Absolute Soft Set Approach for Mining Association Patterns

Saakshi Saraf
Department Of Computer Applications,
MANIT, Bhopal (MP) India.

Neeru Adlakha
Department of Applied Mathematics and Humanities,
S.V.N.I.T, Surat, (Guj.) India.

Sanjay Sharma
Department of Computer Applications,
MANIT, Bhopal (MP) India.

ABSTRACT

Molodtsov initiated the concept of soft set as a new mathematical tool for dealing with uncertainties. In 2003, Maji put forward several notions on Soft Set Theory. In this paper, absolute soft set approach has been developed for mining association patterns from a transactional data set. This approach is used for mining associations has been illustrated with the help of an example and experiment on a real world data set. In particular, the work demonstrates that absolute soft set theory can be applied to problems that contain uncertainties especially in decision making problems. The proposed approach gives better picture of association relationship, confidence levels and is helpful in addressing the absolute association patterns.

Keywords: Association Pattern mining, soft set, absolute soft set, Associations rules.

1. INTRODUCTION

Most of our traditional tools for formal modeling, reasoning and computing are crisp, deterministic and precise in character. However, there are many complicated problems in economics, engineering, environment, social science, medical science etc. that involve uncertainties. A number of techniques are reported in the literature [1-8] for data mining. Association rule mining is one of the important types of data mining which has attracted attention of various research workers. A good number of attempts are reported in the literature [1-4] on development of algorithms for association rule mining in deterministic situations. Apriori algorithm [1, 2] is the first algorithm for mining frequent item sets for association rules. This algorithm is based on support of an itemset determined by just counting the number of occurrences of the item set in every record of transaction, without any consideration to the number of items in a record of transaction. All association rule mining algorithms based on this concept suffer from the problem of under prediction or over prediction of association rules. However, based on human intuitive, it should be considered that the larger number of items purchased in a transaction means that the degree of association among the items in the transaction may be lower. This concept was proposed and discussed well in[3, 4]. Real world data in the field of engineering, science, technology, medical and social sciences, economics etc. pose major challenges due to uncertain or imprecise data which are inherent and pervasive these applications[5]. Various types of theories like theory of probability, fuzzy set theory[6] , intuitionistic fuzzy sets [7], vague sets[8], theory of interval mathematics[9], rough set theory[10] etc. have been proposed for dealing with such systems in an effective way. However, all these theories are associated with an inherent limitation, which is the inadequacy of the parameterization tool are associated. Molodtsov [6] initiated a novel concept of soft set theory as a new mathematical tool for dealing with uncertainties. It associates a set with a set of parameters and thus is free from the above limitations. A general mathematical tool − soft set theory [11-22] offer solutions for dealing with uncertain, fuzzy, not clearly defined objects. In soft set theory, membership is decided by adequate parameters, rough set theory employs equivalence classes, whereas fuzzy set theory depends upon grade of membership. Although three theories are quite distinct yet deal with vagueness. To deals uncertainty in association rule mining, a number of attempts are reported for development of fuzzy set [4, 13, 14],rough set[23]and vague set[24] approaches for association rule mining. A wide range of applications of soft sets have been developed in many different fields, including the smoothness of functions, game theory, operations research, Riemann integration, Perron integration, probability theory and measurement theory and its applications. Maji et al.[25] discussed the application of soft set theory in a decision making problem. Herawan et. al.[26] have developed a soft set approach for mining association patterns in transactional datasets but no attempt is reported in the literature on development of Absolute soft set approach for association patterns in transactional data set. In view of above in the present paper an Absolute soft set approach is developed for mining association patterns in a transaction dataset. The approach is demonstrated with the help of suitable examples and experiments in real world dataset.

The paper is organized as follows. Section 2 describes fundamental concept of soft set theory and association rules mining. Section 3 describes absolute soft set approach for mining association pattern .Section 4 shows the results of experiment on real data set of air pollution. Finally, the conclusion of this work.

2. DEFINITION AND PRELIMINARIES

In this section, we recall some basic notion of soft set theory introduced by Molodtsov et al.[27] and some useful definition from Maji et al.[28, 29]. Here, U to be an initial universal set and E to be a set of parameters and $A, B \subset E$ [30].

Definition 2.1 (Soft set)

A pair (F, E) is called a soft set (over U) if and only if F is a mapping of E into the set of all subsets of the set U. In other words, the soft set is a parameterized family of subsets of the set U. Every set F(e), e ∈ E, from this family may be considered as the set of e -approximate elements of the soft set. Let us consider the following example.

