

# Generalization Driven Fuzzy Classification Rules Extraction using OLAM Data Cubes

**B.Raghuram**

Department of Computer Science & Engineering, Kakatiya Institute of Technology and Science, Warangal, TS-506015, India.

## Abstract

An fuzzy classification rules extraction model for online analytical mining (OLAM) was explained in this article. The efficient integration of the concept of data warehousing, online analytical processing (OLAP) and data mining systems converges to OLAM results in an efficient decision support system. Even after associative classification proved as most efficient classification technique there is a lack of associative classification proposals in field of OLAM. While most of existing data cube models claims their superiority over other the fuzzy multidimensional data cubes proved to be more intuitive in user perspective and effectively manage data imprecision. Considering these factors, in this paper we propose an associative classification model which can perform classification over fuzzy data cubes. Our method aimed to improve accuracy and intuitive ness of classification model using fuzzy concepts and hierarchical relations. We also proposed a generalization-based criterion for ranking associative classification rules to improve classifier accuracy. The model accuracy tested on UCI standard database.

**Key words—** Associative Classification, OLAM, Classifier accuracy, Fuzzy Data Cubes, Generalization.

## I. Introduction

OLAM is an analytical mining process that combines features of on-line analytical process (OLAP) and data mining techniques. On-line analytical mining is first proposed by J.Han [1]. The conventional mining proposal proved highly complex for multidimensional data mining and also provides results with interpretation difficulties. Consideration of user defined meta rules is also a difficult task in traditional mining approaches. On the other side OLAP [2] is proved as a powerful data analysis technique for multi-dimensional analysis. Considering the effective and easiness of OLAP technology in data analysis, the proposal for On-Line Analytical Mining came which combines the features of OLAP operations with data mining technique in order to perform effectively on multidimensional databases and data cubes. The advantage of OLAM is that using OLAP operators it provides interactive data mining by allowing user to

select required dimensions, abstraction levels of data and also make it possible to view results using different visualization techniques. The consistent input data that proved to be shows significant effect on accuracy of data mining task will be provided by data warehouse.

The OLAM also improves the computational efficiency of mining process by using pre computed aggregation values and pre calculated support values across data cubes [3]. In OLAM mining accuracy can be improve by considering complex inter dimensions and multi level hierarchies with help of OLAP operations.

Features of using Data cubes for OLAM will play important role in providing different functionality and applications. Critical issues like considering null values, efficient indexing of data, treating data imprecision, including fuzzy logic [4] and storing

specialized data like XML, object oriented data [5] will be handled by specialized data cubes only. While most of the existing OLAM models claim their accuracy and superiority over other models based on fuzzy data cubes proved to be more intuitive in user perspective. In addition fuzzy data cubes proved to be effective in managing data imprecision which is caused by merging data from different sources. These factors played an important role in motivating us to propose classification model that works on data cubes which can support data imprecision.

Efficient implementation and fast response is the major challenge in the realization of OLAM. This proposal focuses on efficient implementation of associative classification on fuzzy multidimensional data cubes. The classification is a significant technique in data mining with applications in industrial and scientific applications. The classification techniques based on decision tree method and neural networks are mainly used for classification of data cubes in OLAM models. The efficiency of a classification model is evaluated by two parameters, namely the accuracy and the interpretability of the model. Many studies show that Associative Classifiers give better accuracy and interpretability than other traditional classification models [6]. The idea is to utilize frequent patterns and association relationships between elements and class labels in training data set to do classification. If strong associations among some frequent patterns and class labels can be observed in training data set, the future object of similar patterns can be classified. Since association rules explore highly confident associations among multiple variables, it overcomes the constraints introduced by a decision-tree method which examines one variable at a time. Associative classification process is interactive and easily understandable to end-user.

Even after associative classification proved to be accurate there is lack of proposed methods that can perform associative classification on data cube. In addition recent studies show that association classifiers may also suffer some weaknesses inherited from association rule mining such that, large set of mined rules, maintaining same support and confidence values for all level of hierarchies, difficulty in handling imprecision which cause due to margin data from

unreliable sources, problems in handling continuous data and crisp structure maintained for aggregation all these may not appear intuitive to end the end user.

