



## An Empirical Algorithm for High and Low Correlative Association Rule Mining

Hemant Kumar Soni<sup>1\*</sup>    Sanjiv Sharma<sup>2</sup>    Arvind Kumar Upadhyay<sup>1</sup>

<sup>1</sup>*Department of Computer Science and Engineering, Amity School of Engineering and Technology,  
Amity University, Gwalior, Madhya Pradesh, India*

<sup>2</sup>*Department of Computer Science and Engineering,  
Madhav Institute of Technology and Science, Gwalior, Madhya Pradesh, India*

\* Corresponding author's Email: [hksoni@gwa.amity.edu](mailto:hksoni@gwa.amity.edu)

---

**Abstract:** Support and confidence based Association rules mining algorithms have certain problems. Although other metrics like interest factor, comprehensibility, lift, correlation etc. are available to measure the interestingness of association rules. All the objectives are not suitable for each and every situation. All the objectives which were proposed in the literature, have some drawback, like correlation analysis gives equal importance to the items those are present and absent in transaction database. Resultant the rules generated by this, sometime mislead decision makers. Hence there is a strong need to define some new objectives for association rules that support in effective decision making. In this paper, authors proposed two novel objectives, high correlation and low correlation for 2-variables and 3-variables. These novel objectives clearly indicate that how much or how less two/three items are correlated. On the basis of this, decision makers can form their business strategies. An empirical algorithm for high and low correlative association rules generation is also proposed. With numeric evolution and experiments on the real-life data set, effectiveness has been measured and found that proposed algorithm gives better results.

**Keywords:** Association rule mining, High correlation, Low correlation, Multi-objective.

---

### 1. Introduction

Identification of the relationship among items in a transactional database is a significant task of data mining through association rule mining. Application domain of association rule mining is very vast including fraud detection, recommender system, intrusion detection etc. The interesting rules generation in an effective and efficient manner is a driving force in the data mining research community. Starting from Apriori algorithm [1], there is a number of the algorithm has been proposed for association rules generation [2].

Identification of co-occurrence of items in a transactional database is the main task of association rule mining. The market basket analysis is a distinctive example of association rule mining. In this, the buying patterns of the customer are analyzed. The interesting and comprehensible patterns generated by this can lead to effective

decision making. On the basis of association and/or correlation among items, decision makers form their strategies to enhance business growth. Most of the association rules mining algorithms are based on only two objectives- support and confidence. Whereas association rules mining is a multi-objective problem. In literature some other objectives are also defined as correlation, interestingness, comprehensibility, lift, j-measure etc [3]. All the objectives are not suitable for each and every situation. All the objectives which were proposed in the literature, have some drawback, like correlation analysis gives equal importance to the items those are present and absent in transaction database. Resultant the rules generated by this, sometime mislead decision makers. Hence there is a strong need to define some new objectives for association rules that support in effective decision making.

The main contributions of this work are as follows:

- a. We introduce two new objectives namely high correlation and low correlation to measure the actual relationship between items.
- b. We proposed an empirical algorithm which generates high and low correlative association rules. Very few researchers discuss and discover positive and negative association rules, rare association rules, multi-objective association rules. The proposed algorithm uses a different interestingness measure and it generates high and low correlative association rules.
- c. We compare our algorithm with other existing algorithms and discuss their performances.

The rest of the paper is divided into 9 sections. Section 2 explains basic concepts and terminology. Section 3 discusses various objectives for association rule mining. The drawback of these objectives is given in section 4. Section 5 introduces 2 new objectives. A novel algorithm is proposed in section 6 and literature review is given in section 7. Experimental results are discussed in section 8. Paper concludes in section 9 as a conclusion and future scope.

## 2. Basic concepts and terminology

Data mining's prime techniques include generating association rules. The association rules mining first introduced in [1, 4-5]. The objective of data mining is to fetch appealing correlation, regular pattern and relation amongst items in the transaction database. In association rule mining, only those items, whose minimum support and confidence above the threshold are taken as an association rule. The entire association rule mining process is divided into two subsections. In the first subsection, the process identified those items whose repetition is above pre-decided threshold. These itemsets called frequent or large itemsets. The second subsection of the process identified association rule by large items sets which qualify minimum confidence [6-7].

### 2.1 Association rules

For a given transaction database T, An association rule is an implication of the form  $X \Rightarrow Y$ , where  $X \subset I, Y \subset I$ , and  $X \cap Y = \Phi$ , i.e. X and Y are two non-empty and non-intersecting itemsets.

