International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-2 Issue-3, February 2013

Domain Driven Multi-Feature Combined Mining for Retail Dataset

Arti Deshpande, Anjali Mahajan

Abstract*— Association Mining is used to generate the patterns from static data available. But from the business perspective, usefulness and understandability of those rules are more important. Through classical association mining many redundant rules are generated which may be not useful for business analysis. The proposed framework helps in generating the combined rules which gives informative knowledge for business by combining static and transactional data. This paper gives pruning method to remove the redundant rules before generating the combined rules. Finally Rule Clusters are generated for similar group customer or similar transaction characteristics which provide more interesting knowledge and actionable result than traditional association rule. Experimental result demonstrate the proposed techniques.

***Index Terms*— Domain Driven Data Mining, Combined Patterns, Association Rule, Pruning**

I. INTRODUCTION

Impact-targeted activities are rare but they may have a significant impact on the society. For example, isolated terrorism activities may lead to a disastrous event, threatening the national security. Similar issues can also be seen in many other areas. Therefore, it is important to identify such particular activities before they lead to having a significant impact to the world. However, it is challenging to mine impact-targeted activity patterns due to their imbalanced structure.

Longbing Cao et.al [1] developed a technique for discovering such activity patterns. First, the complexities of mining imbalanced impact-targeted activities are analyzed and then strategies for constructing impact-targeted activity sequences are given. Algorithms are developed to mine frequent positive-impact-oriented (P -> T) and negative-impact-oriented (P -> T') activity patterns, sequential impact-contrasted activity patterns (P is frequently associated with both patterns P ->T and P -> T' in separated data sets), and sequential impact-reversed activity patterns (both P -> T and PQ -> T' are frequent). Activity impact modeling is also studied to quantify the pattern impact on business outcomes.Social security debt-related activity data is used to test the proposed approaches. The outcomes show that they are promising for information and security informatics (ISI) applications to identify impact-targeted activity patterns in imbalanced data.

The DMCA (Microarray Data Classification Accuracy) [2] technique is proposed in which the main objective is reducing the number of genes needed for accurate classification. The proposed technique is a combination of two feature selection techniques, f-score and entropy-based, and a powerful classifier, Support Vector Machines. DMCA achieved promising results and is characterized by being flexible in all of its stages. When applied to two public microarray datasets, DMCA succeeded in reducing the number of gene expression values needed to classify a sample by 71.29% and guaranteed reliable classification accuracy.

A formal view of actionable knowledge discovery (AKD) from the system and decision-making perspectives is presented in [3]. AKD is a closed optimization problem-solving process from problem definition, framework/model design to actionable pattern discovery, and is designed to deliver operable business rules that can be seamlessly associated or integrated with business processes and systems. To support such processes, we correspondingly propose, formalize, and illustrate Multisource Combined-Mining-based AKD (MSCM-AKD). A real-life case study of MSCM- based AKD is demonstrated to extract debt prevention patterns from social security data. Substantial experiments show that the proposed frameworks are sufficiently general, flexible, and practical to tackle many complex problems and applications by extracting actionable deliverables for instant decision-making. The method in [4] investigates the characteristics and challenges of activity data, and the methodologies and task of activity mining based on case study experience in the area of social security. Activity mining aims to discover high impact activity patterns in huge volumes of unbalanced activity transactions. Activity patterns identified can be used to prevent disastrous events or improve business decision making and processes. The above issues and prospects were illustrated in mining governmental customer contacts data to recover customer debt. Longbing Cao [5] proposes combined mining as a general approach to mining for informative patterns combining components from either multiple data sets or multiple features or by multiple methods on demand. It summarizes general frameworks, paradigms, and basic processes for multifeature combined mining, multisource combined mining, and multimethod combined mining. Novel types of combined patterns, such as incremental cluster patterns, can result from such frameworks, which cannot be directly produced by the existing methods. A set of real-world case studies has been conducted to test the frameworks. They identify combined patterns for informing government debt prevention and improving government service objectives, which show the flexibility and instantiation capability of combined mining in discovering informative knowledge in complex data The paper [6] shows the flexibility and instantiation capability of combined mining in discovering more informative and actionable patterns in complex data.