In order to tackle the above problems, this work reports about the development of an associative classification model on fuzzy multidimensional model which manages data imprecision. Generalized association rules aggregates most of low level rules which overcome the complexity of large number of rules and rule over fitting which is caused because using very less number of rules for classification. The use of fuzzy data cube provides intuitive association rules and also effectively manages data imprecision. Along with that this model also possess the advantages of OLAM system. Our approach follows Classification Based on Association (CBA) model [6]. As per that our proposed model first generates heuristic associative classification rules using training data cube to which class labels for each transaction is known. In next step a compact set of high confidence associative classification rules will be selected by applying generated rules on training data cube.

Our proposed model uses a fuzzy multi dimensional model proposed by Carlos Molina et al [7]. The model provides aggregations intuitively to end user by means of fuzzy logic. The user has to explicitly specify the threshold value for aggregation operator to which fuzzy level (value 0 to 1) taxonomies should be performed. All OLAP operators of model can effectively manages data imprecision resulting from merging of data from heterogeneous sources both in facts and dimensions. The model provides low fuzzy confidence value for data from unreliable source comparing to data from reliable sources. The information with confidence more or equal to user specified threshold will be consider by aggregation operators. Motivating from the efficiency of the model in performing fuzzy taxonomies and managing imprecision we adopted for our model.

The rules, which may not obtain in base level may be possible at generalized higher abstraction level. Many of proposed associative classification models would not consider relations at higher abstractions. In order to overcome that our proposed model consider rules at multi level hierarchies. In order to realize that for association rule extraction our model adopts frequent item set

generation model proposed by Nicolas Marin et al [8]. The approach calculates support for item sets at base level first if the item would not prove frequent at this step it generalized items to its higher abstraction and calculates support again. In order to overcome using same confidence

level for all abstraction the model also provides interesting mechanism for calculating threshold support at higher level of abstractions.

To the best of our knowledge there are no proposals to perform classification on fuzzy data cubes using effective associative classification method. Considering this lack of proposal and with intention to overcome the problem faced by regular associative classification we propose an associative classification model to classify fuzzy multidimensional data cube. Our proposal has the following goals.

- To provide choice to user for selecting required dimensions and hierarchies for classification.
- To provide classification intuitively in end user point of view.
- To provide an accurate and less complex final classification rule set.
  - To manage data imprecision in process. The outline of the remaining paper is as follows.

Section II reviews the related associative classifying methods together with their main weaknesses. The Architecture of proposed model and the algorithm are presented in section III. The experimental results are presented in section IV. This paper ends with conclusion and future work.

## II. Related Work

The Classification Based on Associations (CBA), proposed by Liu et al [6] is a first association rule-based classifiers reported in the literature, generates all the classification association rules (CARs) that satisfy user-defined support and confidence thresholds. A similar method, called msCBA [9], was introduced later in which multiple classes with minimum support values were used. A new type of associative classifier was proposed by Dong et al, Classification by Aggregating Emerging Patterns

[10] that use emerging patterns to provide a local description of the data. The Large Bayesian algorithm proposed by Meretakis & Wuthrich [11], also based on item sets. Association based decision tree methods were introduced by Wang [12].

Instead of generating a large number of candidate rules, Yin proposal based on Classification Predictive

Association Rules (CPAR) in [13] which adopt a greedy algorithm to generate rules directly from training data. Correlated Association Rule Mining for Classification proposed by Zimmermann [14] also finds rules by employing a branch-and-bound technique. Recently, a multi-class classification algorithm proposed by Thabtah [15], where vertical data format are used to discover the frequent item sets.

