

# YAMI: Incremental Mining of Interesting Association Patterns

Eiad Yafi<sup>1</sup>, Ahmed Al-Hegami<sup>2</sup>, Afshar Alam<sup>1</sup>, and Ranjit Biswas<sup>3</sup>

<sup>1</sup>Department of Computer Science, Jamia Hamdard University, India

<sup>2</sup>Faculty of Computers and Information Technology, Sana'a University, Yemen

<sup>3</sup>Department of Computer Engineering, Jadavpur University, India

**Abstract:** Association rules are an important problem in data mining. Massively increasing volume of data in real life databases has motivated researchers to design novel and incremental algorithms for association rules mining. In this paper, we propose an incremental association rules mining algorithm that integrates shocking interestingness criterion during the process of building the model. A new interesting measure called shocking measure is introduced. One of the main features of the proposed approach is to capture the user background knowledge, which is monotonically augmented. The incremental model that reflects the changing data and the user beliefs is attractive in order to make the over all Knowledge Discovery in Databases (KDD) process more effective and efficient. We implemented the proposed approach and experiment it with some public datasets and found the results quite promising.

**Keywords:** KDD, data mining, incremental association rules, domain knowledge, interestingness, shocking rules.

Received February 4, 2010; accepted August 10, 2010

## 1. Introduction

Association rule mining is one of the most important techniques of data mining. It was first introduced in [1]. It aims to extract interesting correlations, frequent patterns, associations among sets of items in the transaction databases. The task of association rules mining usually performed in a two step process. The first step aims at finding all frequent itemsets that satisfy the minimum support constraint. The second step involves generating association rules that satisfy the minimum confidence constraint from the frequent itemsets. Since finding the frequent itemsets is of great computational complexity, the problem of mining association rules can be reduced to the problem of finding frequent itemsets.

One of the main drawbacks with the classical association rules algorithms is that they do not consider the time in which the data arrive. In practice, data is acquired in small batches over the time. In such scenario a combination of old and new data is used to build a new model from scratch.

As time advances, some old transactions may become obsolete and thus are discarded from the database. Consequently, some Previously Discovered Knowledge (PDK) becomes invalid while some other new rules may show up. Researchers therefore have been strongly motivated to propose techniques that update the association rule model as new data arrives, rather than running the algorithms from scratch [9, 11, 12], resulting in incremental models. Incremental algorithms build and refine the model as new data

arrive at different points in time, in contrast to the traditional algorithms where they perform model building in batch manner [11, 12]. The incremental association rules algorithms that reflect the changing data trends and the user beliefs are attractive in order to make the overall KDD process more effective and efficient.

We propose an incremental algorithm, YAMI (YAMI is derived from the names of the authors), based on the premise that unless the underlying data generation process has changed dramatically, it is expected that the rules discovered from one set are likely to be similar (in varying degrees) to those discovered from another set [9]. Interesting measures can be used as an effective way to filter the rule set discovered from the target data set thereby, reducing the volume of the output. Our work extends the approaches presented in [13, 18] and integrates it into an association rule algorithm in an incremental manner. The proposed approach is a self-upgrading filter that keeps known knowledge PDK and the user Domain Knowledge (DK) rule base updated as new shocking rules discovered. The shocking interestingness presented in [18] is quantified on the basis of determining significant attributes and then the degree of Shocking Rules (SHR) of the newly discovered rules with respect to the known knowledge. The idea of shocking rules came from the latest disasters which have encountered the world recently, such as the increasing number of earthquakes, tornados and Tsunami waves. The more interesting rules are those which are unexpected and novel as well, so

shocking rules have the highest degree of interestingness. They are novel since they do not exist in the PDK, unexpected as they have the highest degree of significant attributes that indicates SHR with respect to the rules in PDK and actionable as they enable the decision maker to make actions to their advantages. A rule is shocking if it overthrows all the expectations of the user. It's unprecedented, never expected and happens suddenly in a way that it shocks the user and put him in an unenviable situation.