Manuscript published on 30 February 2013.
***** Correspondence Author (s)
Coimbatore, India.
Arti Deshpande*, Research Scholar , Department of Computer Engineering, G. H. Raisoni College of Engineering, Nagpur, India.
Dr. Anjali Mahajan, Professor and Head of Department Computer Science & Engineering , Priyadarshini Institute of Engineering and Technology, Nagpur , India.

It also presents combined patterns in dynamic charts, a novel pattern presentation method reflecting the evolution and impact change of a cluster of combined patterns and supporting business to take actions on the deliverables for intervention.

The technique [7] has designed a novel notion of combined patterns to extract useful and actionable knowledge from a large amount of learned rules. It also presents definitions of combined patterns, design novel metrics to measure their interestingness and analyze the redundancy in combined patterns. Experimental results on real-life social security data demonstrate the effectiveness and potential of the proposed approach in extracting actionable knowledge from complex data.

A method [8] to analyse links between binary attributes in a large sparse data set is proposed. Initially the variables are clustered to obtain homogeneous clusters of attributes. Association rules are then mined in each cluster. A graphical comparison of some rule relevancy indexes is presented. It is used to extract best rules depending on the application concerned. The proposed methodology is illustrated by an industrial application from the automotive industry with more than 80 000 vehicles each described by more than 3000 rare attributes.

The methodology [9] of Domain Driven Data aims to construct next-generation methodologies, techniques and tools for a possible paradigm shift from data-centered hidden pattern mining to domain-driven actionable knowledge delivery.

Reference [10] have done a project on improving income reporting to discover the patterns of those customers who were or are in debt to Centrelink. Two data models were built respectively for demographic data and activity data, and decision tree and sequence mining were used respectively to discover demographic patterns and activity sequence patterns of debtors. The project produced some potentially interesting results, and paved the way for more data mining applications in Centrelink in near future.

Xiaoxin Yin[11] proposes a new approach, called CrossMine, which includes a set of novel and powerful methods for multirelational classification, including 1) tuple ID propagation, an efficient and flexible method for virtually joining relations, which enables convenient search among different relations, 2) new definitions for predicates and decision-tree nodes, which involve aggregated information to provide essential statistics for classification, and 3) a selective sampling method for improving scalability with regard to the number of tuples. Based on these techniques, two scalable and accurate methods for multirelational classification: CrossMine-Rule, a rule-based method and CrossMine-Tree, a decision-tree-based method are propsed. Comprehensive experiments on both real and synthetic data sets demonstrate the high scalability and accuracy of the CrossMine approach.

Dempster-Shafer's theory [12] of evidence combination for mining medical data is applied. It considered the classification task in two domains: Breast tumors and skin lesions. Classifier outputs are used as a basis for computing beliefs. Dynamic uncertainty assessment is based on class differentiation. We combine the beliefs of three classifiers: k-Nearest Neighbor (kNN), Naïve Bayesian and Decision Tree. Dempster's rule of combination combines three beliefs to arrive at one final decision. Experiments with k-fold cross validation show that the nature of the data set has a bigger impact on some classifiers than others and the classification based on combined belief shows better overall accuracy than any individual classifier. Comparison of performance of Dempster's combination (with differentiation-based uncertainty assignment) with those of performance-based linear and majority vote combination models is shown.

Bearing in mind the usefulness and understandability of the application from a business perspective, combined rules of multiple patterns derived from different repositories, containing historical and point in time data, were used to produce new techniques [13] in association mining applied to debt recovery. Initially debt repayment patterns were discovered using transactional data and class labels defined by domain expertise, then demographic patterns were attached to each of the class labels. After combining the patterns, two type of rules were discovered leading to different results: 1) same demographic pattern with different repayment patterns, and 2) same repayment pattern with different demographic patterns. The rules produced are interesting, valuable, complete and understandable, which shows the applicability and effectiveness of the new method.