One of the main problems in above mentioned association rule mining is that the number of discovered rules is huge. This could defeat the purpose of readability of the classification model. Classification based on Multiple class-Association Rules (CMAR) introduced by Li [16], applied FP-growth method. The importance of the minimum rule sets for classification is also outlined by Yang [17]. They present their minimum set rule algorithm and investigate the relationship between error rates and minimum support/ confidence. In [18] Hu & Li suggested an optimal association classifier (OAC) that is less sensitive to the missing values in an unseen test data.

The problem how to set and tune the basic mining parameters outlined by Zaiane & Antonie [19] and a new rule pruning approach is presented applying the so-called ARC-AC classifier developed by Zaiane et al [20]. The effect of using different threshold values on association rule-based classification accuracy is also analyzed by Coenen & Leng [21]. Some works replaced the support–confidence framework with other approaches to obtain more efficient rule selection methods. Antonie & Zaiane [22] introduced both positive and negative association rules for building an efficient classifier. An information gain based association rule classification (GARC) method was also proposed by Chen et al [23].

The above specified all associative classifications suffer from the following problems like huge set of rules which are hard to interpret, not considering rule across multi levels which may reduce accuracy because rule not found in lower level may found in higher abstraction levels and crisp logic based rules presented by these rules may not give intuitive look to end user.

## III. Architecture And Algorithm

The architecture of proposed generalization based associative classification model (GBAC) shown in

figure 1. This model performs classification on fuzzy multidimensional data cube [7], which can support data imprecision resulting due to integration of data from different sources. The proposed GBAC model consist of three parts, at first part frequent fuzzy item sets will be generated using proposed item set generation algorithm. Then using frequent item sets associative classification rules generation and pruning will be performed. In the next step associative classification rule will be ranked as per proposed ranking model and in final step is for classifier builder which is based on CBA-CB: M2 algorithm [6]. These three steps will be implemented by OLAM engine on fuzzy multidimensional data cube. At starting point user can select his required dimensions and hierarchies using OLAP operations. The user should mention threshold value for support and confidence for acceptance of item sets and rules. In the frequent fuzzy item set generation process the algorithm will make use of OLAP operations for generalization purpose. In generalization process data cube perform fuzzy logic based aggregation. The data imprecision will be managed implicitly by fuzzy multidimensional data cube [7]. The model provides less number of intuitive and easily interpretable associative classification rules by which data cubes can be classified accurately.

- 2) Generate and prune rules using the item sets found in the previous step.

Proposed algorithm shown in section 3.2. Initially proposed algorithm follow Apriori algorithm [24] like method ‘candidateGen’ to obtain frequent one item sets at the base level of all the dimensions. Each item set is in form of  $\langle \text{condset}, L \rangle$  where condset is a set of items and L is class label. While generating item sets the algorithm will consider same items with different class labels as different items sets in order to obtain frequent item sets specialized for class labels. The candidate item support count and rule support count for iterative passes to data cubes calculate each item set. The number of iterations is equal to number of item sets generated. The candidate item support is equal to number of occurrences of that item in data cube. In the same scan algorithm will calculates the rule support count of items using ‘ruleSubSet’ function. If a candidate item class and transaction class matches then improves respective item rule support count by one. The confidence value is calculated for rule items by formula

$$\text{Confidence} = (\text{rulesupCount} / \text{condsupCount}) * 100 \% .$$

(1)

A rule item is frequent if its rule support is equal to or greater than threshold value. For the base level, the algorithm uses a value given by user.

If the item set is not frequent, then algorithm will generalize item to its higher abstraction level previously specified in cube. While generalizing to higher abstraction level the cube will follow fuzzy taxonomies. The new item sets obtained are considered as candidates and rule support count is calculated again. This process is repeated until the rule item set is accepted as frequent or no longer can generalize.

