Securing Informative Fuzzy Association Rules using Bayesian Network

Muhammad Fahad[1], Khalid Iqbal[2], Somaiya Khatoon[3], Khalid Mahmood Awan[4]
Department Computer Science
COMSATS University Islamabad,
Attock Campus

***Abstract*—In business association rules being considered as important assets, play a vital role in its productivity and growth. Different business partnership share association rules in order to explore the capabilities to make effective decision for enhancement of business and core capabilities. The fuzzy association rule mining approach emerged out of the necessity to mine quantitative data regularly present in database. An association rule is sensitive when it violates few rules and regulation for sharing particular nature of information to third world. Like classical association rules, there is a need for some privacy measures to be taken for retaining the standards and importance of fuzzy association rules. Privacy preservation is used for valuable information extraction and minimizing the risk of sensitive information disclosure. Our proposed model mainly focuses to secure the sensitive information revealing association rules. In our model, sensitive fuzzy association rules are secured by identifying sensitive fuzzy item to perturb fuzzified dataset. The resulting transformed FARs are analyzed to conclude/calculate the accuracy level of our model in context of newly generated fuzzy association rules, hidden rules and lost rules. Extensive experiments are carried out in order to demonstrate the results of our proposed model. Privacy preservation of maximum number of sensitive FARs by keeping minimum perturbation highlights the significance of our model.**

***Keywords*—Fuzzy association rules; privacy preservation; fuzzification; sensitive rules; Bayesian network; perturbation**

I. Introduction

Data mining is a systematic way for extracting useful information from large repositories of data using many effective tools and techniques [1]. Association rule mining being one of most effective techniques of data mining is used for extraction of interesting association rules from large databases [2]. Association Rule Mining (ARM) is focused on extraction of recurrent item sets from crisp data but association rule mining usually looks for attribute that are categorical by nature. Real life data is combination of both categorical and numerical data and it does not focus on either categorical or numerical only. The general scheme used is to transform numerical data into categorical attributes using different ranges of data but this conversion leads to more uncertainty and multiple sharp boundary problems. To cope with this problem, fuzzy methods are taken into account which helps in transformation of numerical data into fuzzy categorical data [3], [4]. This method gives surety for zero or negligible loss of information regardless of considering particular value of numerical attributes. Consequently, the inherent uncertainty that exists in numerical database is also appropriately taken care off. Fuzzy logic [5] is an efficient technique that has been used in many fields for dealing with uncertainty that lies in every type of data. Data is uncertain by nature when it is related to humans. Therefore, there must have a suitable and useful approach to build for handling uncertainty in combination of fuzzy logic to protect privacy of association rules. Fuzzy set theory concept and association rule mining integration results in formation of fuzzy association rules which covers the drawback and sharp boundary problems of classical association rules. Apart from extraction of information or knowledge from database, few techniques are required for safeguard of information or extracted patterns. Basically, knowledge sharing will possibly involve intimidations and threats that privacy or sensitive information (in any form) could be unintentionally revealed. That is, as knowledge sharing increases it become more important for people to keep information more secure and safe regardless of the fact for what purpose and in which form it is being shared. To reduce the information disclosure risk, sensitive information from shared database should be kept out of sight. Therefore, PPDM techniques are widely used in knowledge discovery. The main goal of PPDM is to lessen the risk of ill use of data while keeping quality of data mining practices. Privacy preserving data mining was first familiarized by Agrawal [2]. Basically, there are two problems being address in PPDM [6]; First is privacy of data and second is privacy of sensitive rules (knowledge) extracted from the data. The latter problem which is called Knowledge Hiding in Database (KHD), tells how to conceal sensitive rules keeping the extraction of non-sensitive rules normal in record. This (Association rule hiding) problem is most substantial and hotspot in research field now a days. In [7], [8], [9], a detailed view of privacy preservation of data mining techniques for lessening the disclosure of sensitive information is presented. Support and confidence are two significant measures for discovery of item sets for association rules. Association rules which satisfy the required support confidence threshold, such rules are considered as interesting rules. These measures have been focused by researchers to improve privacy of association rules by manipulating support, confidence such as process of increasing support of antecedent of rules (ISL) and decrease in support of consequent of rule (DSR) [10], Decrease of Support Rule (DSR) and manipulation with support of LHS and RHS of rule [11], Decrease Support Confidence (DSC) algorithm introducing Pi-tree in [12], algorithm dealing with support and confidence framework naming Decrease Support of Sensitive Items (DSSI) [13] for improving issues of [12], [10]. Advanced Decrease Support of Sensitive Item ADSSI algorithm introduced for tackling security problem in effective way [14]. In 1965, Zadeh first introduced fuzzy set theory. This

theory is under focus of many researches of data mining to find interesting fuzzy association rules or sequential patterns in transaction data with quantitative values [15], [16], [17], extraction of fuzzy rules between key phrases in documents using Extractor Package [18] and defining significance factor and certainty factor in item sets and association rules [19]. Limitations in the support and confidence-based methodologies gave researchers the path to work on perturbation and other different approaches [8], [23] for minimizing the limitation of aforementioned framework-based techniques and enhancing the work. Firstly, these various techniques result into some flaws and limitation of generating lost and ghost rules, multiple database scans, incomplete transformation of database and side effects of adding noise in original database. Secondly, all such techniques tend to hide sensitive rules on assumption bases instead of proper certainty of particular node. ARs are considered as sensitive if they disclose valuable and critical information to third party. Thus, data mining techniques must hide such sensitive rules with proper strategy and selection with proper formulation rather than assumption. All aforementioned techniques became unable to give a solid picture for sensitive item identification for perturbation. On contrary, our proposed methodology not only identifies sensitive items accurately, but also recommends particular transactions for modification. To report this challenging problem, our work proposes a PPDM model for securing fuzzy association rules that are sensitive. Proposed agenda comprise of few steps; using fuzzy logic and membership function for preprocessing of original data set and converting it into fuzzy dataset, preparing input for K2 algorithm and apriori algorithm, determining the sensitive node/item using Bayesian network developed by K2 algorithm, hiding the considered sensitive rules, performing minimum perturbation for max. degree transaction in database using sensitive item/node. Bayesian network has been widely used in fields [19], [20]. Bayesian networks basically helps in discovering information about uncertainty of particular domain by building a probabilistic graphical structure. This particular graphical model contains vertices (V) representing items and edges (E) representing the dependencies of variables probabilistically on parallel items. Items and nodes are variables/attributes of proposed work. Bayesian networks can be developed using K2 algorithm [19]. K2 algorithm discovers relationship among items/nodes in an increasing order.