An efficient mechanism called Combined Mining, which makes use of Multi Feature Combined Mining to improve the accuracy of discovering informative knowledge from complex data sets. However, from Experimental results, it is observed that this Combined Mining Scheme is focused to improve the prediction accuracy to extract and discover the knowledge from complex Data Sets. However, it fails to find the missing value or noise, which leads to compromise classification accuracy. To address this issue, an efficient Classifier called Hybrid Combined Mining Technique is proposed [14] which is the combination of both the Missing Value Estimation Technique and Combined Mining Technique. From our experimental results, we have demonstrated that our proposed Hybrid Classifier achieves higher Classification Accuracy as compared with the existing Combined Mining Classifier

II. PROPOSED FRAMEWORK

A. Generation of combined data

Business database has both transactional and static data. Suppose there are two datasets, for e.g. Customer demographic dataset A which is static and Transactional dataset B. From the static dataset A, customers can be classified based on certain characteristics. Using classification algorithm like ID3, classify the customer for two groups {Y, N}. Before generating the combined data, domain knowledge is given to Feature Selection Method which selects the characteristics of static/demographic dataset as well as transaction dataset. For example consider targeting customers of age group 20-30 and gender as male in supermarket data, then only these characteristics of demographic data can be considered for selection . If the interesting product in campaigning are p1 and p2 then only those product from transaction data be selected before generating combined data. This Feature Selection Method reduces the overall size of dataset before applying traditional association mining to generate combined rules. Domain knowledge is the concept in the proposed work where the user can select the features from the dataset for both static and transactional data. This helps to reduce the total number of rules generated using classical rule mining.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Feature selection method (see Figure 1) takes the domain knowledge as a input before generating the combined data.

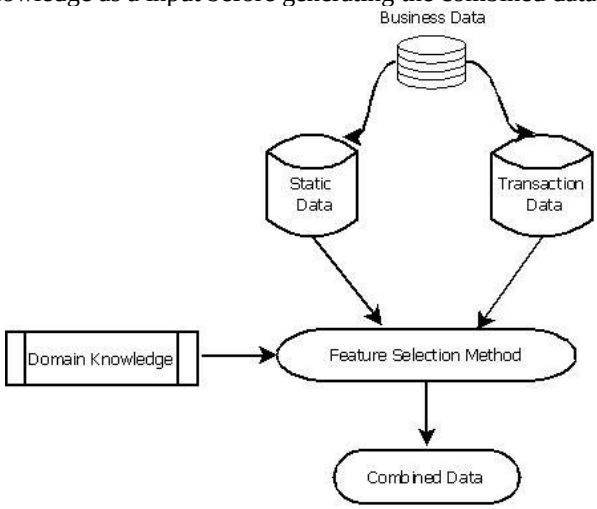

Fig. 1. Combine data generation.

Domain Knowledge Method :

1. List the transactional and demographic tables at run time
2. Select the tables in which the user interested
3. List the characteristics/features of selected table
4. List the distinct values of each selected characteristics of selected table
5. Select only the interested values of listed features

Domain knowledge method reduces the selection of data for feature selection method to generate the combine data. So the time complexity and space complexity for computation get reduced.

B. Pattern Generation with pruning

After generation of combined data, dynamically features can be selected with domain knowledge of the business. Dynamic feature selection method allows to select the interesting measures of the combined data which reduces the overall size of the dataset. This dataset is exposed to association rule mining to generate the patterns. A large number of patterns get generated through this method so to get the interesting patterns pruning is proposed. Pruning reduces the overall generated pattern with association rule mining . If the target problem has demographic dataset d and transactional dataset t. Dataset d has features like A={a1,a2,a3…….an} and dataset t has features like B={b1,b2,b3…….bm}. So a combined association rule R is in the form of X → T. Such that X may be subset of {A} or {B} or { A ^ B }, Where A and B are the interesting features selected from dataset d and t respectively. T ≠0 is a target item or class . If T is the set of two classes {Y,N} , then the combined association rules discovered will be as shown in Table1 .

TABLE I- PATTERNS GENERATED

Sr. No.	Rules
1	a1 → Y
2	b1 → Y
3	a1 ^ b1 → Y
4	a1 ^ b2 → N
5	a1 ^ a2 → N
6	a2 ^ b1 ^ b2 → Y
7	a3 ^ b4 → N
8	a1 ^ b4 → Y
9	b2 → N

III. PRUNNING

For each generated rule, interesting measures like Confidence, Lift and Leverage is calculated as explained in section IV. Pruning select only those rules which are having both the properties of static dataset d as well as transactional dataset t and lift >=1. In this step many of the rules get omitted as they are not showing properties of both dataset d and t. With reference to Table 1, rules 1,2,5 and 9 are not useful to generate combined patterns. Only the rules given in table 2 are considered for the generation of combined rules.