Elements at higher levels may group several values at the base level. The problem with most of multilevel association rule generation approach is they consider same confidence value in all hierarchical levels which may reduce efficiency of data mining task by including huge set of rule sets. This problem is known as rule over fitting. So for obtain an intuitive and accurate classifying model support threshold for higher abstraction levels should be higher than support at lower abstraction level. In order to avoid this problem, we use the automated threshold generation of higher

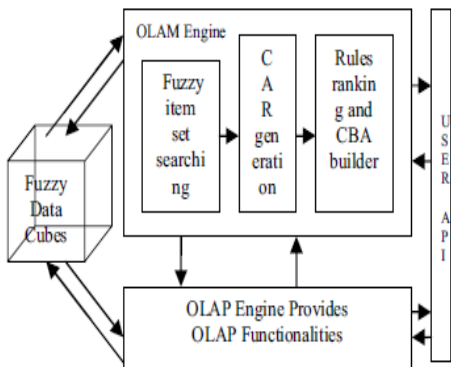


Fig. 1. The Architecture of GBAC model for classifying Fuzzy Multidimensional Data Cubes.

### 3.1. Associative Classification Rule Generation

In this phase the proposed algorithm extracts associative classification rules at different dimensions and multiple levels by making multiple passes to data cubes. The algorithm performs following steps:

- 1) Obtain frequent candidate rule items with and without specific class labels.

abstraction model proposed by Nicolas Marin [8] is used. The model calculates threshold as follows: for an aggregated item set the support threshold is defined as:

$$\text{Threshold}_I = \text{threshold}_{\text{sup}} + (1 - \text{threshold}_{\text{sup}}) * A(I). \quad (2)$$

Where  $\text{threshold}_{\text{sup}}$  is the support threshold established by the user for the basic levels and  $A(I)$  is abstraction value [0 to 1] of item I. Abstraction value for a item provided by data cube depending up on abstraction level of item. In a data cube elements at higher abstraction will present a higher abstraction value (A

(I) value) than elements at lower abstraction. The abstraction of the rule item set is average of all individual item set abstraction values.

After all one item sets are processed in the above manner the algorithm collects the rule items whose support count is more than threshold values. Using this frequent rule items the algorithm produce the associative classification rules using the 'genRule' function performs pruning operations using confidence value. At first pruning will be performed using rule items that have the same condition set. Out of the rule items with same condition set the highest confidence set is chosen as the possible rule items. Finally out of possible rule item set the item set which pass minimum confidence threshold is consider as associative classification rule.

These rules will be pruned again by 'pruneRule' function using pessimistic error rate method specified in C4.5 model [25]. The pessimistic error rate method is efficient pruning model which provides more concise rule items for further processing steps. This pruning method is optional one and can be replaced with any other standard pruning method.

In next iteration the algorithm generates item sets using a rule item which has passed the threshold of minimum support count. The above specified steps will

be repeated on all item set generated. The process will continue until all the possible item sets have been processed. Finally all the obtained pruned associative classification rules are made into one set.

### 3.2 Algorithm for Generalization Based Associative Classification Rule Generator

Input:

- $\text{threshold}_{\text{sup}}$  : Support threshold to consider an candidate item as frequent
- C: DataCube
- $\text{min}_{\text{sup}}$  : Support threshold to consider an

rule item as frequent

Output:

- set of  $\text{CAR}_k$ ,  $\text{PrCAR}_k$  Algorithm:
- 1.  $C_f \emptyset$
- 2.  $k \ 0$
- 3.  $c_1 =$  elements at the base le dimensions
- 4. for (  $k=2$  ;  $C_{fk-1} \emptyset$  ;  $k++$ )
  - a)  $C_k =$  candidateGen( $C_{fk-1}$ )
  - b) while  $C_k \ 0$ 
    - i) I = first element of  $C_k$
    - ii)  $C_k = C_k - I$
    - iii) for each transaction  $T \in C$  do
      - A) if (ruleSubSet(I,T)) then
        - I) I. candidateSupcount++
      - B) if (T.class = I.class ) then I) I.ruleSupcount++
    - iv) End
    - v) if (I.ruleSupcount  $\text{threshold}_{\text{sup}} + (1 - \text{threshold}_{\text{sup}}) * A(I)$ ) then
      - A)  $C_{fk} \ C_{fk} \cup \{I\}$
      - vi) else
        - A)  $C_k \ C_k \cup \text{Generalize}(I)$
  - c) End
  - d)  $\text{CAR}_k = \text{genRules}(C_{fk})$
  - e)  $\text{PrCAR}_k = \text{pruneRules}(\text{CAR}_k)$ 
    - 5. End
    - 6.  $\text{CAR}_s = \cup_k \text{CAR}_k$
    - 7.  $\text{PrCAR}_s = \cup_k \text{PrCAR}_k$