The proposed incremental association rule algorithm operates on the incremental training set and builds a model. During the frequent itemsets generation, the algorithm computes the shocking interesting measure against the known knowledge and prunes the items that do not meet the user interest. A detailed description of computation of shocking interestingness measure can be found in [18]. Further iteration of the algorithm is performed only for the frequent itemsets that have shocking interestingness measure higher than a user specified threshold. This is a useful feature in which the user may need to trade off some accuracy for shocking interestingness that may arise in some domains where the user wants a rough picture about the domain rather than an optimal model that contains a lot of details. The incremental nature of the proposed algorithm makes it advantageous to discover shocking patterns at current time with respect to the previously discovered patterns (rules), rather than exhaustively discovering all patterns.

## 2. Related Work

Several approaches have been proposed for developing incremental algorithms of association rules mining [6, 7, 8, 10, 15, 20] and mining of the frequent item sets [6, 7, 8]. The main assumption of these approaches is to update the discovered model when new data arrive. In [10], DEMON algorithm is proposed that works effectively and efficiently with evolving data over the time [15]. Proposed algorithm for monitoring the changes in the data stream environments [16]. Another incremental algorithm, Lee and Cheung uses statistical methods for updating process with DELI algorithm. DELI algorithm applies a sampling technique to estimate the support counts using an approximate upper/lower bounds on the amount of changes in the set of newly introduced association rules. A low bound would mean that changes in association rules are small and there should be no maintenance. In addition, these algorithms are incremental in nature as they use and reuse the previously discovered knowledge and integrated when new data occur. Zaki *et al.* [20] designed a parallel approach that reduces the computational requirement in the algorithm. Cheung *et al.* [7, 8] proposed Fast Update (FUP) algorithm for the maintenance of discovered association rules in large databases. They have proposed to handle incremental

database by scanning the database to check whether there are large itemsets or not. FUP algorithm is introduced for quantifying the large itemsets in the updated database. The goal of this algorithm is to solve the efficient update problem of association rule in updated database. They have extended their work to FUP\* and FUP2 to k-pass algorithms, they scan the database  $k^{\text{th}}$  time.

In this paper, we propose a new approach to dynamic mining association rules in evolving databases. The proposed approach integrates the shocking interestingness criterion [18] during the model building process to form a constraint in order only to discover shocking rules.

The rest of the paper is organized as follows. The problem statements are given in section 3. Section 4 presents the shocking interestingness measure used in the proposed algorithm. In section 5, our proposed approach for incremental generation of association rules is introduced. Section 6 presents the algorithm used in the proposed approach. Section 7 shows the effectiveness of the proposed approach through a detailed example. Experimental results and implementation are presented in section 8. Finally, section 9 contains the conclusion.

## 3. Problem Statement

Given a dataset  $D$  collected over the time  $[t_0, t_1, t_2, \dots, t_n]$ . At  $t_0$ ,  $D_0$  represents an empty database. At time instance  $t_i$ , an incremental dataset  $D_i$ ,  $i \in \{1, \dots, n\}$ , is collected such that  $D = D_1 \cup D_2 \cup \dots \cup D_i$ . Let  $T_i$  and  $T_{i+1}$  be two models discovered at time instances  $t_i$  and  $t_{i+1}$  from datasets  $\bigcup_{j=1}^i D_j$  and  $\bigcup_{j=1}^{i+1} D_j$  respectively. The major

volume of discovered rules in  $T_i$  and  $T_{i+1}$  would be similar to some extent. A small set of rules, which are either present or absent in  $T_{i+1}$  represents change in data characteristics. The objective is to update  $T_i$  to  $T_{i+1}$  using  $D_{i+1}$  and  $T_i$ .  $T_i$  the model discovered at time  $t_i$  now represents PDK.  $T_{i+1}$  is the up-to-date model obtained by adding interesting rules discovered from  $D_{i+1}$ . This is achieved by constructing a model  $T_{i+1}$  from  $D_{i+1}$  such that association rules in  $T_{i+1}$  have user specified degree of shocking interestingness with respect to the rules in  $T_i$ . Subsequently,  $T_{i+1}$  is used to update  $T_i$  to  $T_{i+1}$ .

## 4. Shocking Interestingness Measure

The shocking interestingness of a rule is quantified by computing the significant attributes and then the degree of SHR of the antecedent and the consequent at conjunct level and subsequently the significance at conjunct level is combined to compute the significance at rule level. A high significant attributes indicates a high degree of SHR. The significant attribute is computed by measuring the deviation for the

antecedent and the consequent at conjunct level and subsequently, combine the conjunct level deviation to compute rule level deviation. The detailed description of our proposed approach to quantify shocking interestingness can be found in [18].