Example 2.1.1: A soft set (F, E) describes the attractiveness of the bikes which Mr. X is going to buy [31].

U is the set of bikes under consideration. E is the set of parameters. Each parameter is a word or a sentence.

E = (e1= stylish; e2 =heavy duty; e3= light; e4 =steel body; e5 = cheap; e6 = good mileage; e7 =easily Started; e8 = long driven; e9=costly; e10 = fibre body)

In this case, to define a soft set means to point out stylish bikes, heavy duty bikes, and so on.

Example 2.1.2 : Let U={u1,u2,u3,u4,u5} be a universal set and E={x1,x2,x3,x4} be a set of parameters. If A={x2, x3, x4} and then the soft set F_A is written by $F_A = \{(x2,\{u2,u4\}),(x4,U)\}$

Definition 2.2 (Operation with soft sets)

Suppose a binary operation denoted by *, is defined for all subsets of the set *U*. Let (*F, A*) and (*G, B*) be two soft sets over *U*. Then the operation * for the soft sets is defined in the following way: (*F, A*) * (*G, B*) = (*H, A×B*) Where *H* (α , β) = *F*(α) * *G*(β) , α ∈ *A*, β ∈ *B* and *A×B* is the Cartesian product of the sets *A* and *B* .

Definition 2.3 (NOT set of a set of parameters)

Let E= {$e_1, e_2, e_3, \dots e_n$} be a set of parameters. The NOT set of E denoted by $\neg$E and is defined by $\neg$E= { $\neg e_1$, $\neg e_2$, $\neg e_3 \dots$, $\neg e_n$} where $\neg e_i$ = not e_i for all i . It may be noted that $\neg$ and $\neg$ are two different operations.

Definition 2.4 (Complement of a soft set)

The complement of a soft set (*F, A*) is denoted by (*F, A*)^c and is defined by (*F, A*)^c = (*F^c* , $\neg$*A*) where *F^c* : $\neg A \rightarrow P(U)$ is a mapping which is defined by $F^c(\alpha) = U - F(\neg \alpha)$, for all α ∈ $\neg A$.

Definition 2.5 (Relative complement of a soft set)

The relative complement of a soft set (F, A) is denoted by (F, A)^r and is defined by (F, A)^r = (F^r , A) where F^r : A →P(U) is a mapping given by F^r(α)=U - F(α), for all α ∈ A .

Definition 2.6 (NULL soft set)

A soft set (*F, A*) over *U* is said to be a NULL soft set denoted by Φ, if *for all ε ∈A, F (ε)* = ϕ (null-set).

Definition 2.7 (Relative NULL soft set)

A soft set (*F, A*) over *U* is said to be relative NULL soft set with respect to parameter set *A* denoted by Φ *A if* ε ∈*A, F (ε)* = ϕ (null set).

Definition 2.8 (Relative whole soft set)

A soft set (*F, A*) over *U* is said to be relative whole soft set (with respect to parameter set *A)* denoted by U*A*, if for all ε ∈*A, F (ε)* = *U.*

Definition 2.9 (Absolute soft set)

The relative whole soft set U(E) with respect to the universe set of parameters E is called the absolute soft set over U.

Example 2.9.1:[12] Suppose that, U is the set of wooden houses under consideration is the set of parameters. Let there be five houses in the universe U given by U= {hl, h2, h3, h4, h5} and B = {not brick; not muddy; not steel; not stone}. The soft set (G, B) describes the "construction of the houses". The soft sets (G, B) is defined as
G (not brick) means the houses not built by brick,
G (not muddy) means the not muddy houses,
G (not steel) means the houses not built by steel,

G (not stone) means the houses not built by stone.
The soft set (G, B) is the collection of approximations as below:
(G, B) = {not brick built houses = {hl, h2, h3, h4, h5},
not muddy houses = {hl, h2, h3, h4,h5},
not steel built houses = {hl, h2,h3,h4, h5},
not stone built house = {h1,h2,h3,h4,h5} }
The soft set (G, B) is the absolute soft set.

Definition 2.10 (AND operation on two soft sets)

If (*F, A*) and (*G, B*) be two soft sets then (*F, A*) AND (*G,B*) denoted by (*F, A*) ∧ (*G,B*) and is defined by (*F, A*) ∧ (*G,B*) = (*H, A× B*) where H(α, β) = F(α) ∩ G(β) for all (α, β)∈ A× B .