### 3.3. Improved Associative Classifier

In order to build a classifier using generated associative classification rules we are using CBA-CB [6] model. In order to improve accuracy of CBA-CB model with overlapping rules we adopted one more rule (3 in following list) to three rules for rule ranking method proposed in CBA-CB.

Given two rules  $r_i$  and  $r_j$ ,  $r_i > r_j$  (i.e.,  $r_i$  precedes  $r_j$  or  $r_i$  has higher precedence over  $r_j$ ) if one of the following holds good:

- 1) The confidence of  $r_i$  is greater than that of  $r_j$
- 2) Their confidences are the same but support of  $r_i$  is greater than that of  $r_j$
- 3) **Generalized Rules:** all the above constraints are the same but sum of abstraction values of rule items in  $r_i$  greater than sum of abstraction values of rule items in  $r_j$  ( $A(r_i) > A(r_j)$  ). The rules that are more general are preferred over rules which are more specific. That is, the rules, which cover potentially more instances, will be preferred.
- 4) All the above constraints are the same but  $r_i$  is generated before  $r_j$

By introducing third constrain, better accuracy can be achieved. The introduction of the new criteria has reduced this randomness in the same confidence and support rule selection, and hence the improvement in the predictive accuracy. After

finalizing ranks of rules an interesting and compact rule set will be collected by applying rules on test data cube using standard CBA-CB: M2 [6] algorithm for Classifier Builder. The model is heuristic hence it will generate classifier with maximum efficiency. The resulted classifier can be applied for classifying fuzzy data cubes to which class labels or known.

#### IV. Experimental Results

The implementation of our algorithm is a java program. To evaluate the accuracy of the proposed classifier generated by our GBAC model, we tested this algorithm on the Heart deices data namely Hungarian processed heart data sets obtained from UCI Machine Learning Repository.

The aim of experiment is to generate associative classification rules that classify the data cube transactions depends up on severity low (0) or high (1) by analyzing given dimensions that can show effect on heart deices. The experiment conducted in three steps. At first step of experiment proposed algorithm implemented on training fuzzy data cube for which class labels of each transaction are known to generate associative classification rules. In next step generated rule sorted using the proposed rule ranking model and set of effective rules collected using CBA-CB: M2 algorithm. In final step classifier applying obtained rule set on test DataCube collects efficiency. The details of experiment given below.

##### 3.1. Data Cubes

At first we implemented a fuzzy multidimensional data cube defined over heart data, for that purpose we obtain multidimensional schemas that model the information in fuzzy perspective. Each transaction in data cube is classified into two classes that is severity low (0) or high (1). We briefly explain the structure of the multidimensional schema. Dimensions:

Data has been analyzed in the following dimensions. The possible values at base level and pre specified aggregations are stated here.

- Age: in this dimension the most detailed levels consider age for 20 to 100, in next hierarchy the age 20 to 35 aggregated as young, age 35 to 50 aggregated as mid, age 50 to 100 considered. Finally All grope. Age = ({20—100}, {young, mid, old},{All})
- sex: This dimension can took two values male(1) and female(0) and can aggregated to human Gender = ({0,1},{All})