### 2.2 Support

A transaction T is said to support an item  $i_k$ , if  $i_k$  is present in T. T is said to support a subset of items

$X \subseteq I$ , if T support each item  $i_k$  in X. An itemset  $X \subseteq I$  have a support s in D, denoted by  $s(X)$ , if s% of transactions in D support X.

$$\text{Support}(X \Rightarrow Y) = \text{Support}(X \cup Y) / |D| \quad (1)$$

### 2.3 Confidence

The confidence of association rule is measure on the basis of the fraction of transactions in D that contain both X and Y. The confidence also shows the strength of the rule and represent by the following expression:

$$(X \Rightarrow Y) = \text{Support}(X \cup Y) / \text{Support}(X) \quad (2)$$

## 3. Other objectives for association rule mining

Other than Support and confidence, some other objectives such as Comprehensibility, Coverage, Cosine, lift, Laplace, Jaccard, J-measure, prevalence, surprise, recall, conviction, surprise and so on are also available in the literature[3, 8].

### 3.1 Comprehensibility

Comprehensibility of an association rule is quantified by the following expression:

$$\text{Comprehensibility} = \frac{\log(1+|Y|)}{\log(1+|X \cup Y|)} \quad (3)$$

where |itemset| means the number of attributes involved in the itemset.

### 3.2 Interestingness

One of the most important objectives of association rule mining is to extract some hidden information and interestingness. Interestingness is used to quantify how much the rule is surprising to the user. Interestingness can be represented by the following expression :

$$\text{Interistngness} = \left( \frac{\text{Support}(X \cup Y)}{\text{Support}(X)} \right) \times \left( \frac{\text{Support}(X \cup Y)}{\text{Support}(Y)} \right) \times \left( 1 - \frac{\text{Support}(X \cup Y)}{\text{Support}(Z)} \right) \quad (4)$$

where Support(Z) is the number of records in the database [9-10]. However, most researchers have adopted Piatetsky-Shapiro's [11] argument.

### 3.3 Lift

Lift compute the ratio between the rule's confidence and the support of the itemset in the rule

consequent. Lift is equivalent to the ratio of the observed support to that expected if X and Y were statistically independent.

$$Lift = \frac{Confidence(x \rightarrow y)}{Support(y)} \quad (5)$$

### 3.4 Interest factor

For binary variables, lift is equivalent to another objective called interest factor which is defined by the following expression :

$$I(A, B) = \frac{S(A,B)}{S(A) \times S(B)} = \frac{N f_{11}}{f_{(1+)} f_{(+1)}} \quad (6)$$

These parameters are explained in Table 2.

### 3.5 Correlation analysis

Correlation analysis is a statistical based technique for analyzing the relationship between a pair of variables. For continuous variable, correlation is defined using Person correlation coefficient. For binary variables correlation can be measured as –

$$\phi = \frac{f(11)f(00) - f(10)f(01)}{\sqrt{f(1+)f(+1)f(0+)f(+0)}} \quad (7)$$

The value of correlation range from -1(perfect negative correlation) to +1 (perfect positive correlation) [12]. If the variables are statistically independent than  $\phi = 0$ .

## 4. Drawback of the objectives

The drawback of association rules generated by using support and confidence objectives are, if the threshold value for support and confidence is low then it will generate too many rules. Whereas if the threshold value is high, it will generate few rules [13]. This resultant some important association rules may be missed. Sometimes association rules based on support and confidence, mislead us [14]. There is a lack of interestingness and comprehensiveness are also found in this kind of association rules. Thus goodness, effectiveness and interestingness of association rules cannot be evaluated only on the basis of support and confidence [15]. Almost every objective has some drawback and every objective is also not applicable to every situation [3].

## 4.1 Limitation of correlation analysis

The occurrence of word pair {P, Q} and word pair {R,S} in the same document, is given in below tables.

From the word association example given in Table 1, the shortcoming of correlation can be easily observe. In spite of co-occurrence of {P,Q} is more than {R,S}, the  $\phi$  -coefficient for word pairs {P,Q} and {R,S} are identical, i.e.  $\phi(P,Q) = \phi(R,S) = 0.232$ . Since  $\phi$ -coefficient confer the same weight to both co-presence and co-absence of items. Hence  $\phi$ -coefficient much appropriate for analyzing symmetric binary variables. If the sample size has been changed proportionately, the value of  $\phi$ -coefficient will remain same. This is another drawback of this measure.

## 5. High and low correlation objectives

For assessing the worthiness of the association rule, objectives are the only measure. A contingency table consists the frequency count, which may use to measure objectives.

Table 2 is an example of 2-dimensional contingency table for 2-variables, A and B. nA (Negative A) and nB (Negative B) indicate the absence from a transaction. The cells consist of f(i, j) represents the frequency count. Cell value f(1,1) represent the co-occurrence of A and B together whereas f(0,1) illustrate the absence of A but presence of B in transaction. Cell f(1,+) shows the support of A and cell f(+,1) demonstrate the support of B.