Definition 2.11 (OR operation on two soft sets)

If (F, A) and (G,B) be two soft sets then (F, A) OR (G,B) denoted by (F, A) ∨ (G,B) is defined by(F,A) ∨ (G,B)= (O,A×B) where O(α, β) = F(α) ∪ G(β) for all (α, β) ∈ A× B .

3. ASSOCIATION RULES [26]

Let I = {i1, i2, . . ., i|A|}, for |A| > 0 refers to the set of literals called set of items and the set D = {t1, t2, . . ., t|U|}, for |U| > 0 refers to the transactional dataset, where each transaction t ∈ D is a list of distinct items t = {i1, i2, . . ., i|M|}, $1 \leq M \leq A$ and each transaction can be identified by a distinct identifier TID. Let, a set X ⊆ t ⊆ I called an itemset. An itemset with k-items is called a k-itemset. The support of an itemset X, denote by sup (X) is defined as a number of transactions containing X. An association rule between sets X and Y is an implication of the form X ⟹ Y, where X ∩ Y = ∅ . The itemsets X and Y are called antecedent and consequent, respectively. The support of an association rule X ⟹ Y, denoted by sup (X ⟹ Y), is defined as a number of transactions in D containing X ∪ Y. The confidence of an association rule X ⟹ Y, denoted by cfi (X ⟹ Y) is defined as a ratio of the number of transactions in D containing X ∪ Y to the number of transactions in D containing X. Thus,
$$\textbf{Cfi (X} \Rightarrow \textbf{Y)} = \frac{\sup(X \cup Y)}{\sup(X)}.$$

A huge number of association rules can be found from a transactional dataset. To find the interesting association rules in a transactional dataset, define a specified minimum support (called minsup) and specified minimum confidence (called minconf). The itemset Y ⊆ I is called frequent itemset if sup (X) ≥ minsup. It is known that a subset of any frequent itemset is a frequent itemset, a superset of any infrequent itemset is not a frequent itemset. Finally, the association rule X ⟹ Y holds if conf (X ⟹ Y) ≥ minconf. The association rules are said to be strong if it meets the minimum confidence threshold. However, while association rules provide means to discover many interesting associations, they fail to discover other, no less interesting associations, which are also hidden in the data.

4. ABSOLUTE SOFT SET APPROACH FOR ASSOCIATION RULE MINING:

An Illustrative Example 4.1 is given to understand well the concept of the soft set association rule mining[26, 32] which is

performed step by step. The process is started from a given transactional database as shown in Table 1

Suppose that δ arbitrarily equals to 3; that means qualified transaction is regarded as a transaction with no more than 3 items purchased in the transaction. There is a data set consisting of 16 transactions which contains two categories i.e., T= {Countries, Items},

Where Countries = {India, USA, Iran} and Items= {I1, I2, I3}. Absolute soft set of {I1} is Not I2, Not I3 items purchased by countries, Only I1 Item is purchased.

Result of the step is a set of qualified transaction where M ={ $T_1, T_2, T_3, T_4, T_5, T_6, T_7, T_8, T_9, T_{10}, T_{11}, T_{12}, T_{13}, T_{14}, T_{15}, T_{16}$ }

Table: 1 A Qualified Data Transaction (M)

Trans_ID	List of topics with countries	Trans_ID	List of topics with countries
T_1	{India,I1}	T_9	{India,Iran,I1,I3}
T_2	{USA,I2}	T_{10}	{USA,I2}
T_3	{India,I1}	T_{11}	{India,I1,I3}
T_4	{USA,I2}	T_{12}	{USA,I2,I3}
T_5	{India,I1}	T_{13}	{India,I1,I2,I3}
T_6	{Iran,I3}	T_{14}	{USA,I2,I3}
T_7	{USA,I2}	T_{15}	{Iran,I1,I3}
T_8	{India,USA,I2,I3}	T_{16}	{Iran,I3}

Employing soft set theory and Apriori algorithm on transactions in Table 1 calculate candidate itemsets and frequent itemsets. Finally, calculate support and confidence of frequent itemsets. The itemsets and their supports are given below:

1 Itemsets

$\{India\} = \{NotUSA, NotIran\} = \{T_1, T_3, T_5, T_{11}, T_{13}\} = 5$

$\{USA\} = \{NotIndia, NotIran\} = \{T_2, T_4, T_7, T_{10}, T_{12}, T_{14}\} = 6$

$\{Iran\} = \{NotIndia, NotUSA\} = \{T_6, T_{15}, T_{16}\} = 3$

$\{I1\} = \{NotI2, NotI3\} = \{T_1, T_3, T_5\} = 3$

$\{I2\} = \{NotI1, NotI3\} = \{T_2, T_4, T_7, T_{10}\} = 4$

$\{I3\} = \{Noti1, NotI2\} = \{T_6, T_{16}\} = 2$

Minimum Support is $\beta_1 = 2$.

2 Itemsets

$\{India, I1\} = \{NotUSA, NotIran, NotI2, NotI3\} = \{T_1, T_2, T_5\} = 3$

$\{India, I2\} = \{NotUSA, NotIran, NotI1, NotI3\} = \{Nil\} = 0$

$\{India, I3\} = \{NotUSA, NotIran, NotI1, NotI2\} = \{Nil\} = 0$

$\{USA, I1\} = \{NotIndia, NotIran, NotI2, NotI3\} = \{Nil\} = 0$

$\{USA, I2\} = \{NotIndia, NotIran, NotI1, NotI3\} = \{T_2, T_4, T_7, T_{10}\} = 4$

$\{USA, I3\} = \{NotIndia, NotIran, NotI1, NotI2\} = \{Nil\} = 0$

$\{Iran, I1\} = \{NotIndia, NotUSA, NotI2, NotI3\} = \{Nil\} = 0$

$\{Iran, I2\} = \{NotIndia, NotUSA, NotI1, NotI3\} = \{Nil\} = 0$

$\{Iran, I3\} = \{NotIndia, NotUSA, NotI1, NotI2\} = \{T_6, T_{16}\} = 2$

{India,I2},{India,I3},{USA,I1),{USA,I3},{Iran,I1},{Iran,I2} cannot be considered for further process because their support is less than $\beta_2 = 2$.

Only those itemset are considered whose support is greater or equal to minimum support. The support and confidence of frequent itemsets are given in Tables 2,3 respectively.

Table 2:L1 ($\beta_1 = 2$.)

Min_Sup =2	
1-itemsets	**Support**
{India}	31%
{USA}	37.5%
{Iran}	18.75%
{I1}	18.75%
{I2}	25%
{I3}	12.5%

$Cfi\ (X \Rightarrow Y) = \dfrac{\sup(X \cup Y)}{\sup(X)}$.

Table 3:L2 ($\beta_2 = 2$)

Min_Sup=2		
2-itemsets	**Support**	**Confidence**
{India,I1}	18.75%	60%
{USA,I2}	25%	66%
{Iran,I3}	12.5%	66%

5. EXPERIMENT ON AIR POLLUTION DATASET [33]

We further explain the concept and process of absolute soft set approach for mining association patterns by performing experiment on real data set of air pollution from Maharashtra pollution control board[33] of India.The data of air pollution of Pune city of India for the period of two years during 1January, 2010 to 31 December 2011 is used to perform the experiment[33].It is based on the observation of the air pollution data of two areas of Pune city,namely Chinchwad and Karve Road, that have 570 and 692 transactions respectively.It contains the data on concentration in Micrograms per cubic meter($\mu g/m^3$) of three pollutants SO_2,NO_x and RSPM in the atmosphere of above two areas of Pune.The standards of acceptable limit of concentration of the above three pollutants is also given in the dataset available on websites of Maharashtra Pollution Control Board.The data is presented as the average amount of each data item per day.

Using the given acceptable limit, the taxonomy of data is prepared as given below:
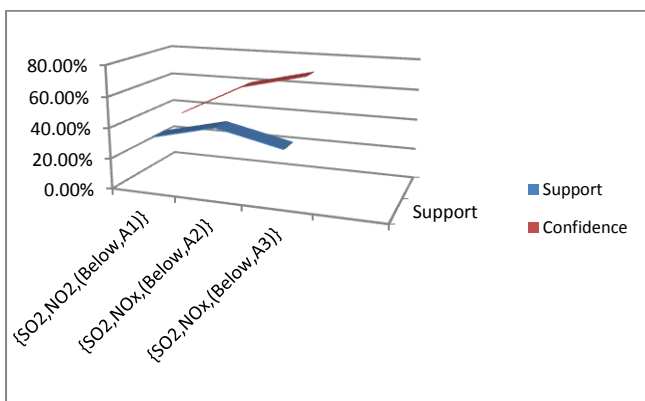
T={below acceptable limit,above acceptable limit,Area}