Remaining part of paper is organized as follows: Section II contains related work done by previous researchers, in Section III we proposed our methodology, Section IV is with experimental section and Section V presents conclusion.

II. Related Works

Different techniques have been used by researcher for privacy preservation of association rules and fuzzy association rules. Most of algorithms in privacy preservation mostly rely on support and confidence manipulation of rules and transformation in databases. Two algorithm increasing support of L.H.S (ISL) and Decreasing support of R.H.S (DSR) proposed by Wang et al. [10] ISL increase the support of antecedent of rule below minimum threshold while DSR with no hiding failure, lowers the support and confidence of rule for hiding the sensitive rules. Wang et al. [10], proposed two hiding algorithms ISL (Increase Support of LHS) and DSR (Decrease Support of RHS) algorithm. ISL through rising support of rules' LHS confidence will be reduced under the threshold; as a result, the sensitive association rules will be unseen. DSR decreases the whole rule's support and confidence below the threshold to hide sensitive association rules. DSR has no hiding failure; notwithstanding, whereas ISL will fail if there is no suitable transaction to add. Two more algorithm introduced by Wang et al. [21] naming Decrement in Confidence by Decrease support (DCDS) and decrement in confidence by increase support (DCIS). This algorithm puts combined recommendation association rules under curtain. Rule hiding is done by decreasing support of suggested item keeping others items support constant and increment in support of non-suggested items while keeping support of suggested items constant in (DCDS) and (DCIS) algorithms respectively. Wang et al. [12] presented another support confidence-based approach (DSC) Decrease support and confidence with multiple scan of database in Pi-tree structure. In [14], Chang and Chen proposed an improved version of [12] for securing complex rules sets naming DSSI (Decrease support of sensitive item set) [13] and also covering its limitation. DSSI is more effectual as compared to another algorithm as it requires only one scan of database and left no rule unhidden. However, it results in removal of some non-sensitive rules which minimizes the impact of its benefit over other algorithm. Another support confidence-based approach ADSSRC Advance Decrease Support of Right-Hand Side item of Rule Cluster) is proposed by Modi et al. [22] which start hiding process after makes cluster of assumed sensitive items picked from RHS hence reducing the number of alterations in original database. A Bayesian network-based model has been used as well in privacy preservation [23]. In [11], different fuzzy based mapping approaches are analyzed in context of privacy preservation aspect and abilities to maintain relationship with other fields. In [24], a technique in privacy preserving manner is proposed to find out comprehensive fuzzy rules with the similar traits from share data. In [29], a fuzzy c-regression method is used to produce artificial data generation process which helps third parties to statistical calculation with a narrow disclosure risk. Hong et al. [26], proposed an algorithm that extracts exciting fuzzy association rules from given transaction set by incorporating concepts of apriori algorithm and fuzzy set idea. This algorithm focused only the particular fuzzy region having more support than minimum support used to frame the rules. this algorithm claims better time complexity. Berberoglu [28], proposed a unique scheme to hide sensitive fuzzy association rules. This process is done by increment in support threshold of antecedent part of rule which later, results in lowering confidence of particular rule. Manoj Gupta et al. [30], proposed quantitative rules hiding by lowering the support after performing fuzzification with random membership function. This method needs predefined membership function for fuzzification and are typically constructed by human experts. Hameed et al. [16] proposed a framework for privacy preservation of fuzzy association rules. This framework (PPFAR) is based on fuzzy correlation study. Interesting association rules are highlighted and considered as sensitive by fuzzy set integration with aprioiri and fuzzy correlation study. Experimental results demonstrate that PPFAR approach tends to hide informative rules with minor level alterations and sustains quality of the modified dataset. Another technique for hiding sensitive FAR is introduced in [21] which extract fuzzified data with help of modified aprioiri algorithm. DSR (Decrease support of Right-hand side of rule) approach is used for hiding sensitive

rules. Chan and Au introduced an algorithm F-APACS in [27] for extraction of fuzzy association rule. Transformation from quantitative to linguistic data is performed in this algorithm. Furthermore, process of accustomed difference analysis is used. It helped in finding out the most interesting associations among variables. In [34] a secure framework for privacy preserving fuzzy co-clustering is proposed for handling both vertically and horizontally distributed co-occurrence matrices. A method to hide fuzzy association rule is proposed in [31] using modified apriori algorithm in order to identify sensitive rules to be hidden. A perturbation approach, algorithm Fast hiding sensitive association rules (FHSAR) proposed by Weng et al. [25]. This algorithm minimizes the execution time for hiding sensitive rules by taking single scan of database. All transactions are assessed one by one to capably select items for modification purpose. In addition to hide sensitive rules, Index table-based transformation of frequent item set is used by framework proposed by Wu et al. [32]. It is done for retrieving rules rapidly. In technique [20] process of binning and field rotation are used for securing sensitive information after perturbation in database. This methodology sustained the novelty of data even after recognition of items required for perturbation. In [33], [34], multiple rules hiding technique is proposed that requires two scans of database irrespective of sensitive items. Index table files are generated for speed up of locating sensitive transaction in firs scan whereas hiding algorithm is applied in dataset in second scan. A border-based approach is introduced by X. Sun [35] which is used to efficiently evaluate influence of any manipulation or alteration to the original database throughout hiding procedure. Alteration is performed with minimum side effects keeping the quality of database constant. Later on Sun [35] introduced enhanced usefulness of the work presented in [36]. Hybrid approaches like [37], EMO algorithm is also used for overcoming the failure of previous techniques-based algorithms by combining hybrid technique (e.g. distortion and genetic algorithm). Despite of this, optimal solution is NP-hard problem but there is always focus on optimal solution or methods for the discussed problem.