- Blood pressure (in mm Hg): in this dimension blood pressure level 100 mm Hg to 180 mm Hg consider and is aggregated in three groups 100 to 130 as normal, 125 to 150 high, 145 to 180 sever. Blood pressure = ({100-180}, {normal, high, severe})
- Serum cholesterol (in mg/dl): the cholesterol level 100 mg/dl Hg to 450 mg/dl consider and is aggregated in three groups 100 to 250 as normal, 200 to 350 high, 300 to 450 sever. Serum cholesterol = ({100-450}, {normal, high, severe})
- Thalach: Maximum heart beat achieved can take values of 80 to 180 these things are grouped as low < 110, normal 105 to 165, high > 160. Thalch = ( {80-180}, {low, nor, high} )
- Old peak: this dimension can take values from 0 to 3 these value aggregated as 0 normal,1 to 2 high, 1.5 to 3 sever. Old peak = ( {0 -5},{nor, high, sever})

The remaining dimensions cp, fbs, restecg, exang, slope and thal, taken without any aggregations. The final field num will specify severity class of transaction by values low (0) or high (1).

Measures: The only fact or measure value we considered in this experiment is the number of transactions with the same values for all considered dimensions. There fore while calculating the support of a dimension in a cubes transaction we should add its fact value to the support count of that dimension.

Age	Sex	Blood Pressure	Serum Cholesterol	Thalach	Old Peak	Severity
20-35	Male	100-130	100-250	80-110	0-1	0
35-50	Female	125-150	200-350	105-165	1-2	1
50-100	Male	145-180	300-450	>160	1.5-3	1

Fig. 2. Generalized Associative classification rules for Severity high (1) class

##### 3.2. Classification Builder

The minimum support count observed for this experiment is 2% of available transactions. The confidence value for rule acceptance in this

experiment is 50%. Rules generated from algorithm sorted as per proposed rule sorting model and applied on heart data cube transaction whose class labels are known and obtain most effective rule set using CBA:m2 [6] model. Figure 2 shows the screen shot of generalized associative classification rule set for severity high (1) class.

### 3.3. Evaluation

From this experiment we observed that our proposed ranking model efficiently deal with rules consisting similar support and confidence values. It generated less number easily interpretable and generalized rules.. This generated classification model shown accuracy of 82.81% on test data cube of UCI Hungarian heart dataset based cubes with less number of generalized rules. The Table 1 shows accuracy of basic associative classifiers CBA [6], CMAR [16], CPAR [13] and OAC [18] models on UCI heart database (collected form literature) and accuracy of proposed GBAC model. On comparing accuracy of GBAC model to other associative classification model accuracy on same UCI heart database proves that our proposed GBAC model is effective associative classifier. Apart from accuracy our proposed model has got advantages provided by OLAM systems.

Model Name	CBA	CMAR	CPAR	OAC	GBAC
Accuracy	81.5	82.2	82.3	81.5	82.81

Table. 1. Classification accuracy of Models

### V. Conclusion And Future Work

The aim of this paper is to develop an accurate and intuitive associative classifier in field of OLAM. For that purpose we proposed a model and algorithm to extract associative classification rules over fuzzy data cubes. It attempts to improve efficiency and intuitiveness of classification by using fuzzy concepts and hierarchical relations. The fuzzy multidimensional data cube we used in this proposal also supports data imprecision implicitly. As per our knowledge this is first approach for associative classification in field of OLAM. We also proposed generalization-based criteria for associative classification rule ranking. The experimental result obtained proves the efficiency of algorithm.

In proposed algorithm generalization is performed only in the item set generation process, but it is proved that performing generalization at rule generation process

will provide less number and more interpretive results. Studying associative classification builder using such rules will be interesting.