In this section, authors proposed 2 new objectives – High Correlation and Low Correlation, to calculate positive correlation and a negative correlation between items for 2 variables and 3 variables.

Table 1. Contingency table for the word pairs {P,Q} and {R,S}

	<b>P</b>	<b>nP</b>			<b>R</b>	<b>nR</b>	
<b>Q</b>	880	50	930	<b>S</b>	20	50	70
<b>nQ</b>	50	20	70	<b>nS</b>	50	880	930
	930	70	1000		70	930	1000

Table 2. A 2-dimensional contingency table for 2-variables

<b>Variable</b>	<b>B</b>	<b>nB</b>	<b>Sum</b>
<b>A</b>	f(1,1)	f(1,0)	f(1,+)
<b>nA</b>	f(0,1)	f(0,0)	f(0,+)
<b>Sum</b>	f(+,1)	f(+,0)	N

### 5.1 High and low correlation objectives for 2-variables

$\phi$ - coefficient gives equal importance to both co-presence and co-absence of items in transactions. It is, therefore, more suitable for analyzing symmetric binary variable. To overcome the drawback of correlation analysis, in this paper, we proposed two new objectives namely High Correlation ( $h\phi_2$ ) and Low Correlation ( $l\phi_2$ ) for 2-variables.

#### 5.1.1. High correlation for 2-variables

Analyzing the relationship between a pair of variables. It gives the importance to the co-presence of the variables. High Correlation compute as the ratio of difference between Support(A, B) and Support( A, nB), with square root of Support(A, -) and Support(-, B)

$$h\phi_2 = \frac{f(1,1)-f(1,0)}{\sqrt{f(1,+)f(+,1)}} \quad (8)$$

#### 5.1.2. Low correlation for 2-variables

Analyzing the relationship between a pair of variables. It gives the importance to the co-absence of the variables. Low Correlation compute as the ratio of difference between Support(nA, nB) and Support(nA, B), with square root of Support(nA, -) and Support(-, nB)

$$l\phi_2 = \frac{f(0,0)-f(0,1)}{\sqrt{f(0,+)f(+,0)}} \quad (9)$$

### 5.2 High and low correlation objectives for 3-variables

In the literature, most of the objectives are defined for the relation between 2 variables. In this paper, authors proposed 2 more new objectives which show the relation among 3 variables. To illustrate the concept, we are using a three-dimensional contingency table for A,B and C as shown below.

Table 3. A 3-Dimensional Table for 3-variables.

C	B	nB	
A	f(1,1,1)	f(1,0,1)	f(1,+,1)
nA	f(0,1,1)	f(0,0,1)	f(0,+,1)
	f(+,1,1)	f(+,0,1)	f(+,+,1)

nC	B	nB	
A	f(1,1,0)	f(1,0,0)	f(1,+,0)
nA	f(0,1,,0)	f(0,0,0)	f(0,+,0)
	f(+,1,0)	f(+,0,0)	f(+,+,0)

The cells of Table 3, represent the presence and/or absence of the items in a transaction. For example f(I,J,K) shows the particular combination of items A,B and C. The cell value f(1,0,1) illustrates the presence of A and C but the absence of B in number of transactions. Whereas the cell f(1,+,1) represent the presence of A and C, irrespective of presence or absence of B in number of the transaction.

#### 5.2.1. High correlation for 3-variables

Analyzing the relationship among 3 variables. It gives the importance to the co-presence of the variables.

$$h\phi_3 = \frac{f(1,1,1)-f(1,0,1)-f(1,1,0)-f(1,0,0)}{\sqrt{f(1,+,1)f(1,+,0)f(+,1,1)f(+,1,0)f(+,+,1)}} \quad (10)$$

#### 5.2.2. Low correlation for 3-variables

Analyzing relationship among 3 variables. It gives the importance to the co-absence of the variables.

$$l\phi_3 = \frac{f(0,0,0)-f(0,0,1)-f(0,1,0)-f(1,0,0)}{\sqrt{f(0,+,1)f(+,0,1)f(+,0,0)f(0,+,0)f(+,+,0)}} \quad (11)$$

## 6. Proposed algorithm

In this section we proposed an empirical novel algorithm- **hlcARM** for generating high and low correlation association rules. Traditionally, the association rules mining algorithms process divided into two phases – extracting frequent patterns and generating strong association rules. In **hlcARM**, we combine both the phases and generating the significant rules while analyzing high and low correlation within each candidate's itemset. This avoids calculating combinations of item redundantly. In the closing state, we get all the significant association rules that either for highly correlated items or low correlated items.

We calculate the high and low correlation between item combinations x and y by the formula derived by us. The value of high and low correlation decided that the rules fall into which category. The generated rules classify the items as highly correlated or low correlated even if the support is not higher than the minimum support.