TABLE 2
Rules after prunning

Sr.No.	Rules
3	a1 ^ b1 → Y
4	a1 ^ b2 → N
6	a2 ^ b1 ^ b2 → Y
7	a3 ^ b4 → N
8	a1 ^ b4 → Y

Algorithm for pruning :

// To get the pruned rules
Input : Confidence , Lift , Patterns Generated on combined data above or equal to minimum support

1. For each pattern p
   a. If Confidence(p) > =confidence and Lift(p) >=1
      i. If (Check_Pattern_String(p))
      ii. Select the Pattern for Combined Mining
      iii. End if
   b. End if
2. Next pattern

Output : set of pruned rules

// Pattern string checking function
Check_Pattern_String(p)
1. For each pattern p (X → T)
   a. Check for left hand side of the rule
   b. If X contains {A1 ^ B1 }
      i. Consider it for combined rule
      ii. Return true
   c. Else
      i. Skip that rule
      ii. Return false

2. End for

Before going for pruning process, closed frequent patterns or maximal frequent patterns can be generated which reduces the number of patterns for pruning. A frequent itemset is one that occurs in at least a user-specific percentage of the database. That percentage is called support. An itemset is closed if none of its immediate supersets has the same support as the itemset. An itemset is maximal frequent if none of its immediate supersets is frequent. So the number of rules get reduced for pruning. The same algorithm can be applied for closed or maximal frequent patterns.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

IV. INTERESTINGNESS MEASURES FOR COMBINED PATTERN

A. *Support, confidence, Lift, Leverage and* I_{rule}

Assume that the rule is a ^ b → T

support: The support of the rule, that is, the relative frequency of transactions that contain a ^ b and T.

support(a ^ b → T) = *support(a+b+T)*

confidence: The confidence of the rule .

confidence(a ^ b → T) = *support(a+b+T) / support(a+b)*

lift: The lift value of the rule is the additional interestingness measures on the rules. These measures can then be used to either rank the rules by importance (and present a sorted list to the user) or as an additional pruning criterion .

lift(a ^ b → T) = *confidence*(a ^ b → T) / *support(T)*

Two new lifts for measuring the interestingness of combined association rules are [7] as

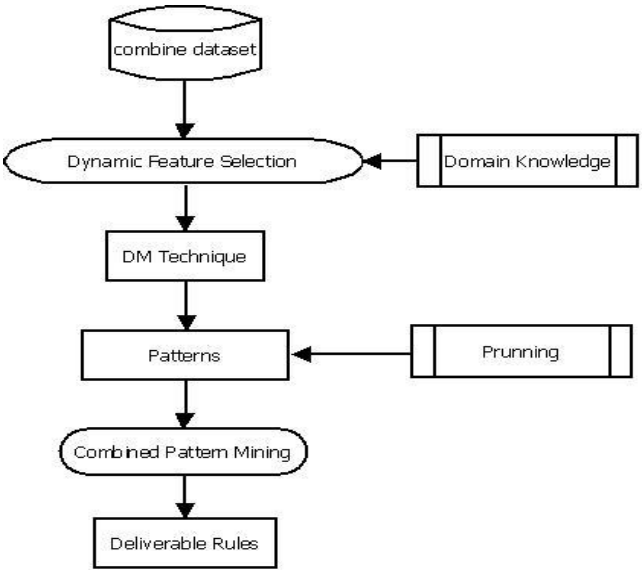

Fig. 2. Framework for multi-feature combined mining.

$Lift_a$(a ^ b → T) =*Lift* (a ^ b → T) /*lift* (*b* → *T*)

$Lift_b$ (a ^ b → T) =*Lift* (a ^ b → T)/*Lift* (*a*→ *T*)

Where $Lift_a$(a ^ b → T) is the lift of a with b as a precondition, which shows how much *a* contributes to the rule. $Lift_b$ (a ^ b → T) gives the contribution of *b* in the rule. Based on the above two new lifts, the interestingness of combined association rules is defined as Irule

Irule indicates whether the contribution of *a* (or *b*) to the occurrence of *T* increases with *a* (or *b*) as a precondition

I_{rule}(a ^ b → T) = $Lift_a$(a ^ b → T)/ *lift* (*a*→ *T*).