### References

- [1] J.Han. "Towards on-line analytical mining in large databases ", ACM Special Interest Group on Management of Data, SIGMOD vol. 27, no. 1, pp. 97–107, 1998.
- [2] C.E.F.Codd, S.B.Codd and C.T.Salley "Providing OLAP to User-Analysts", E.F.Codd Associate, 1993. Jim Gray, Surajit Chaudhuri," Data Cube: A Relational Aggregation Operator Generalizing Group-By, Cross-Tab, and Sub-Totals", MIT, Data Mining and Knowledge Discovery, PP 29–53, 1997.
- [3] J.Int "Querying fuzzy multidimensional database: Unary operators and their properties", Uncertain Fuzziness Knowledge Based System. Vol 11, no. Supplement, pp. 31-45, 2003.
- [4] Jiawei Han and Shojiro Nishio "Generalization Based data mining in object-oriented databases using object cube model", IT, Data and Knowledge Engineering, 1998.
- [5] B.Liu, W..Hsu & Y.Ma. "Integrating classification and association rule mining". Knowledge discovery and data mining ,pp. 80–86,1998
- [6] Molina, D. Sanchez, M. A. Vila, and L. Rodríguez-Ariza, "A new fuzzy multidimensional model," *IEEE Trans. Fuzzy Syst.*, vol. 14, no. 6, pp. 897–912, Dec. 2006.
- [7] Nicolas Marin, Carlos Molina, José M. Serrano, and M. Amparo Vila. "A Complexity Guided Algorithm for
- [8] Association Rule Extraction on Fuzzy DataCubes ", IEEE Transactions on Fuzzy Systems, VOL. 16, NO. 3, pp 693-714, june 2008.
- [9] B.Liu, C. K .Wong, "Improving an association rule based classifier", Principles of data mining and knowledge discovery, pp 4–509, 2000.
- [10] G.Dong, X.Zhang , L.Wong, & J.Li., "CAEP: Classification by aggregating emerging patterns", Discovery Science, pp. 30–42, 1999.

- [11] D.Meretakis & B.Wuthrich, "Extending naive bayes classifiers using long item sets". Knowledge discovery and data mining, pp. 165–174, 1999.
- [12] Wang & Y.He, "Growing decision tree on support less association rules", knowledge discovery and data mining pp 265- 269, 2003.
- [13] J.Han & X.Yin, "CPAR: Classification based on predictive association rules". Proceedings of 2003 SIAM international conference on data mining, 2003.
- [14] A.Zimmermann, & L.D.Raedt, "CORCLASS: Correlated association rule mining for classification". Discovery science, 7<sup>th</sup> international conference, Italy ,pp 60–72, 2004
- [16] F.Thabtah, P. Cowling, I. Peng, "A study of predictive accuracy for four associative classifiers", Journal of Digital Information Management, vol 3, pp202–205, 2005.
- [17] W.Li, J. Han & J.Pei, "CMAR: Accurate and efficient classification based on multiple class-association rules ". ICDM, pp. 369–376, 2001.
- [18] J.Yang, "Classification by association rules: The importance of minimal rule sets". The twentieth international conference on machine learning, ICML, 2003.
- [19] H.Hu, & J.Li," Using association rules to make rule-based classifiers robust". Proceedings of the sixteenth Australasian database conference pp. 47–54, 2005.
- [20] O.Zaiane, R. & M.-L. Antonie." On pruning and tuning rules for associative classifiers". Ninth international conference on knowledgebased intelligence information and engineering systems, KES'05, pp. 966–973, 2005.
- [21] O.Zaiane,M.L.Antonie &A.Coman, " Mammography based classification by an association rule-based classifier". ACM, SIGKDD, pp. 62–69, 2002.
- [22] F.Coenen, & P.Leng, "The effect of threshold values on association rule based classification accuracy". Data & Knowledge Engineering, vol.60 (2), pp 345–360. 2007
- [23] M.L.Antonie, & O.Zaiane, "An associative classifier based on positive and negative rules". Proceedings of the 9th ACM SIGMOD pp. 64–69. 2004
- [24] G.Chen, H.Liu & Zhang, "A new approach to classification based on association rule mining ", Decision Support Systems, Vol. 42(2), pp 674–689, 2006.
- [25] R.Agrawal and R. Srikant, "Fast algorithms for mining association rules in large databases ", in Proc. 20th Int, Conf, VLDB, pp. 478-499, 1994.
- [26] J.Quinlan, "C4.5: programs for machine learning", Morgan Kufman, 1993.