III. Proposed Methodology

Our proposed framework presented in Fig. 1 contains quantitative dataset. Initially original dataset is converted into fuzzified data using membership function $MF(i) : D(i) \rightarrow [0, 1]$. This transformation is resulted after fuzzy set intervals of every attribute of original dataset $D_{(org)}$. The resulting data is again processed into binary data $D_{(bin)}$ and transactional dataset $D_{(T)}$. Binary table represents presence and absence of item using binary digit 0 and 1. Transactional data is used as input for apriori algorithm to generate efficient fuzzy association rules FARS. Binary table D(bin) is prepared as input for K2 algorithm. K2 algorithm is used to get the corresponding node in order of the occurrence of items. Conditional Probability scores are generated using random orders of variables of binary table. Those scores help in measuring items dependency column by column on each other. Thus, this table is recorded during the execution of K2 algorithm. Probabilistic dependency and highest computation of particular item throughout this Bayesian network structure is fundamentally used to locate the sensitive item (I_S) and transform the transactional bi-vector dataset after minimum perturbation. The released or reformed dataset (D_R) is used as input again to apriori algorithm keeping the support and confidence constant to extract modified and non-sensitive rules (R_{NS}).

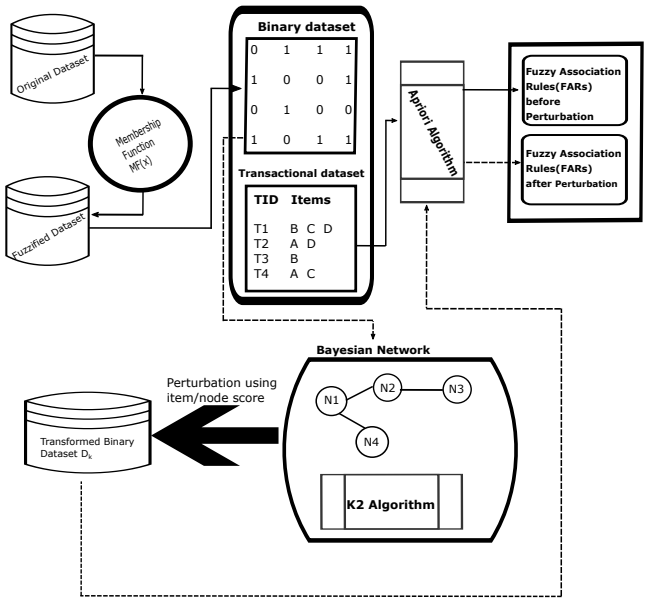

Fig. 1. Bayesian network-based Privacy preservation of sensitive fuzzy association rules

A. Important Components of the Proposed Model

After brief discussion of proposed model, few important components of this model are discussed below:

1) Apriori algorithm: Apriori algorithm, first introduced by agarwal [2] is extensively used and was one of the first attempt for extracting association rules from large datasets. This algorithm is count sub class of DFS algorithm. Though generating interesting association rules, this algorithm is unable to protect disclosure of sensitive rules.

2) Membership function MF(x): Membership function introduced by Lofti A. Zadeh, is a curve that determines how ever input value is mapped to degree of membership from 0 to 1. Gaussian membership function is effective function for achieving the smoothness in curve and removing uncertainty in data series. Gaussian membership function is computed in eq. (1):

$$gmf(x) = e^{-0.5(y)^2}; y = 8[\frac{X - X_{min}}{X_{max} - X_{min}}]4, x > 0 \quad (1)$$

Where X_{min} = Minimum value in given data series; X_{max} = Largest value in given data series; X= current value that is in queue to be fuzzified.

3) Fuzzy table: Fuzzy table is collection of record which is fuzzy in nature and is calculated after applying membership function using multiple intervals (expert domain based) as shown in Table I.

TABLE I. FUZZIFIED DATA EXAMPLE

TID	Age			Income		
	Low	Young	Old	Low	Average	High
T1	0.3	0	0	0.7	0	0
T2	0.122	0.55	0	0.8	0.1	0.002
T3	0	0.7	0.12	0	0.45	0.34
T4	0	0	0.67	0	0	0.9
T5	0	0.78	1	0.1	0.2	0

4) Fuzzy association rule (FARs): Fuzzy association rule mining is used for extracting association rules with fuzzy set concept such that the quantitative values can be controlled. Fuzzy association rules use fuzzy logic to convert the numerical, quantitative and crisp values into linguistic terms. They basically split information into multiple domains and represent it into more descriptive form.
$Customer_{age=young,marital\_status=widow} \rightarrow Loan_{Balance=low}$

5) Binary table of transactions: Table II represents particular item's presence and absence in form of 0s and 1s in every transaction.