Or

*I*rule(a ^ b → T) = $Lift_b$(a ^ b → T)/ *lift* (*b*→ *T*).

The value of I_{rule} falls in [0,+∞). When I_{rule} > 1, the higher I_{rule} is, the more interesting the rule is.

Once the I_{rule} is calculated for each rule, arrange them in descending order, The rule with I_{rule} > 1 and higher I_{rule} value is more interesting.

Leverage measures the difference between the support for *a* ^ *b* → *T* and the support that would be expected if a ^ *b* and *T* were independent.

Leverage=support (a ^ b → T) – support((a ^ b) x support(T)

B. *Combined Rule Pairs and Cluster*

A combined rule pair is generated from the two contrasting rules. Assume that the rule is a ^ b → T where a is showing demographic/static features and b is showing transactional feature. Then a rule pair is composed with same demographic characteristics and different transactional characteristics which can have same or different target class i.e. T1 = T2 , T1 ≠ T2. The rule pair can also be generated for different demographic/static characteristics with same transactional characteristics.

From Table 2, rule pair can be generated for rule 3 and 4 (if their $I_{rule > 1}$)

$Rule_{Pair}$(p) : { r3 : a1 ^ b1 → Y , r4 : a1 ^ b2 → N }

$Rule_{Pair}$(p) measures the contribution of two different parts in antecedents to the occurrence of different classes . It helps to get a group of customers with the same demographic or the same transaction pattern. Such knowledge is vital to take business decisions for particular campaign to improve the business process.

Combined Rule clusters are then generated from rule pairs and the other combined association rules. Rule Cluster is composed of n combined association rules r1,r2……rn.

From $Rule_{Pair}$(p) and table 2 , Rule cluster C can be generated as

$Rule_{Cluster}$(C) = { $Rule_{Pair}$(p) ,
a1 ^ b4 → Y
}
= { r3 : a1 ^ b1 → Y ,
r4 : a1 ^ b2 → N,
r8 : a1 ^ b4 → Y
}

Rules from $Rule_{Cluster}$ (C) have same demographic characteristics but different transaction characteristics with different classes. It helps to get the knowledge about the similar customer group that how the customer changes the class with respect to different transactional patterns . Similar type of Rule cluster can be generated by changing the group of customer and by keeping the transactional pattern same.

TABLE 3 - Traditional Association Rules For Transactional Dataset

Rule (B->T)	Support	Confidence(%)	Lift
Bluetooth Headphones,Home Theater System->Moderate	13 /104	43	1.96
Bluetooth Headphones,Home Theater System->Low	17 /104	57	1.03
Televisions,VCD & DVD->Moderate	13 /104	22	0.98
Televisions,VCD & DVD->Low	37 /104	62	1.13

Table 4- Traditional Assoiation Rules for Demographic Dataset

Rule (A->T)	Support	Confidence(%)	Lift
M,Female->High	12 /104	38	1.62
M->High	19 /104	31	1.35
Female->High	16 /104	30	1.31
S,Male->Low	15 /104	68	1.24
S,Male->Moderate	6 /104	27	1.23
S->Low	28 /104	65	1.19

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

V. EXPERIMENTAL RESULTS

The proposed technique is tested on the subset of retail demo dataset used for presenting Microsoft Business Intelligence products which is available on microsoft.com. The sample subset of data contained 290 transactional records of 104 customers for 7 different products. Customer data was classified as High, Moderate and Low customers based on their features. The aim of the experiment was to find the association of demographic features of customer , product buying pattern and the class of customers which could help to give a promotional campaign on different products based on customer class and their demographic feature.

We used SQL server 2008 and dotnet technology for implementation purpose. For experiment purpose 7 products from transactional dataset and from the customer table two features Gender (Male, Female) and Marital Status(S-Single, M-Married) were selected through domain knowledge concept.

First traditional association mining is applied on transactional and demographic data separately.Subset of traditional association Rules for Transactional dataset and Customer dataset are shown respectively in Table 3 and Table 4. Minimum support was set to 10% and minimum confidence was set to 20% . We got 56 rules for transactional dataset 19 rules for demographic dataset which are above minimum support, confidence and Lift>=1.