Below figure describe the pseudo-code for hlcARM. In line 1, both sets of high and low correlated association's rules are set to empty. Initial scanning of the database and frequent 1-itemsets are generated in line 2. All frequent k-itemsets generated through iteration and stored in  $F_k$  (line 8).  $F_k$  is confirmed from a set of candidate item  $C_k$  in

line 4. The iteration which starts from line 2, will stop when no further itemsets are achievable.

In contrast with Apriori algorithm, this algorithm, join the frequent itemsets at level  $k-1$  with the frequent 1-itemset, as shown in line 4. This helps in analyzing the high and low correlation of more item combinations. In line 7, support for candidate items has been checked although we check the correlation even if the support is below than the minimum support, as in line 9 to 22.

In line 10 and 11, the high and low correlation for all generated candidate itemsets are computed. If the  $high\_corr$  is greater than the  $low\_corr$ , a high correlation association rule of the type  $x \Rightarrow y$  is generated, if the  $supp(x \cup y)$  is above the  $minsupp$  and the confidence is greater than  $minconf$  otherwise a low correlation association rule generated. Simultaneously if the support is less than minimum support and confidence is higher than minimum confidence and  $high\_corr$  is less than  $low\_corr$ , a low correlation association rule generated. At the end, a set of high and low correlated association rules will found.

**Algorithm** High and Low Correlative Association Rules Mining

**Input** TD,  $minsupp$ ,  $minconf$

**Output** AR: High and Low Correlative Association Rules.

**Method:**

- (1)  $hCAR \leftarrow \emptyset; lCAR \leftarrow \emptyset$  /\*high and low correlative Association Rule sets\*/
- (2) scan the database and find the set of frequent 1-itemsets ( $F_1$ )
- (3) **for** ( $k = 2$   $F_{k-1} \neq \emptyset$ ;  $k + +$ ) {
- (4)  $C_k = F_{k-1} \bowtie F_1$
- (5) **for each**  $i \in C_k$  {
- (6)  $s = Support(TD, i)$
- (7) **if** ( $s > minsupp$ ) **then**
- (8)  $F_k \leftarrow F_k \cup i$  /\*item  $i$  is added to  $F_k$ \*/
- (9) **for each**  $x, y (i = (x \cup y))$  {
- (10)  $high\_corr = h\phi_2(x, y)$  /\* calculate high correlation between  $x$  and  $y$ \*/
- (11)  $low\_corr = l\phi_2(x, y)$  /\* calculate low correlation between  $x$  and  $y$ \*/
- (12) **if** ( $s \geq minsupp$ ) **then**
- (13) **if** ( $confidence(x, y) \geq minconf$ ) **then**
- (14) **if** ( $high\_corr(x, y) > low\_corr(x, y)$ ) **then**
- (15)  $hCAR \leftarrow hCAR \cup \{x \rightarrow y\}$
- (16) **else**  $lCAR \leftarrow lCAR \cup \{x \rightarrow y\}$
- (17) **if** ( $s < minsupp$ ) **then**
- (18) **if** ( $confidence(x, y) \geq minconf$ ) **then**
- (19) **if** ( $high\_corr(x, y) < low\_corr(x, y)$ ) **then**

(20)  $lCAR \leftarrow lCAR \cup \{x \rightarrow y\}$

(21) **else**  $hCAR \leftarrow hCAR \cup \{x \rightarrow y\}$

(22) } }

(23) return  $\{hCAR, lCAR\}$

## 7. Related work

In the literature, various algorithms are proposed by the researchers to generate the association rules. These algorithms generated association mainly based on Apriori and Tree-based approach [2]. Apriori-based algorithms work in two phases, first to identify frequent patterns and second generate rules from these frequent patterns [7]. Association rule mining helpful in the various domain including e-commerce and time-series data analysis [19-20].

Various authors proposed the variation of association rule mining approach through positive and negative association rule [21-22] rare association rule [23] and association rules by using various objectives[18] including right interesting measures[24].

Wu et al. [21] proposed a method to which generate both positive and negative association rules. These association rules belong to the form of  $A \Rightarrow \neg B$ ,  $\neg A \Rightarrow B$  and  $\neg A \Rightarrow \neg B$  which indicate negative association between items. In their approach, authors takes minimum support, minimum confidence and minimum interestingness as parameters and after the first phase, received the set of frequent itemsets of interest and set of infrequent itemsets of interest. Authors used conditional-probability increment ratio function and in phase two, generate the positive and negative association rules. Authors claim their approach is effective, efficient and promising on the basis of the experimental results.