TABLE II. BINARY TABLE

TID	A			B		
	A,1	A,2	A,3	B,1	B,2	B,3
T1	1	0	0	1	0	0
T2	1	1	0	1	1	0
T3	0	1	1	0	1	1
T4	0	0	1	0	0	1
T5	0	1	1	0	1	0

6) K2 Algorithm with Bayesian network: K2 algorithm [19] can be used to establish Bayesian network structure in increasing order. Bayesian networks are denoted in a (DAG) Directed Acyclic Graph according to given eq. (2)

$$f(i,\pi_i)=\prod_{i=1}^{n}\prod_{j=1}^{q_i}\frac{(r_i-1)!}{(N_{ij}+r_i-1)!}\prod_{k=1}^{r_i}\alpha_{ijk}! \qquad (2)$$

π_i: Parents of xi(node) $q_i = |\varphi_i|$ φ_i : List of all possible representation of the parents of x_i in Database. i.e, if $p_1, ..., p_s$ are the parents of x_i then φ_i is the Cartesian product $\{v_1^{p_1}, ..., v_r^{p_i}\}...\{v_1^{p_1}, ..., v_r^{p_s}\}$ of all possible values of attribute p_1 through p_s. $r_i = |v_i|$ v_i: All possible values of the attribute x_i α_{ijk}: Number of items in D in which the trait x_i is represented with its k^{th} value, and the parents of x_i in π_i are instantiated with the j^th instantiation in φ_i. $f(i,\pi_i)$; Database D probability that parents of attribute x_i are π_i.

7) Maximum degree transaction (Tmax): It is a transaction in data set which contains highest numbers of item in it. Maximum degree transaction is utilized for perturbation process.

8) Sensitive rule (RS)/ item(IS): A rule is sensitive (RS) when it discloses any private information. A sensitive rule (RS) is determined by scores obtained after k2. In this model, a sensitive node is identified through K2 algorithm on basis of Bayesian Network. A node is recognized as sensitive by taking its frequent occurrence or maximum likelihood calculation into account throughout this network structure. So the item having maximum column by column dependency after generating number of scores from their random orders, is used to perturb the transactional dataset.

9) Algorithm working and implementation: Before explanation of proposed model here some main steps of models are described that are to be followed:

STEP-1: Initially, text file contain quantitative data is read.
STEP-2: This quantitative data is then converted into fuzzified form by applying Gaussian membership function (gmf(x)).
STEP 3: Binary data is obtained from fuzzified data generated in STEP-2.
STEP-4: Transactional data (after manipulation of binary data with symbolic representation)is used as input to Apriori algorithm for generating fuzzy association rules.
STEP-5: Binary table from STEP 3 is given as input to K2 algorithm [34] for generating Bayesian Network of different random orders.
STEP-6: After K2 algorithm, scores on basis of conditional probability of items on each other is calculated and store in table from STEP-5.
STEP-7: An item with its symbol in transactional table having maximum score after different order based probability and dependency is identified.
STEP-8: Perturb the transactional table using STEP-7.
STEP-9: Transactional data altered in STEP-8 is used by Apriori algorithm.
STEP-10: Generate output of result based on STEP-9 (Hidden Informative FAR).

Proposed algorithm with each step is shown in Fig. 2. This algorithm initially takes original data as input for fuzzification process/Gaussian membership function being most productive in this domain. Gmf(x) is used to convert original data D into fuzzy data using various fuzzy sets. A membership function converts the crisp data into fuzzified form thus eliminating the factor of uncertainty and sharp boundary problem in quantitative or binary data. Furthermore fuzzified data is converted into binary table after applying filtration process of threshold of 0.1% to all fuzzy values and as this threshold is enough to prune negligible fuzzy values. This binary table is given as input to K2 algorithm for making Bayesian network and finding out probability computation of every variable. On the basis of this computation a node is identified as sensitive node. A node is a basically attribute/variable of binary table. A symbolized transactional data, created from binary table, is given input to apriori algorithm to produce efficient fuzzy association rules. Minimum perturbation is performed using maximum degree transaction in transactional table to remove sensitive item and pass the modified data set to generate non sensitive association rules.

Major factor of our proposed algorithm is firstly, highlighting the privacy in fuzzy association rule mining unlike previous researches and secondly, restriction in change of either support confidence threshold. Different dataset samples from literatures are used to test effectiveness and accuracy of our technique.

```
1. Input: Data.txt file, Support Threshold.
2. Output: Non sensitive Fuzzy Association Rules
3. Data < ---- readtxt(Orig_Data.txt)
4. while(sizeof_Trans>=j)
5. i=1;
6. while( sizeOf Fuzz_Col>=i)
7. while(Data_matrixSize(1)>=itrOfData)
8. Fuzz_data <----- Guass MF( 0;1)
9. Fuzz_data = Current_value
10. end
11. Bin_Data[j,i]
    ←(Fuzz_data[Bin_Data(j,i)>0.001]
12. Tran_Doc.txt←Bin_Dta[row, col]
13. Trans_Data←readdata(Tran_Doc.txt)
14. Num_of_Trans← length (Trans_Doc.tra)
15. for k = 1 to num_of_Trans
16. All nodes← [all Nodes,Trans_Data.tra,Node]
17. end
18. All_Nodes=unique(all_Nodes)
19. Node_Symbols←char(64+(1:length(all_Nodes)
20. for k =1 to length(all_Nodes)
21. nodeLookUpTable{k,1}←all_Nodes{k}
22. nodeLookUpTable{k,1}← node_Symbols(k)
23. end
24. BinaryTable←zeros(num_of_Trans.length(all_
    nodes)
25. List_Of_Trans←cell(num_of_Trans,1)
26. for k =1to num_of_Trans
27. Cur_Trans_Nodes←Trans_Data.tra(k).Node
28. for m=1 to length(all_Nodes)
29. nodeIndex←strmatch(all_Nodes{m}.cur_Tra_
    Nodes)
30. if~(isempty)(nodeindex))
31. List_Of_Trans{k}←[List_Of_Trans{k},nodeLo
    okUpTable{m,2}]
32. BinaryTable{k,m}←1
33. end
34. end
35. end
36. Bay_Network_Nodes<call K2{Bin_Data}
37. BayesScore←[K2Score,Order']
38. maxScore← max(abs(BayesScore(:,1)))
39. id# ←find(abs(BayesScore(:,1))>=maxScore)
40. Score_Item_No←BayesScore(id,2)
41. Score_Item_Symbol←
    itemLookUpTable{Score_Item_No,2}
42. Sensitive_Node←Score_Item_Symbol
43. Tran_And_Len←[];trans_Counter←0
44. for k = 1 to length (List_Of_Trans)
45. if
    any(List_Of_Trans{k}←Score_Item_Symbol)
46. Trans_Counter←trans_Counter+1
47. Tran_And_Len(trans_Counter,1)←k
48. Tra_And_Len(trans-
    Counter,2)←length(List_Of_Trans){k})
49. end
50. end
51. Max_length← max(tran_And_Len(:,2))
52. Tran_Delete←tran_And_Len(find(tran_And_L
    en(:,2)=Max_length),1)
53. for m = 1 to length(Tran_Delete)
54. Trans←List_Of_Trans){Tran_Delete(m)}
55. Trans(trans=Symbol_Item_Symbol)←[]
56. List_Of_Trans{Tran_Delete(m)}←Trans
57. end
58. call Apriori(List_Of_Trans) //use apriori
    algorithm to hide Sensitive FAR on transformed
    List_Of_Trans.
```