After pruning and combined mining , the rules generated with both transactional and demographic features are shown in Table 5. Total 329 combined rules have been generated. From those combined rules 8 clusters were formed with same demographic feature and different transactional products with same or different class label as shown in Table 6. This result shows that how the different class customer with different demographic characteristics changes the product pattern. In table 6 , r3 : "Home Theater System , M , Male->Low" shows 70% confidence and 1.27 Lift. Though the Irule is same , Lift_B>Lift_A which suggest that the contribution of B with respect to Class is more as compared to A. Also the leverage of A and B in the same rule is 0.015 which is near to 0, therefore "Home Theater System" and " M , Male" are independent. Such type of knowledge help to give the promotional campaign on products with respect to the class of customer and their demographic features. The rules are readable and understandable to human to take business decision and can reduce the cost of promotion. For specific products, the rules can be generated with domain knowledge
technique. On the framework designed , the concept of feature technique. On the framework designed , the concept of feature selection is given which reduces the computational cost and space complexity also.

VI. CONCLUSION

This paper presents the new idea about pruning the association rules before making the rule pairs or rule clusters. Th domain driven concept gives the idea of selection of data features before generating the frequent patterns. So the user have freedom to select the product on which the company wants to give promotional campaign before generating the rules. Through domain driven user can select the customer characteristics from static customer data. The rule pairs can also be generated from pruned frequent patterns , or closed frequent patterns or maximal frequent patterns. So the number of rules generated get reduced and also readable and understandable to the user. This new technique is giving the cluster of rules for similar type of customers with their change of class as the transactional characteristic changes. This proposed technique gives more actionable rules than traditional association technique which help to improve business process.

Table 5 – Combined Association Rules

Rule_no	Rule	support	Conf(%)	Lift	Leverage	A	B	Class	Lift_A	Lift_B	Irule_A	Irule_B
r1	Bluetooth Headphones,Female,Home Theater System ->Moderate	7 /104	50	2.26	0.022	Female	Bluetooth Headphones,Home Theater System	Moderate	1.92	2.4	2.04	2.04
r2	Female,Home Theater System,MP4&MP3 ->Moderate	7 /104	50	2.26	0.022	Female	Home Theater System, MP4&MP3	Moderate	2.31	2.4	2.45	2.45
r3	Home Theater System,M,Male ->Low	10 /104	83	1.52	0.015	M,Male	Home Theater System	Low	1.13	1.37	1.01	1.01
r4	S,Home Theater System->Low	19 /104	76	1.39	0.013	S	Home Theater System	Low	1.03	1.17	0.87	0.87
r5	Bluetooth Headphones, Home Theater System,S ->Moderate	6 /104	30	1.36	0.013	S	Bluetooth Headphones, Home Theater System	Moderate	1.15	1.3	1.1	1.1
r6	S,Televisions, VCD & DVD->Low	23 /104	74	1.35	0.013	S	Televisions, VCD & DVD	Low	1.19	1.13	1	1
r7	Female,Home Theater System ->Low	16 /104	70	1.27	0.012	Female	Home Theater System	Low	0.94	1.41	1.05	1.05
r8	Bluetooth Headphones, M,Televisions ->High	8 /104	28	1.2	0.011	M	Bluetooth Headphones, Televisions	High	1.67	0.89	1.23	1.23
r9	Bluetooth Headphones,M,Male ->Low	8 /104	62	1.12	0.011	M,Male	Bluetooth Headphones	Low	0.99	1.01	0.89	0.89
r10	Bluetooth Headphones, M, MP4&MP3 ->Moderate	7 /104	24	1.09	0.010	M	Bluetooth Headphones, MP4&MP3	Moderate	1.11	1.14	1.16	1.16
r11	M,Televisions,VCD & DVD ->Moderate	7 /104	24	1.09	0.010	M	Televisions, VCD & DVD	Moderate	1.11	1.14	1.16	1.16
r12	S,Home Theater System ->Moderate	6 /104	24	1.09	0.010	S	Home Theater System	Moderate	0.92	1.04	0.88	0.88
r13	Bluetooth Headphones, Female, MP4&MP3 ->Low	17 /104	57	1.03	0.010	Female	Bluetooth Headphones, MP4&MP3	Low	0.91	1.14	1.01	1.01
r14	Bluetooth Headphones,M,Male ->High	3 /104	23	1	0.074	M,Male	Bluetooth Headphones	High	1.39	0.95	1.32	1.32