In this work we integrate the approach proposed in [18] into the mining association rule algorithm. The shocking measure is pushed into the classical Apriori algorithm to form a constraint in order to discover only shocking interesting patterns. The shocking measure is applied at each iteration of the frequent itemsets generation. For each itemset, a set of rules are extracted to form strong partial rules. The strong partial rules are those rules with  $\min\_confidence$ . These partial rules are subjected to the shocking interestingness criterion resulting in computation of the degree of shocking interestingness which is assigned to each partial rule. The partial rules are shocking if the degree of their interestingness is higher than a user interestingness threshold value.

The process of frequent itemsets generation is further continuing taking into account the rules that are only interesting and pruning the itemsets that are uninteresting. This strategy ensures that only shocking interesting itemsets are eligible to be candidate during the next iteration of frequent itemsets generation.

## 5. Incremental Mining of Association Rules

The proposed algorithm uses shocking interestingness measure as a constraint during the model building in order to discover association rules that are shocking (interesting) for the user. The advantage of pushing such constraints inside the algorithm is that the search space is reduced and the algorithm discovers relatively smaller sized model. Further, the discovered knowledge reflects the user's requirement of interestingness.

One of the main features of the proposed algorithm is to deal with time changing data and user beliefs. This is a useful functionality in situations when two datasets have arrived at different points of time or from different geographical locations. Certainly, it is attractive to update the discovered knowledge each time new data arrive.

For every stage of frequent itemsets generation, partial rules<sup>1</sup> are generated from the frequent itemsets at that stage. These partial rules are evaluated using confidence measure and prune the partial rules that do not satisfy this measure resulting in a set of strong partial association rules. The strong rules are subjected to the shocking criterion [18] in order to decide either these rules are interesting or not. The Shocking Interestingness (SI) of a partial rule  $P$  is computed with

<sup>1</sup>We use the term partial rule because the complete association rule has not been generated yet, and addition of items is certain to be added to the rules.

respect to the closest rule  $R$  in the existing model  $T_i$ . The  $SI_P^R=0$ , indicates that the partial rule if expanded is likely to create a rule which is already present in  $T_i$ . If the partial rule is found to be shocking interesting i.e.,  $SI_P^R=1$ , the algorithm expands the current frequent itemsets to next level frequent itemsets like a normal Apriori algorithm. In case a partial rule is found to be uninteresting, it computes the Significance Factor (SF) of the partial rule  $P$  to decide whether the partial rule is to be expanded further or not. Computation of SF of  $P$  indicates the relevance of rule  $R$  with respect to current training set  $(D_i+I)$ . If the SF is acceptable, the partial rule is not expanded further.

Figure 1 shows the environment in which the proposed algorithm operates. At time  $t_i+1$ , database  $D_i+1$  is pre-processed and subjected to the algorithm. The algorithm takes into account the existing model  $T_i$  representing the known association rules. The algorithm expands only those frequent itemsets that are likely to lead to shocking interesting rules. This results into discovering of  $T_i+1$ . For each frequent itemsets, a rule is extracted and used to update the model  $T_i+1$ .

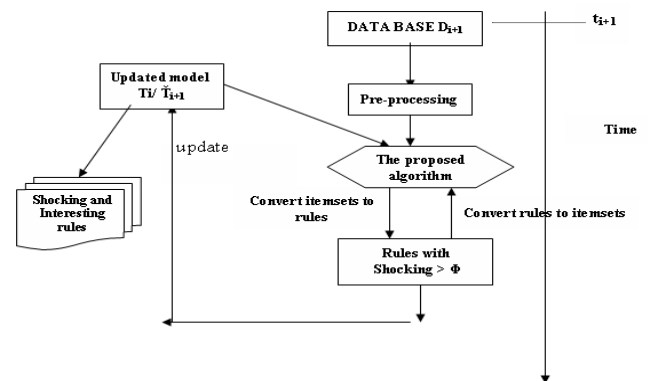


Figure 1. Operating environment of the proposed algorithm.

a. *Frequent Itemsets Generation*: The proposed algorithm is based on the generic Apriori methodology [1]. It dynamically decides whether an itemset is to be used in the next iteration of candidate generation or not taking into account the partial rule that is obtained by converting the frequent itemsets, which have confidence higher than confidence threshold value, into rules. At each frequent itemsets, the following tasks are performed:

1. Extracting the partial rules which have confidence higher than the confidence threshold value.
2. Computation of SI of the partial rules  $P$  with respect to the existing model  $T_i$ .
3. Computation of SF of the partial rule  $P$  with  $SF_P^R = 0$ .