Fig. 2. Proposed algorithm with detailed steps.

IV. Experimental Result

To assess the working of proposed model is tested on different sample data sets based on other literatures to hide sensitive fuzzy association rules with backing of K2 and apriori algorithms. Our proposed methodology, initially reads a sample dataset containing five attributes of quantitative data in .txt format to apply fuzzification process. Resulting fuzzified data (example in Table I) is converted into binary table representing the presence or absence of particular item. Table II is showing (generic example of) binary data obtained from fuzzified data. Transactional items set Fuzzy association rules are produced from apriori algorithm. A sensitive node (i.e. "B", 2nd node) is identified using scores of different orders obtained from K2 algorithm. Thus FARs containing item "B" are considered as sensitive item and are presented in Table III. Minimum perturbation, based on sensitive item "B" is applied on maximum degree transaction i.e. 3rd row of transactional item set. Fig. 3 and Fig. 4 show the item dependency graph on each other for given sample dataset. Keeping the support constant (i.e. min_support is 20%) Apriori algorithm is again applied on perturbed transactional data to generate non sensitive association rules. As a result, 6 out of 7 rules from sensitive fuzzy association rules based on identified sensitive item, are hidden that are mentioned in Table IV.

In our proposed model, sensitive fuzzy association rules (FARs) are considered according to the identified sensitive item (Is), so these rules are not required to be shared in combined platform of business or publicly. Such rules need to be hidden which is successfully done by our proposed methodology. In Table V, summary of our proposed model on sample data set D is presented. We used HEART data set from UCI repository to evaluate our proposed methodology. This data set contains seven attributes out of which four attributes naming, "Age", "Cholesterol Level", "Blood pressure" and "Maximum

TABLE III. Sensitive Fuzzy Association Rules (FARs) using Sensitive Item

Rule No.	Fuzzy Association Rules	Support%	Confidence%	Lift Ratio%
1	B → D	25	66.6	16
2	B → H	25	66.6	13
3	B → I	25	66.66	22.5
4	B, D → H	25	100	20
5	B, D → O	25	100	33.3
6	B, H → O	25	100	33.33
7	B, D, H → O	25	100	33.33

TABLE IV. Hidden Sensitive Fuzzy Association Rules (FARs) using sensitive item

Rule No.	Fuzzy Association Rules	Support%	Confidence%	Lift Ratio%
1	B → D	25	66.6	16
2	B → H	25	66.6	13
3	B → I	25	66.66	22.5
4	B, D → H	25	100	20
5	B, H → O	25	100	33.33
6	B, D, H → O	25	100	33.33

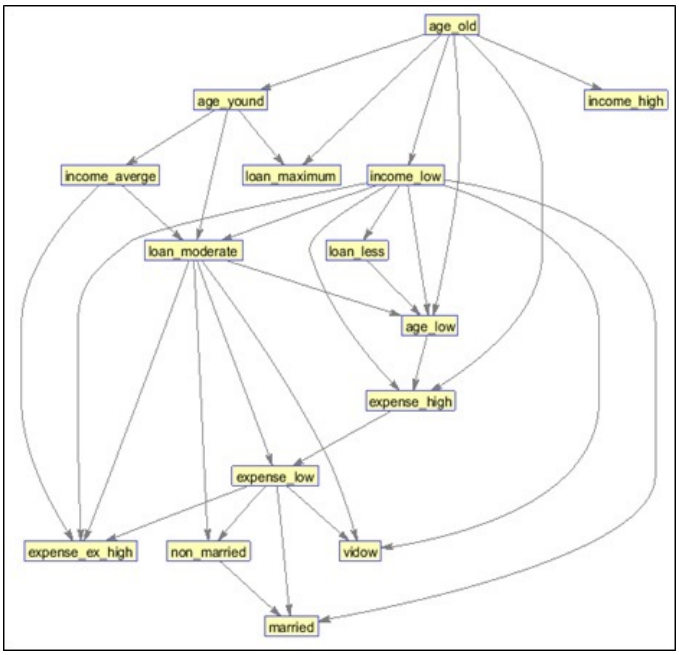

Fig. 3. Bayesian Network on given dataset (D)