Table 5 – Combined Rules with Cluster

Clusters	Rule_no	Rule	support	confi(%)	Lift	Leverage	A	B	Class	Lift_A	Lift_B	Irule_A	Irule_B
c1	r1	Bluetooth Headphones,Female,Home Theater System->Moderate	7 /104	50	2.26	0.022	Female	Bluetooth Headphones,Home Theater System	Moderate	1.92	2.4	2.04	2.04
c1	r2	Female,Home Theater System,MP4&MP3 ->Moderate	7 /104	50	2.26	0.022	Female	Home Theater System,MP4&MP3	Moderate	2.31	2.4	2.45	2.45
c1	r7	Female,Home Theater System ->Low	16 /104	70	1.27	0.012	Female	Home Theater System	Low	0.94	1.41	1.05	1.05
c1	r13	Bluetooth Headphones,Female,MP4&MP3->Low	17 /104	57	1.03	0.010	Female	Bluetooth Headphones,MP4&MP3	Low	0.91	1.14	1.01	1.01
c2	r8	Bluetooth Headphones,M,Televisions->High	8 /104	28	1.2	0.011	M	Bluetooth Headphones,Televisions	High	1.67	0.89	1.23	1.23
c2	r10	Bluetooth Headphones,M,MP4&MP3->Moderate	7 /104	24	1.09	0.010	M	Bluetooth Headphones,MP4&MP3	Moderate	1.11	1.14	1.16	1.16
c2	r11	M,Televisions,VCD & DVD ->Moderate	7 /104	24	1.09	0.010	M	Televisions,VCD & DVD	Moderate	1.11	1.14	1.16	1.16
c3	r3	Home Theater System,M,Male ->Low	10 /104	83	1.52	0.015	M,Male	Home Theater System	Low	1.13	1.37	1.01	1.01
c3	r9	Bluetooth Headphones,M,Male->Low	8 /104	62	1.12	0.011	M,Male	Bluetooth Headphones	Low	0.99	1.01	0.89	0.89
c3	r14	Bluetooth Headphones,M,Male->High	3 /104	23	1	0.074	M,Male	Bluetooth Headphones	High	1.39	0.95	1.32	1.32
c4	r4	S,Home Theater System->Low	19 /104	76	1.39	0.013	S	Home Theater System	Low	1.03	1.17	0.87	0.87
c4	r5	Bluetooth Headphones,Home Theater System,S ->Moderate	6 /104	30	1.36	0.013	S	Bluetooth Headphones,Home Theater System	Moderate	1.15	1.3	1.1	1.1
c4	r6	S,Televisions,VCD & DVD->Low	23 /104	74	1.35	0.013	S	Televisions,VCD & DVD	Low	1.19	1.13	1	1
c4	r12	S,Home Theater System->Moderate	6 /104	24	1.09	0.010	S	Home Theater System	Moderate	0.92	1.04	0.88	0.88

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

REFERENCES

[1] Longbing Cao, Yanchang Zhao, Chengqi Zhang, "Mining Impact Targeted Activity Patterns in Imbalanced Data" *IEEE Trans* Knowledge And Data Engineering, Vol. 20, No. 8, August 2008, pp 1053-1066.

[2] Dina A. Salem, Rania Ahmed A. A. Abul Seoud, and Hesham A. Ali, "DMCA: A Combined Data Mining Technique for Improving the Microarray Data Classification Accuracy", in Proc International Conference on Environment and BioScience, IPCBEE vol.21 (2011), pp. 36-41.

[3] Ambikavathi .V, Veeraiah.A, Prabhu.R, Actionable Knowledge Discovery", International Journal Of Computational Engineering Research, Vol. 2, Issue No.1, pp 56-59.

[4] Longbing Cao, Yanchang Zhao, Chengqi Zhang and huaifeng zhang, "Activity Mining: From Activities to Actions" International Journalof Information Technology and decision making, Vol . 7 No 2, pp 259-273.