The process of candidate generation is continued as the tradition Apriori algorithm taking into account the only shocking interesting frequent items.

b. *Dynamic Pruning Based on Shocking Interestingness*: The characteristic feature of the approach is its ability to facilitate dynamic pruning based on shocking interestingness [18]. The objective is to reduce the size (complexity) of the frequent itemsets generation with assurance that the resulting rules does not compromise in terms of accuracy and provides the user with shocking interesting association rules. The algorithm computes SI at each iteration of frequent itemsets generation to determine whether an itemset is likely to lead to an interesting rule or not. An itemset becomes a candidate for next level frequent itemsets generation if its SI value is one or the significance factor of the closest rule in  $T_i$  is less than the significance factor threshold value. An interestingness value of 1, of the partial rule indicates that this rule is unlikely to expand to any existing association rule. A SI value of 0 of the partial rule indicates that the partial rule is likely to expand to some existing association rule. The threshold used to compute shocking interestingness can be set dynamically according to the user requirement. This flexibility to dynamically change the threshold is a useful feature in a situation where the user has a rough picture about the domain and is in learning phase.

c. *SF of a Partial Rule*: The algorithm computes the SF at each partial rule P with  $SI_P^R=0$ , to judge significance of the rule in the current training set  $D_i$ . The computation of SF is required in order to be assured that the expected expansion of the current partial rule is significant in the current increment  $D_{i+1}$  of the database. A higher significance factor of the expected expansion of a partial rule with respect to the training data indicates that the rule is still significant at current time  $t_{i+1}$ . A smaller significance factor, on the other hand, indicates that the expected expansion of the partial rule is now obsolete and is not valid in this dataset. As a result, the itemset must be further expanded and may lead to an interesting rule. The significance factor computation is done using the following methodology. Given a rule  $A \rightarrow R$ , the subset of the training set corresponding to A is called cover of A ( $\Gamma A$ ). The SF of  $A \rightarrow R$  is given as follows:

$$SF(A \rightarrow R) = \frac{|\Gamma(A \cap R)|}{|\Gamma A|} \quad (1)$$

Where  $|\Gamma(A \cap R)|$  denotes the number of tuples that contain both A and the class R, and  $|\Gamma A|$  is the number of tuples that contain antecedent A. Let RP be the partial rule and  $RS(A \rightarrow R)$  be the closest rule in  $T_i$  such that  $SI(R^P, R^S)=0$ . Then having computed the significance factor of the partial rule, the algorithm expands the itemset if the significance

factor is lower than the specified significance factor threshold value and stops expanding otherwise.

$$SF(R^P) = \frac{|\Gamma(A \cap R)|}{|\Gamma A|} \quad (2)$$

## 6. Algorithm

In this paper, we present a new algorithm that efficiently generates incremental association rules in the updated database by applying the Apriori property. The problem of mining association rules is decomposed into two sub-problems. Firstly, generating all large itemsets in the database and secondly, generating association rules according to the large itemsets generated in the first step. Figure 2 shows that traditional apriori algorithm proposed by [1].

```

• Find the frequent itemsets: the sets of items that have
  minimum
  support:
  Ck: Candidate itemset of size k
  Lk: frequent itemset of size k
  L1 = {frequent items};
  for (k = 1; Lk ≠ ∅; k++) do begin
    Ck+1 = candidates generated from Lk;
    for each transaction t in database do
      increment the count of all candidates in Ck+1 that are
      contained in t
    Lk+1 = candidates in Ck+1 with min_support
  end
  return ∪k Lk;
• Use the frequent itemsets to generate association rules.

```

Figure 2. The pseudo-code of apriori algorithm.