Antonie et al. [22] describe an algorithm that extends the support-confidence framework with a sliding correlation coefficient threshold. In this algorithm, for each generated itemset, compute all possible combinations of items to analyze their correlation. At last only those rules is considered which generated from item combination of strong correlation. Simultaneously, if the correlation between items is negative then the algorithm generated the negative correlation rules in which either the antecedent or consequent part is negated. Authors also perform experiments on real datasets and compare the performance with the existing algorithm and proved that their algorithm can discover strong patterns and all types of confined rules also.

Surana et al. [23] suggested a set of properties

Table 4. Performance of proposed algorithm on synthetic datasets on various parameters

Parameter / Data Sets	10 K	30 K	50 K	70 K	90 K
<b>Transactions Count</b>	120428	284284	475649	665470	855367
<b>Max Memory Usages (in MB)</b>	4.25	23.40	99.66	115.43	158.91
<b>Frequent Itemsets Count</b>	779	770	779	779	779
<b>Time to generate Frequent Patterns (in ms)</b>	390	484	484	796	968
<b>Number of Association Rules Generated</b>	266	255	255	256	256
<b>Time to Generate Association Rules(in ms)</b>	47	31	31	63	48

Table 5. Performance of proposed algorithm on real datasets on various parameters

Parameter / Data Sets	Dermatology	Marketing	Monk-2	Wisconsin	Shuttle
<b>Transactions Count</b>	358	6876	432	683	57999
<b>No. of Attributes</b>	34	13	6	9	9
<b>Max Memory Usages (in MB)</b>	2.8234	8.452	2.240	3.200	5.538
<b>Frequent Itemset Count</b>	11800	1642	39	96	5
<b>Time taken to generate Frequent Patterns</b>	603	340	41	76	380
<b>No. of Association Rules generated</b>	9052	3882	35	114	5
<b>Time taken to Generate Association Rules</b>	375	161	7	28	6
<b>No. of HCL Rule</b>	684	2644	11	81	1
<b>No. of LCL Rule</b>	8368	1238	24	33	4

that consider to find the interestingness of rare association rules. Authors discuss various interestingness measures. They also explain the interestingness measure as objective and subjective and further symmetric and asymmetric interestingness measure. Authors analyze how various interestingness measures perform in extracting rare association rules. Through experiments, authors demonstrate the similarity between various measures and measure selection for mining association rules. It is observed that a single measure may not be appropriate to mine interesting association rules.

Tan et al. [24] discussed 21 various objective measure from the diversified field including statistic, social science, machine learning and data mining to select the right interesting measure for a given application. It is identified that depending on its properties, some of the measures are useful in some application but not in all situation. Authors also proposed a framework for selecting a right measure. Authors exhibit support based pruning and table standardization technique. At last authors proposed RANDOM and DISJOINT to capture maximum conflict of ranking among different measures.

### 8. Experimental results

The experiments are conducted on Intel Core i3 processor, 4GB RAM and Windows 7 operating system based computer by using Eclipse Java EE

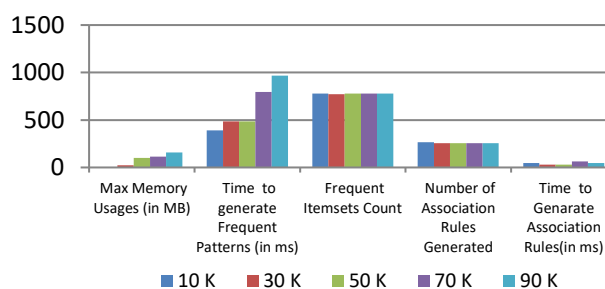


Figure. 1 Performance of proposed algorithm on various parameters

IDE for Web Developers, Version: Luna Service Release 2 (4.4.2). We conducted experiments on synthetic as well as real datasets, to study the performance of the algorithms. The synthetic datasets are 10k, 30k, 50k, 70k and 90k in size. These datasets are available at KEEL [25]. The performance of the algorithm on these synthetic datasets are as follows:

It is clear from the Fig. 1, that proposed algorithm scalable in terms of the time taken to generate frequent patterns. Also, since the number of frequent patterns are almost equal, then the proposed algorithm takes almost same amount of time to generate association rules.

Below Fig. 2, demonstrate the relationship between transaction count and memory uses. Simultaneously, Fig. 3 shows the relation between the time required to generate frequent patterns with the time taken to generate association rules.

The performance of the proposed algorithm is also evaluated on original datasets viz. Dermatology,

Marketing, Monk-2, Wisconsin and Shuttle dataset. These datasets are available at [25, 26] Performance of proposed algorithm on various parameters is given in Table 5.

Fig. 5 shows the performance of the proposed algorithm on time required to generate rules and number of rules generated. Fig. 6 demonstrate the comparison between number of frequent patterns and number of association rules generated. Fig. 7 illustrate the bifurcation between high correlation rules and low correlation rules.