[5] Longbing Cao, Huaifeng Zhang Yanchang Zhao, Dan Luo, and Chengqi Zhang, "Combined Mining: Discovering Informative Knowledge in Complex Data", *IEEE Trans* Systems, Man, And Cybernetics—Part B: Cybernetics, Vol. 41, No. 3, June 2011pp-699-712.

[6] Longbing Cao, Huaifeng Zhang, Yanchang Zhao and Chengqi Zhang, "General Frameworks for Combined Mining: Discovering Informative Knowledge in Complex Data" IEEE Trans Syst Man Cybern B Cybern. 2011 Jun;41(3):699-712.

[7] Yanchang Zhao, Huaifeng Zhang, Longbing Cao, Chengqi Zhang, and Hans Bohlscheid, "Combined Pattern Mining: From Learned Rules to Actionable Knowledge", Lecture Notes in Computer Science, 2008, Volume 5360/2008, pp. 393-403.

[8] Marie Plasse, Ndeye Niang, Gilbert Saporta, Alexandre Villeminot, and Laurent Leblond , "Combined use of association rules mining and clustering methods to find relevant links between binary rare attributes in a large data set", Computational Statistics & Data Analysis Volume 52, Issue 1, 15 September 2007, Pages 596 – 613.

[9] Longbing Cao, "Domain Driven Data Mining (D3M)", IEEE International Conference on Data Mining Workshops, 2008, pp 74-76.

[10] Yanchang Zhao, Longbing Cao, Yvonne Morrow,Yuming Ou, Jiarui Ni, and Chengqi Zhang, "Discovering Debtor Patterns of Centrelink Customers", *in Proc*. AusDM '06 fifth Australasian conference on Data mining and analystics - Volume 61 Pp 135-144.

[11] Xiaoxin Yin, Jiawei Han, Jiong Yang, and Philip S. Yu, "Efficient Classification across Multiple Database Relations: A CrossMine Approach", IEEE Trans On Knowledge And Data Engineering, Vol. 18, No. 6, June 2006. pp 770 -783

[12] Y. Alp Aslandogan, Gauri A. Mahajani,"Evidence Combination in Medical Data Mining", Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004, Volume: 2 pp 465 – 469.

[13] Yanchang Zhao, Huaifeng Zhang, Fernando Figueiredo, "Mining for Combined Association Rules on Multiple Datasets", Workshop on Domain Driven Data Mining (DDDM2007), August 12, 2007

[14] T. Nandhini, and N. K. Sakthivel, "A Smart Hybrid Data Mining Technique for Improving Classification Accuracy in Complex Data Sets", European Journal of Scientific Research, Vol.71 No.2 (2012), pp. 233-242.

[15] Guimei Liu, Haojun Zhang, Limsoon Wong,"Controlling False Positives in Association Rule Mining", *Proceedings of the VLDB Endowment,* Vol. 5, No. 2 (2011)

[16] Sotiris Kotsiantis, Dimitris Kanellopoulos, "Association Rules Mining: A Recent Overview", GESTS International Transactions on Computer Science and Engineering, Vol.32 (1), 2006, pp. 71-82

[17] Evelina Lamma, Fabrizio Riguzzi and Sergio Storari, "Improving the K2 Algorithm Using Association Rule Parameters", c 2005 Elsevier B.V.

Arti Deshpande is working as Assistant Professor in Thadomal Shahani Engineering College , Mumbai and doing her PhD in Computer Science and Engineering from G.H. Raisoni College of Engineering , Nagpur. She is BE and ME in Computer Science and Engineering. She has more than 14 years of experience. Her area of specialization is Data Warehousing and Mining , Business Intelligence . She has published books on Data Warehousing and Mining for Mumbai and Gujrat University .

.

Dr. A. R. Mahajan is working as Professor & Head of Computer Science and Engineering Department, PIET, Nagpur, India. She is BE and ME in Computer Science and Engineering She has completed her PhD in Computer Science and Engineering. She has presented twelve papers in International Journal and one paper in national Journal She has published Twenty eight papers in International conferences and five papers in national conferences .She has more than Eighteen years of experience. Her area of specialization is compiler optimization, Artificial Intelligence, Parallel algorithms.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)