Our proposed incremental algorithm is similar to Apriori algorithm except that after each frequent itemsets generation, the shocking interestingness measure is computed with respect the existing model  $T_i$  and prune uninteresting items that are not significant in the current training set. Figure 3 shows the proposed algorithm.

```

• Find the frequent itemsets: the sets of items that have
  minimum support:
  Ck: Candidate itemset of size k
  Lk: frequent itemset of size k
  L1 = {frequent items};
  for (k = 1; Lk ≠ ∅; k++) do begin
    Ck+1 = candidates generated from Lk;
    for each transaction t in database do
      increment the count of all candidates in Ck+1 that are contained in t
    Lk+1 = candidates in Ck+1 with min_support.
    Extract partial rules Pi from Lk+1 with min_support
    Compute the shocking interestingness of Pi with respect to KK
    For each  $SJ_{P_i}^{KK} < \Phi$ 
      Compute Significance Factor (SF)
      If SF > SF threshold
        Delete Pi from Lk+1
    end
  end
  return ∪k Lk;
• Use the frequent itemsets to generate association rules.

```

Figure 3. The pseudo-code of the proposed algorithm.

### 7. Detailed Example

Given the transaction database  $D_i$ , the traditional Apriori algorithm generates the frequent itemsets as shown in Figure 4. Subsequently extract the rules from these frequent itemsets. Table 1 shows the corresponding set of discovered association rules assuming that the confidence threshold value is 0.6. We consider these rules as previously discovered rules PDK at time  $t_i$ . Figure 5 presents the process of the proposed incremental mining of patterns from transaction database ( $D_{i+1}$ ).

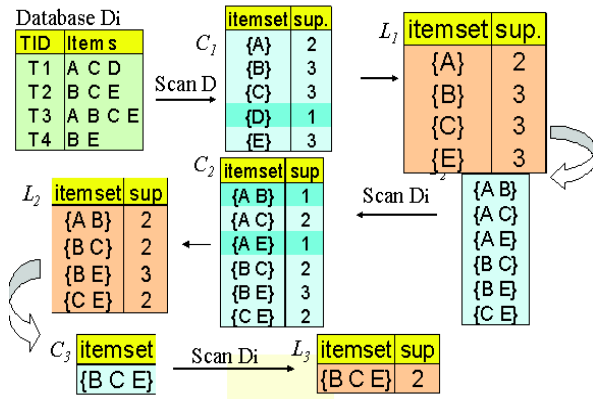


Figure 4. A priori mining process.

Table 1. The discovered association rules at time  $T_1$ .

Rule No.	Association Rules	Confidence
R1	$B \rightarrow C$	$[2/3=0.67]$
R2	$C \rightarrow B$	$[2/3=0.67]$
R3	$B \rightarrow E$	$[3/3=1]$
R4	$E \rightarrow B$	$[3/3=1]$
R5	$C \rightarrow E$	$[2/3=0.67]$
R6	$E \rightarrow C$	$[2/3=0.67]$
R7	$B \rightarrow CE$	$[2/3=0.67]$
R8	$CE \rightarrow B$	$[2/2=1]$
R9	$C \rightarrow BE$	$[2/3=0.67]$
R10	$BE \rightarrow C$	$[2/3=0.67]$
R11	$E \rightarrow BC$	$[2/3=0.67]$
R12	$BC \rightarrow E$	$[2/2=1]$

### 8. Experimental Studies

The proposed approach is implemented and tested using several public datasets available at <http://kdd.ics.uci.edu>. The approach is implemented using c programming language. The datasets are partitioned into three groups representing instances arrive at time  $T_1$ , time  $T_2$  and  $T_3$  respectively with 0.1% and 1% to indicate minimum confidence and minimum support respectively. The proposed approach uses the discovered rules to find out those rules which are interesting and which are conforming. The conforming rules are the rules that confirm the user background knowledge and hence are not interesting.

- *Experiment 1:* The objective of the first experiment is to show the effectiveness of our approach in reducing the number of discovered rules. It is expected that the number of discovered rules keeps on decreasing over the time. We work with five datasets and assume that the shocking interestingness threshold value  $SI=0.6$ . We used WEKA system to discover the patterns. WEKA is a collection of machine learning algorithms for data mining tasks. It is an open source software issued under the GNU General Public License [2]. The values in the third column of Table 2 represent the number of rules discovered, using WEKA system, at a given partition and the values in the fourth column represent the shocking interesting rules discovered by our approach. It is observed that the number of interesting rules decreases in contrast to the number of conforming rules which increases each time new rules discovered as per our expectation. Intuitively, the interesting rules discovered at time  $T_1$  is no more interesting at time  $T_2$ . The conforming rules are at the last column of Table 2.