Heart Rate" are taken into consideration for testing purpose. Our methodology performed the fuzzification operation to convert those attributes into fuzzified values. Binary table is generated to locate the item's presence and absence in particular tuple. Binary table of heart data set is presented in Table VI. This resulting data is passed to K2 algorithm to generate Bayesian network and computing the dependency scores of each fuzzified attribute (as shown in Fig. 5). This score is calculated using multiple order approach of nodes to find out constant maximum score of particular item and hence considered as sensitive node. A sample of identified sensitive item based FARs are shown in Table VII. Apriori algorithm is used on transactional item sets generated from binary table, to create non redundant fuzzified association rules. Minimum

TABLE V. SUMMARY OF PROPOSED MODEL ON SAMPLE DATASET

Total FARs	Sensitive FARs	Hidden FARS	Lost FARs	Ghost FARs	Modified Transactions	Total Trans-actions
86	07	06	00	00	01	07

perturbation is performed on maximum degree transaction in transactional table and in this case two transactions are perturbed as they both contains maximum number of items in it. Modified table is thus passed to Apriori algorithm to generate non sensitive Fuzzy association rules.

TABLE VI. A SAMPLE OF BINARY TABLE ON HEART DATA SET

ID#	A	B	C	D	E	F	G	H	I	J	K	L
1	0	0	1	1	1	0	0	1	0	1	0	0
2	0	0	1	1	0	0	0	0	1	0	1	1
3	0	1	1	1	1	0	0	1	0	1	1	0
4	0	1	1	0	1	0	1	1	0	0	0	0

TABLE VII. SENSITIVE FUZZY ARS USING SENSITIVE ITEM

Rule No.	Fuzzy Association Rules	Support%	Confidence%	Lift Ratio%
29	E→J	20	34.5	1.56
59	B,E→G	24.48	57.14	1.78
61	B,E→L	20.40	47.61	2.26
73	C,E→K	22.44	78.57	2.70
79	D,E→G	20.40	55.55	1.73
80	D,E→H	26.73	100	2.32
87	E,G→H	36.73	100	2.32
88	E,G→K	20.40	55.55	1,91
89	E,G→L	22.44	61.11	2.91
90	E,H→J	20.40	35.71	1.62
91	E,H→K	32.65	57.1	1.97
101	B,C,E→H	24.48	92.3	3.18
102	B,C,E→K	22.44	84.6	2.91
108	B,D,E→K	22.44	78.56	2.70
119	D,E,G→H	20.40	100	2.32
122	E,G,H→K	20.40	55.55	1.91

Pursuing our assessment on different UCI repository data sets proved effectiveness of our proposed methodology and it showed that our model was successful in hiding sensitive rules. Major factor of our model is restriction of ghost rules and lost rules unlike other algorithms like ISR, DSR and ADSSI. No. of lost rules are not found in our technique. In Fig. 5, a comparison analysis of our proposed methodology with different approaches is shown on hiding failure, number of ghost rules, lost rules. Our technique showed tremendous growth with zero hiding failure and ghost rules and minimizing the number of lost rules using least transformation in database unlike other compared approaches.

V. DISCUSSION

Privacy-Preserving Data Mining (PPDM) techniques are one step ahead in extraction of knowledge from large commercial and public data repositories while preventing the disclosure of sensitive information in form of association rules. In this work, we presented a Bayesian network based PPDM strategy for securing the critical fuzzy association rules. We not only summarized the previously researched work of hiding