Area={A1,A2,A3}

A1= Chinchwad, A2=Karve, A3=All(Chinchwad + Karve)

Association patterns for Air Pollution Data along with their Support and confidence .Taking threshold value of Soft set as 30% , the frequent association patterns as shown in Figure 1.
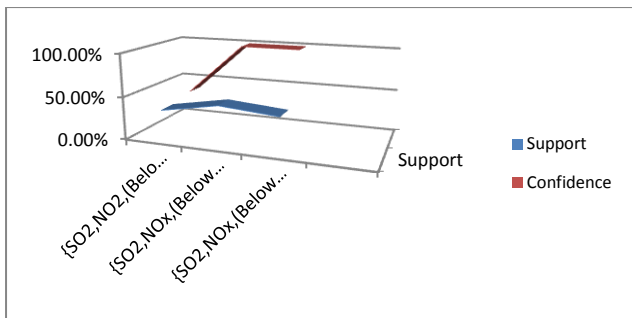
Figure 1: Absolute Soft Set frequent association patterns.

Figure 2: Traditional approach of Association patterns for Air Pollution Data along with their support and confidence .Taking threshold value of traditional approach(apriori) as 30% and get the frequent association patterns as shown in Figure 2.

In this example, calculate support and confidence level of association patterns discovered by the two approaches .In Figure 1 shows the frequent association pattern(absolute soft set) for support threshold 30%.The patterns {So2,Nox,(Below,A1)}, {So2,Nox,(Below,A2}, {So2,Nox,(Below,A3)} is found to be frequent by absolute soft set approach. Here observed that the absolute soft set approach is showing better result and quite useful in addressing the issues of under prediction and over prediction of Association pattern a part of this a soft set approach gives better picture of confidence level as compared to that in traditional approach which is visible in Figure 1 and 2 as significant difference in confidence level of both approaches.

6. CONCLUSION:
In this paper, we have proposed absolute soft set approach of association rules mining from transactional dataset which has been successfully demonstrated using suitable examples. This process is able to incorporate the degree of association among the items or elements of absolute soft set based transaction. The algorithm is based on the concept that the larger number of items purchased in a transaction means the lower degree of association among the items in the transaction. The proposed approach shows a better picture of association relationship and confidence levels as compared to traditional approach.

7. REFERENCES
[1] R. Agrawal , T. Imielinski, and A. Swami, "Mining Association Rules Between Sets of Items in Large Databases," in: proceedings of the 1993 acm sigmod international conference on management of data, washington dc (usa), 1993, pp. 207-216.

[2] R. Agrawal, and R. Srikant, "Fast Algorithms for Mining Association Rules," Proceedings of the 20th VLDB Conference Santiago.

[3] R. Intan, and M. Mukaidono, "Toward a Fuzzy Thesaurus Based on Similarity in Fuzzy Covering," Australian Journal of Intelligent Information Processing, vol. 8, no. 3, 2004, pp. 132-139.

[4] R. Intan, and O. Yenty, "Mining Multidimensional Fuzzy Association Rules from a Normalized Database," Book Mining Multidimensional Fuzzy Association Rules from a Normalized Database, Series Mining Multidimensional Fuzzy Association Rules from a Normalized Database, ed., Editor ed.^eds., 2008, pp. 425-432.

[5] R. Borgohain, and P.K. Das, "Application of Fuzzy Soft set in medical diagnosis using Fuzzy arithmetic operations on Fuzzy Number," SIBCOLTEJO, vol. 05, 2010, pp. 107-116.

[6] L.A. Zadeh, "Fuzzy set," Information and Control, vol. 8, 1965, pp. 338-353.

[7] K.T. Atanassov, "Intuitionistic fuzzy sets," Fuzzy Sets and Systems, vol. 20, 1986, pp. 87-96.

[8] W. Gau, and D. Buehrer, "Vague sets," IEEE Trans System Man Cybernet, vol. 23, no. 2, 1993, pp. 610–614.