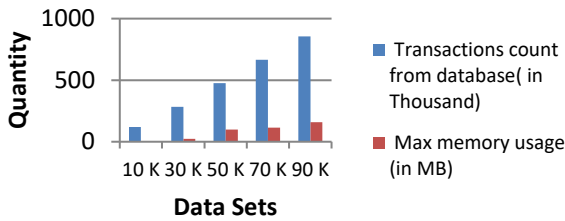


Figure. 2 Relation between transaction count and memory uses

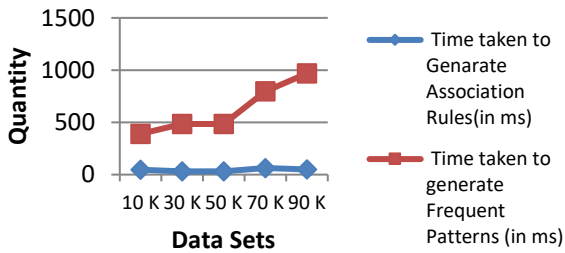


Figure. 3 Relation between time required to generate frequent patterns with time taken to generate association rules

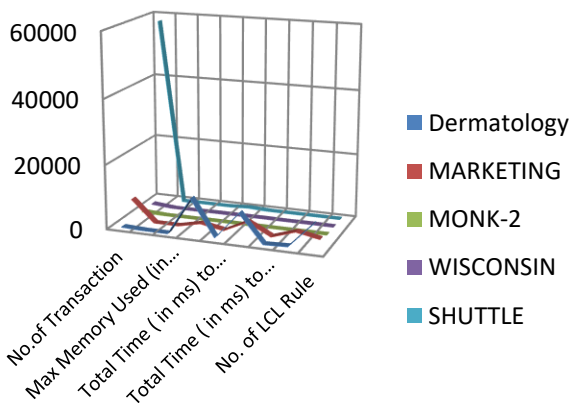


Figure.4 Performance of hlcARM on various parameters of real datasets

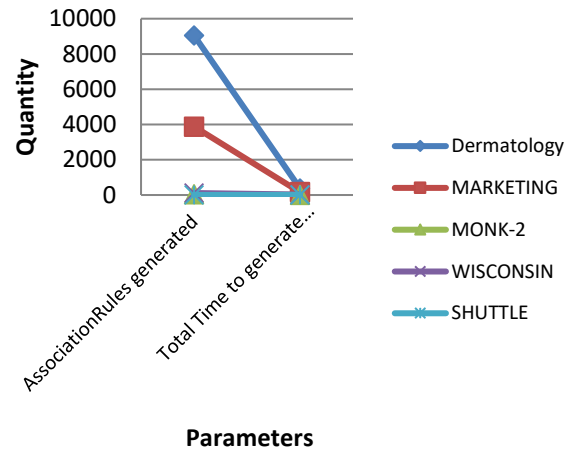


Figure.5 Performance of the proposed algorithm on time required to generate rules and no. of rules generated.

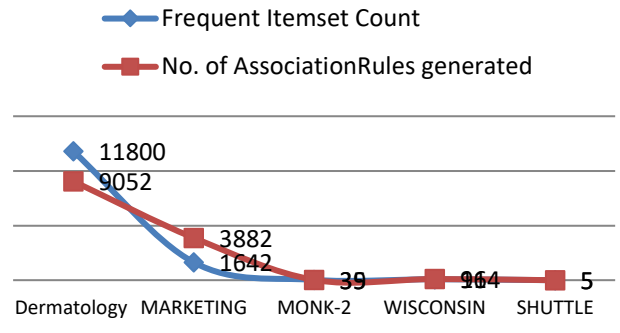


Figure. 6 Comparison between number of frequent patterns and number of association rules generated

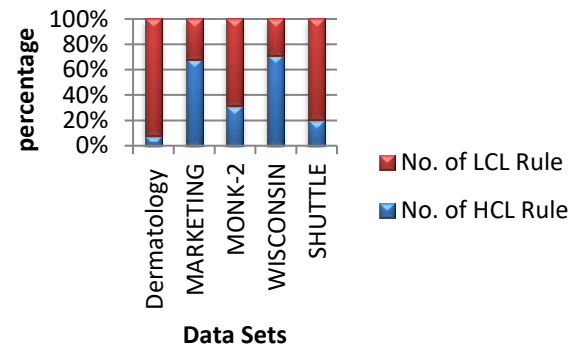


Figure. 7 Number of LCL and HCL rules generated through real datasets

Authors also compare the performance of their algorithm with a variation of an algorithm proposed by [27] as per the code available on [28] and found that the proposed algorithm gives promising results. Table 6 represents the performance comparison of proposed and existing algorithm.