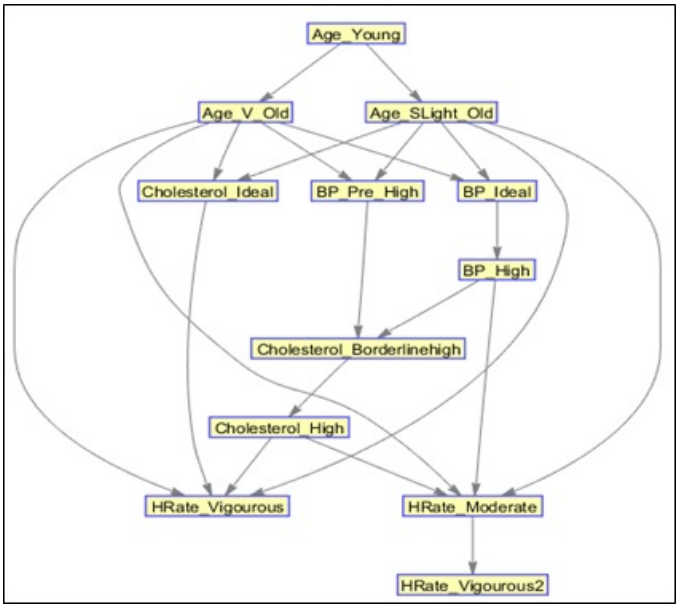

Fig. 4. Bayesian Network by K2 algorithm on HEART Dataset

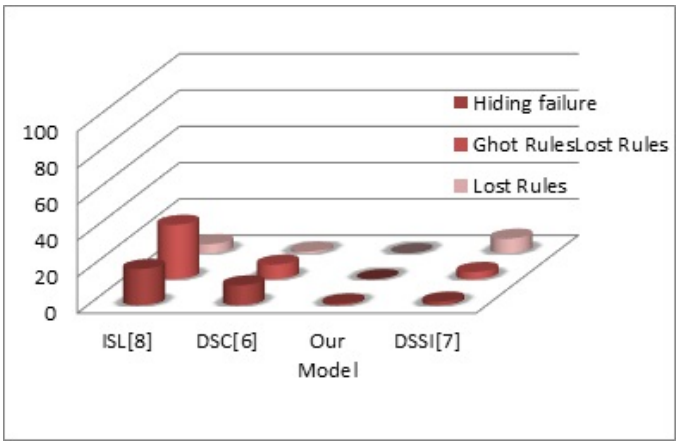

Fig. 5. Comparison of proposed model with few previous approaches

and securing sensitive information procedures like Support confidence, Perturbation techniques but also highlighted the lacking in their findings in context of information securing. After thorough study we came across main factors of previous work; i) Stress on extraction and securing of information from only classical association rule domain; ii) Assumption of sensitive information or target required to be concealed rather than proper formulation. We progressed with prior knowledge of aforementioned techniques to enhance the PPDM securing strategy in fuzzy domain. We mainly addressed privacy preservation of FARs which was not significantly taken into consideration in early contributions. We went through different highly significant phases in our work that includes fuzzification techniques for uncertain data, applying association rule mining technique for items relationships and their correlated behavior. Furthermore, Bayesian network incorporation into decision making procedure for sensitive items identification which later on, lead to creation of non-sensitive association rules after perturbation. The advantage of perturbing in this way, is

to keep the smallest influence to original database with the maximum subsidy with no side effects. Sensitivity analysis is a vital issue in different business and non-business organizations that progress collectively to achieve a given target. This model is proved to be very effective and beneficial in certain research areas which exploit fuzzy association rules like biomedical fields, criminal pattern discovery and financial transactions like supermarket data. When a biologist decides to conceal some sensitive pattern of health samples (may be tissue, patient or tumor), BN based approach can be incorporated to improves the security. In addition, State law enforcement are enduring to call upon data mining techniques to enhance crime analytics and well defend their communities and properties. Real-time solutions can save significant resources and push the capability of law enforcement closer to the pulse of criminal activity. In crime pattern discovery, extracted FARs are more accurate and interesting, and presented and discussed at national level, thus presented model can prove to be effective in dealing security of such high-level sensitive information. Association rules mined from a supermarket database using data mining methods can reveal sensitive information to business contestants. To confine sensitive information, supermarket database should be perturbed for this Bayesian Network-based identified sensitive attributes in a uppermost degree transaction. It is significant to put stress on fuzzy domain considering two facts; firstly, fuzzy logic adaptation to address uncertainty and impreciseness in real-world applications (mentioned above) requiring privacy safeguards and secondly is absence of privacy preservation in the domains. The results of our hiding strategy are quite impressing without any shortcomings and can have valuable research endeavor for future works.

VI. Conclusion

Privacy preservation turns out to be an important aspect in data mining to restrict the disclosure of sensitive information after mining process. In our proposed method, we tend to hide the sensitive fuzzy association rules with minimum or no side effects. Main goal is to uncover those sensitive items which participate in exposing secrets of healthy business to its business contestants during fuzzy association rule generation process. Privacy preservation of maximum number of sensitive FARs by keeping minimum perturbation highlights the significance of our model. Different pervasive tests over case studies and datasets will be held to validate the effectiveness and accuracy of proposed work and hiding sensitive association rules after comparison with current techniques. Support and confidence are extensively taken into account for privacy preservation by many researchers. Our model prospered in performing privacy preservation of association rules without alteration of these measures. The ability to produce the intended results of proposed model are demonstrated in the form of minimizing sensitive fuzzy association rules over different datasets with minimum perturbation to original dataset. In future, probability based multiple sensitive items can be identified to hide sensitive FARs with an aim to keep the minimum perturbation to original data and maximum hidden rules. In addition, machine learning algorithms can be used to discover the relevant and interesting candidate items to increase the effect of securing sensitive FARs.

References

[1] C. Clifton and D. Marks, *Security and Privacy Implications of Data Mining* In ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, 1996, pp. 15–19.

[2] R. Agrawal, T. Imielinski and A. N. Swami, *Mining association rules between sets of items in large databases*, In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, P. Buneman and S. Jajodia, Eds. Washington, D.C., 207-216.

[3] J. Han, J. Pei and Y. Yin, *Mining frequent patterns without candidate generation*, ACM SIGMOD International Conference on Management of Data, 29(2),pp. 1–12, 2000.

[4] T.-P. Hong, C.W. Lin,C.C. Chang and S.-L. Wang, *Hiding Sensitive Itemsets by Inserting Dummy Transactions*, In 2011 IEEE International Conference on Granular Computing, pp. 246-249. IEEE, 2011.

[5] Zadeh L, *Fuzzy sets*, Inf. Control. Vol.8, PP, 338 – 353.

[6] K. Pathak, N. S. Chaudhari and A. Tiwari, *Privacy-Preserving Data Sharing Using Data Reconstruction Based Approach*, SI:International Journal of Computer Applications (0975 – 8887) on Communication Security, No.13, 2012.

[7] C. C. Agarwal and S. Y. Philip, *Privacy Preserving Data Mining: Models and Algorithms*, Springer Science & Business Media., 2008.

[8] W. Xiaodan, C. Chao-Hsien, W. Yunfeng, L. Fengli and Y. Dianmin, *Privacy Preserving Data Mining Research: Current Status and Key Issues*,International Conference on Computational Science,Springer, Berlin, pp. 762-772, 2007.

[9] S. Samet and A. Miri, *Privacy-Preserving Bayesian Network for Horizontally Partitioned Data*, In International Conference on Computational Science and Engineering, Vol. 3, pp. 9–16, 2003.

[10] S.-L. Wang, B. Parikh and A. Jafari, *Hiding informative association rule sets*, In Expert Systemswith Applications 33(2), pp. 316–323, 2007.

[11] P. Kongchai, N. Kerdprasop and K. Kerdprasop, *The Fuzzy Search for Association Rules with Interestingness Measure*, International Journal of Computer Theory and Engineering, Vol.6,No.6,2014.

[12] S. L. Wang, R. Maskey, A. Jafari, and T.P. Hong, *Efficient Sanitization of Informative Association Rules*, Expert Systems with Applications, vol. 37, 2007.