[9] M.B. Gorzalzany, "A method of inference in approximate reasoning based on interval-valued fuzzy sets," Fuzzy Sets and Systems, vol. 21, 1987, pp. 1-17.

[10] Z. Pawlak, "Rough Sets," International. Journal of Computer and Information Sciences, vol. 11, 1982, pp. 341–356.

[11] D. Molodtsov, "Soft set theory-first results," Computers and Mathematics with Applications, vol. 37, 1999, pp. 19-31.

[12] P.K. Maji, R. Biswas, and A.R. Roy, "Soft Set Theory," Computer and Mathematics with Application, vol. 45, 2003, pp. 555-562.

[13] M. Ali, F. Feng, X. Liu, W. Min, and M. Shabir, "On some new operations in soft set theory," Computers and Mathematics with Applications, vol. 57, 2009, pp. 1547-1553.

[14] K. Babitha, and J. Sunil, "Soft set relations and functions," Computer and mathematics with Applications, vol. 60, no. 7, 2010, pp. 1840-1849.

[15] N. Cagman, and S. Enginoglu, "Soft matrix theory and its decision making," Computers and Mathematics with Applications, vol. 59, 2010, pp. 3308-3314.

[16] N. Cagman, and S. Enginoglu, "Soft set theory and uni-int decision making," European Journal of Operational Research, vol. 207, 2010, pp. 848-855.

[17] D. Chen, E.C.C. Tsang, D.S. Yeung, and X. Wang, "The parameterization reduction of soft sets and its applications," Computers and Mathematics with Applications, vol. 49, 2005, pp. 757-763.

[18] X. Ge, and S. Yang, "Investigations on some operations of soft sets," World Academy of Science, Engineering and Technology, vol. 51, 2011, pp. 1112-1115.

[19] Y.B. Jun, C. Hwan, and Park, "Applications of soft sets in ideal theory of BCK/BCI-algebras," Information Sciences, vol. 178, 2008, pp. 2466-2475.

[20] Z. Kong, L. Gao, L. Wang, and S. Li, "The normal parameter reduction of soft sets and its algorithm," Computers and Mathematics with Applications, vol. 56, 2008, pp. 3029-3037.

[21] T.Y. Lin, "A set theory for soft computing a unified view of fuzzy sets via neighborhoods," in Proceedings of the IEEE International Conference on Fuzzy Systems, pp. 1140-1146.

[22] D. Pei, and D. Miao, "From soft sets to Information systems," in Proceedings of the IEEE International Conference on Granular Computing, pp. 617-621.

[23] J.W. Guan, D.A. Bell, and D.Y. Liu, "The rough set approach to association rule mining," in :The proceedings of the Third IEEE International Conference on Data mining(ICDM'03), pp. 529-532.

[24] A. Lu, and W. Ng, "Mining Hesitation Information by Vague Association Rules," Conceptual Modeling - ER 2007,Lecture Notes in Computer Science, vol. 4801, 2007, pp. 39-55.

[25] P.K. Maji, A.R. Roy, and R. Biswas, "An application of soft sets in a Decision making problem," Computers and Mathematics with Applications, vol. 44, 2002, pp. 1077–1083.

[26] T. Herawan, and M.M. Deris, "A Soft Set approach for association rule mining," Knowledge Based Systems, vol. 24, 2011, pp. 186-195.

[27] D. Molodtsov, "Soft set theory-first results," Computers and Mathematics with Applications,vol. 37, 1999, pp. 19–31.

[28] P.K. Maji, A.R. Roy, and R. Biswas, "An application of soft sets in a Decision making problem," Computers and Mathematics with Applications, vol. 44, 2002, pp. 1077–1083.

[29] P.K. Maji, R. Biswas, and A.R. Roy, "Soft Set Theory," Computer and Mathematics with Application, vol. 45 2003, pp. 555-562.

[30] K.T. Atanassov, "Operators over interval valued intuitionistic fuzzy sets," Fuzzy Sets and Systems, vol. 64, 1994, pp. 159-174.

[31] M. Pal, and S. Mondal, "Soft Matrices," African journal of Mathematics and Computer Science Research, vol. 4, no. 13, 2011, pp. 379-388.

[32] T. Herawan, I.T.R. Yanto, and M.M. Deris, "SMARViz:Soft Maximal Association Rules Visualization," In: IVIC, Springer, vol. 5857, 2009, pp. 664-674.

[33] www.mpcb.gov.in/envtdata/demoPage1.php.