure Prediction. – Expert Systems with Applications, Vol. **18**, 2000, 65-74.

19. A s h l o c k, D. Evolutionary Computation for Modeling and Optimization. Springer, 2006.

20. T r i p a t h y, B. K., D. P. A c h a r j y a. Knowledge Mining Using Ordering Rules and Rough Sets on Fuzzy Approximation Spaces. – International Journal of Advances in Science and Technology, Vol. **1**, 2010, No 3, 41-50.

21. A c h a r j y a, D. P., D e b a s r i t a R o y, Md. A. R a h a m a n. Prediction of Missing Associations Using Rough Computing and Bayesian Classification. – International Journal of Intelligent Systems and Applications, Vol. **4**, 2012, No 11, 1-13.

22. P a w l a k, Z. Rough Sets: Theoretical Aspects of Reasoning about Data. London, Kluwer Academic Publishers, 1991.

23. A g r a w a l, R., T. M a r l i n s, A. S w a m i. Mining Association Rules between Sets of Items in Large Databases. – In: International Conference on Management of Data, Proceeding of ACM-SIGMOD, Washington, DC, 1993, 207-216.

24. H i p p, J., U. G u n t z e r, G. N a k h a e i z a d e h. Algorithms for Association Rule Mining – A General Survey and Comparison. – SIGKDD Explorations, Vol. **2**, 2000, No 2, 1-58.

25. W a n g, S h y u e-L i a n g., Yu-H u e i L e e, S. B i l l i s, A. J a f a r i. Hiding Sensitive Items in Privacy Preserving Association Rule Mining. – International Conference on Systems, Man and Cybernetics, IEEE Xplore, Vol. **4**, 2004, 3239-3244.

26. W a l c z a k, B., D. L. M a s s a r t. Rough Set Theory. – Chemometrics and Intelligent Laboratory, Vol. **47**, 1999, No 1, 1-16.

27. S a l e e m D u r a i, M. A., D. P. A c h a r j y a, A. K a n n a n, N. C h. S. N. I y e n g a r. An Intelligent Knowledge Mining Model for Kidney Cancer Using Rough Set Theory. – International Journal of Bioinformatics Research and Applications, Vol. **8**, 2012, No 5/6, 417-435.