[13] Y. C. Chang and S.T. Chen, *Fast Algorithm for Completely Hiding Sensitive Association Rule Sets*, Proceedings of the Cryptology and Information Security Conference, pp. 547-560, 2008.

[14] S. T. Chen, S.M. Lin, C.Y. Tang, and G.Y. Lin, *An Improved Algorithm for Completely Hiding Sensitive Association Rule Sets*, IEEE 2nd International Conference on Computer Science and its Applications, pp.1-6, 2009.

[15] K. Sathiyapriya, G. Sudhasadasivam and N. Celin, *A New Method for preserving privacy in Quantitative Association Rules Using DSR Approach With Automated Generation of Membership Function*, In the Proceedings of World Congress on Information and Communication Technologies, pp.148-153, 2011.

[16] S. Hameed, F. Shahzad and S. Asghar, *A Fuzzy Correlation Scheme For Privacy Preservation In Knowledge Based Systems*, Australian Journal of Basic and Applied Sciences,6,no.9, pp.562-571, 2012.

[17] T.P. Hong, M.J.Chiang and S.L. Wang, *Mining from quantitative data with linguistic minimum supports and confidences*, Proceedings of the 8th IEEE International Conference on Fuzzy Systems,vol.1, pp. 494-499,IEEE,2002.

[18] T.P. Hong, K.Y. Lin and B.C. Chien, *Mining fuzzy multiple-level association rules from quantitative data*, Applied Intelligence, 18(1), pp. 79–90, 2003.

[19] G.-F. Cooper and E. Herskovits,*A Bayesian method for the induction of probabilistic networks from data Machine Learning*, 9(4), pp. 309–347, 1992.

[20] M.-A. Kadampur and D.-V.-L.-N. Somayajulu, *A Data Perturbation Method by Field Rotation and Binning by Averages Strategy for Privacy Preservation*, In Intelligent Data Engineering and Automated Learning–IDEAL, Springer-Ver lag, pp. 250–257, 2008.

[21] S.-L. Wang, D. Patel and A. Jafari, *Hiding collaborative recommendation association rules*, Applied Intelligence, Vol.27, No.1,pp.67-77, 2007.

[22] K. Shah, A. Thakkar and A. Ganatra, *Association Rule Hiding by Heuristic Approach to Reduce Side Effects & Hide Multiple RHS Items*, International Journal of Computer Applications (0975 – 8887), 2012.

[23] K. Iqbal, X.-C. Yin, H.-W. Hao, Q. M. Ilyas and X. Yin, *A central tendency-based privacy preserving model for sensitive XML association rules using Bayesian networks*, Intelligent Data Analysis 18, no. 2, 281-303, 2014.

[24] G. Chen, P. Yan and E.E. Kerre, *Computationally Efficient Mining for Fuzzy Implication Based Association Rules in Quantitative Databases*, International Journal of General Systems,33, pp.163-182 2004.

[25] C.-C. Weng, N.D.U. Taoyuan, S.-T.Chen, H.-C. Lo, *A Novel Algorithm for Completely Hiding Sensitive Association Rules*, In Eight Intenrnational Conference on Intelligent Systems Design and Applications,pp. 202-208, 2008.

[26] T. P. Hong, C. S. Kuo and S. C. Chi, *Mining association rules from quantitative data*, Intelligent Data Analysis,3(5),pp.363–376, 1999.

[27] S.N. Mandal,J.P.CHoudhury and S.R.B. Chaudhuri *In Search of Suitable Fuzzy Membership Function in Prediction of Time Series Data*, International Journal of Computer Science Issues, Vol.9,Issue 3,pp. 293-302, 2012.

[28] T. Berberoglu and M. Kaya, *Hiding Fuzzy Association Rules in Quantitative Data*, In 3rd International Conference on Grid and Pervasive Computing-Workshops, pp. 387-392,IEEE, 2008.

[29] E. Hüllermeier, *Fuzzy methods in machine learning and data mining: Status and prospects*, Fuzzy Sets and Systems, 156, pp. 387-406, 2005.

[30] M. Gupta and R.C.Joshi, *Privacy Preserving Fuzzy Association Rules Hiding in Quantitative Data*, International Journal of Computer Theory and Engineering, Vol.1,No.4,pp.382-388,2009.

[31] J. Han, J. Pei and Y. Yin, *Mining frequent patterns without candidate generation*, ACM sigmod record, Vol.29, No.2, ACM, 2000.

[32] C.-M. Wu, Y.-F. Huang and J.-Y. Chen, *Privacy Preserving Association Rules by Using Greedy Approach*, WRI World Congress on Computer Science and Information Engineering, 4, pp.61–65,2009.

[33] S. R. M. Oliveira and O. R. Zaïane, *Privacy preserving frequent itemset mining*, In Proceedings of the 2002 IEEE International Conference on Privacy, Security and Data Mining (CRPITS), pp. 43–54, 2002.

[34] S.-R.-M. Oliveira and R.-Z. Osmar, *A Unified Framework for Protecting Sensitive Association Rules in Business Collaboration*, Inderscience Publishers 1(3),247–287, 2006.

[35] X. Sun and P.S. Yu, *Hiding sensitive frequent itemsets by a border-based approach*, Journal of Computing Science and Engineering, Vol.1, No.1, pp. 74-94, 2007.

[36] X. Sun, P.S. Yu, *A border-based approach for hiding sensitive frequent itemsets*, Fifth IEEE International Conference in Data mining, p. 1550-4786, 2005.

[37] P. Cheng, J.-S.Pan, C.-W.L. Harbin, *Use EMO to Protect Sensitive Knowledge in Association Rule Mining by Removing Items*, In IEEE Congress on Evolutionary Computation (CEC),2014.