Table 6. Performance comparison of hclARM and existing algorithm on various parameters

Data Sets / Parameters	Memory Consumption (in MB)		Time to generate Frequent Patterns (in ms)		Association Rules Generated	
	by Existing Algo.	by hclARM	by Existing Algo.	by hclARM	by existing Algo.	by hclARM
Dermatology	33.75	32.76	531	499	10052	9052
MARKETING	10.31	9.001	328	328	10820	3882
MONK-2	1.920	2.240	62	31	43	35
WISCONSIN	3.200	3.200	62	46	174	114
SHUTTLE	4.238	4.374	515	452	5	5

Fig. 8 shows the comparison between existing algorithm and hclARM algorithm on memory consumption. It is clear from the Table 6 and Fig. 8 as well, that the proposed algorithm consume less memory. Fig. 9 shows the comparison on the number of frequent patterns generated. It is clear from the results that the hclARM proposed less and only those frequent patterns which generate significant hcl and lcl association rules. Fig. 10 shows the comparison between association rules generated by existing algorithm and hclARM algorithm. The association rules generated by hclARM are interesting and gives proper knowledge that how much the generated rule is highly correlated or low correlated. It means how much the rule is important, significant and useful.

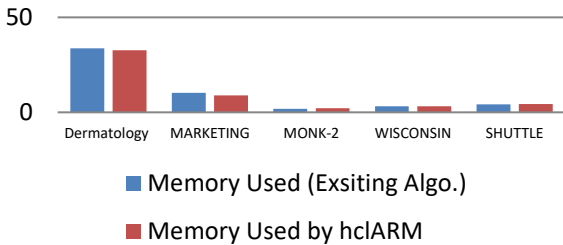


Figure.8 Comparison between old and hclARM on memory consumption

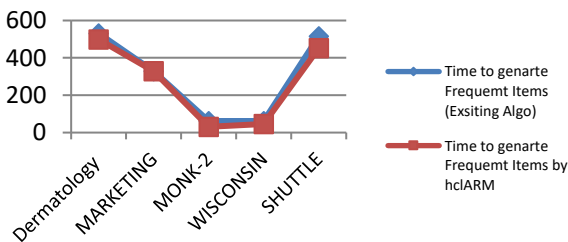


Figure.9 Comparison between old and hclARM on time to generate frequent patterns

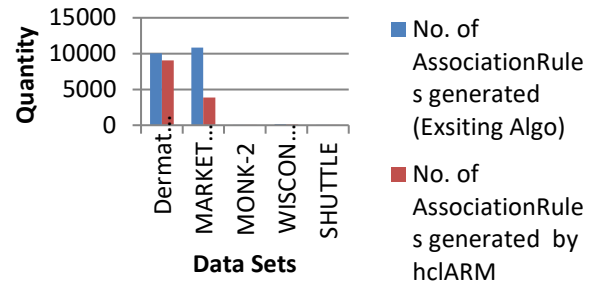


Figure. 10 Comparison between association rules generated by the existing and hclARM algorithm

### 9. Conclusion and future work

Proposed hclARM algorithm shows promising results in comparison of an existing algorithm. Its performance in terms of memory consumption is remarkable. This algorithm also generates less number of frequent patterns which is helpful in generating hcl and lcl rules. Simultaneously the proposed algorithm generates significant, interesting and actual correlation among items, which is very helpful in decision making and making business strategies. For future work some improvement may be done through the following:

- To reduce the number of scanning we may implement the evolutionary algorithm.
- New objectives can be identified to measure more interestingness and comprehensibility.
- Modify this algorithm which can be used for another category of association rule include categorical association rules and fuzzy association rules.
- Enhance this algorithm to support Incremental and interactive association rule mining.

### References

[1] R. Agrawal, T. Imielinski, and A. Swami, "Mining Association Rules between sets of Items in large Databases", In: *Proc. of ACM-SIGMOD Int. Conf. Management of Data*, pp. 207-216, 1993.

[2] M. Kumar and H. K. Soni, "A Comparative Study of Tree-based and Apriori-based



- Approaches for incremental Data Mining”, *International Journal of Engineering Research in Africa*, Vol. 23, pp. 120-130, 2016.
- [3] P.N. Tan, V Kumar, and J. Srivastava, “Selecting the right interestingness measure for association patterns”, In: *Proc. of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 32-41, 2002.
- [4] R. Agarwal and R. Srikant, “Fast Algorithm for Mining Association Rules”, In: *Proc. of Int. conf. Very Large Data Bases*, pp. 487-499, 1994.
- [5] M. Houtsma and A. Swami, “Set-Oriented Mining of Association Rules”, *Research Report RJ 9567, IBM Almaden Research Centre, San Jose, California*, 1993.
- [6] S. Kotsiantis and D. S. Kanellopoulos, “Association Rules Mining: A Recent Overview”, *International Transactions on Computer Science and Engineering*, Vol. 32, No. 1, pp. 71-82, 2006.
- [7] H. K. Soni, S. Sharma, and M. Jain, “Frequent Pattern Generation Algorithms for Association Rule Mining : Strength and Challenges”, In: *Proc. of IEEE International Conference on Electrical, Electronics and Optimization Techniques*, pp. 3744-3747, 2016.
- [8] E. B. Ayed and M. B. Ayed, “A quality model for the evaluation of decision support systems based on a knowledge discovery from data process”, *Journal of Decision Systems*, Vol. 25, No. 2, pp. 95-117, 2016.
- [9] H. R. Qodmanan, M. Nasiri, and B. M. Bidgoli. “Multi objective association rule mining with genetic algorithm without specifying minimum support and minimum confidence”, *Expert Systems with Applications*, Vol. 38, pp. 288–298, 2011.
- [10] L. Davis, (Ed.). *Handbook of genetic algorithm*. New York: Van Nostrand Reinhold. 1991.
- [11] G. Piatetsky-Shapiro, *Discovery, analysis, and presentation of strong rules*. In G. Piatetsky-Shapiro, W. Frawley (eds.) *Knowledge Discovery in Databases*, AAAI/MIT Press, pp. 229–248, 1991.
- [12] B. Xiao, “The measures and mining method of credible association rule”, In: *Proc. of IEEE International Conference on Network*, pp. 322-325, 2009.
- [13] J. Han and M. Kamber, *Data mining concepts and techniques*. Morgan Kaufmann, 2001.
- [14] P. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Addison-Wesley, 2005.
- [15] A. Mukhopadhyay, U. Maulik, S. Bandyopadhyay, and C. A. Coello, “Survey of multi-objective evolutionary algorithms for data mining: Part-II”, *IEEE Transactions on Evolutionary Computation*, Vol. 18 , No. 1, pp. 20-35, 2014.
- [16] S. Brin, R. Motwani, and C. Silverstein, “Beyond market baskets: Generalizing association rules to correlations”, In: *Proc. of SIGMOD Conference*, pp. 265–276, 1997.
- [17] X. Wu, C. Zhang, and S. Zhang, “Efficient mining of both positive and negative association rules”, *ACM Trans. Inf. Syst.*, Vol. 22, No. 3, pp. 381-405, 2004.
- [18] H. K. Soni, S. Sharma, and A.K. Upadhyay, “Two novel pioneer objectives of association rule mining for high and low correlation of 2-variables and 3-variables”, *International Journal of Engineering and Technology*, Vol. 9, No. 2, pp. 695-703, 2017.
- [19] H. K. Soni and S. Sharma, “Plausible Characteristics of Association Rule Mining Algorithms for E-Commerce”, In: *Proc. of the 3<sup>rd</sup> International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics*, pp. 36-39, 2017.
- [20] N. Mishra, H. K. Soni, and S. Sharma, “A Comprehensive Survey of Data Mining Techniques on Time Series Data for Rainfall Prediction”, *Journal of ICT Research and Applications*, Vol.11, No.2, pp. 167-183, 2017.
- [21] X. Wo, C. Zhang, and S. Zhang, “Efficient Mining of Both Positive and Negative Association Rules”, *ACM Transactions on Information Systems*, Vol. 22, No. 3, pp. 381–405, 2004.
- [22] M.L. Antonie and O.R. Zaiane, “Mining Positive and Negative Association Rules: An Approach for Confined Rules”, *Knowledge Discovery in Databases: PKDD*, pp. 27-38, 2004.
- [23] A. Surana, R.U. Kiran, and P.K. Reddy, “Selecting a Right Interestingness Measure for Rare Association Rules”, In: *Proc. of the 15<sup>th</sup> International Conference on Management of Data*, pp. 115-124, 2010.
- [24] P. N. Tan, V. Kumar, and J. Srivastava. “Selecting the right objective measure for association analysis”, *Information System*, Vol. 29, pp. 293-313, 2004.
- [25] <http://sci2s.ugr.es/keel/dataset.php>
- [26] <https://archive.ics.uci.edu/ml/datasets>
- [27] J. Han, J. Pei, Y. Yin, and R. Mao, “Mining frequent patterns without candidate generation: *International Journal of Intelligent Engineering and Systems*, Vol.11, No.3, 2018

A frequent-pattern tree approach”, *Data Mining and Knowledge Discovery*, Vol. 8, No. 1, pp. 53-87, 2004.

[28] <http://www.philippe-fourrier-viger.com/spmf/documentation.php#lift>