Table 2. The discovered medical rules at time  $T_1, T_2$ , and  $T_3$ .

Dataset	Time	Discovered AR's	Interesting Rules	Conforming Rules
Lymph	$T_1$	32000	18230	13770
	$T_2$	28562	12003	16559
	$T_3$	26781	2010	24771
Breast	$T_1$	802	320	482
	$T_2$	725	180	545
	$T_3$	540	73	467
Hepatitis	$T_1$	1207	800	407
	$T_2$	980	430	550
	$T_3$	626	228	398
Heart	$T_1$	987	564	423
	$T_2$	566	320	246
	$T_3$	207	118	89
Sick	$T_1$	4502	2876	1635
	$T_2$	2709	1078	1631
	$T_3$	986	401	585

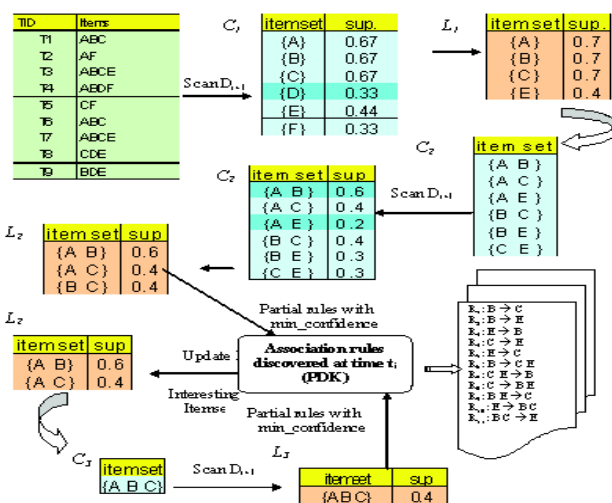


Figure 5. The proposed algorithm mining process from transaction database ( $D_{i+1}$ ).

- *Experiment 2:* The second experiment was performed using Sick, Zoo and Monk datasets available at <http://kdd.ics.uci.edu/>. We considered these datasets as evolving with time, and partitioned them into three increments:  $D_1, D_2$  and  $D_3$  assumed

to have arrived at times  $t_1$ ,  $t_2$  and  $t_3$  respectively. We run our proposed algorithm against the three partitions of each datasets to compute the overall accuracy of the models. It has been shown that the accuracy is increased as the number of training instances increases. Figure 6(a, b, c) shows the change in the accuracy of the generated rules using Sick, Zoo and Monk datasets respectively.

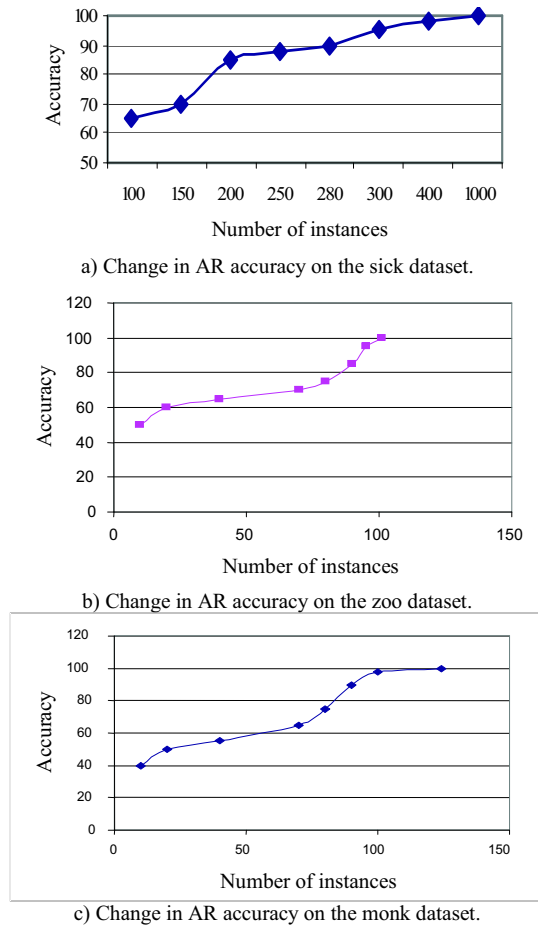


Figure 6. Sick, zoo and monk datasets experiment.

## 9. Conclusions

In this paper, we have proposed an algorithm for incremental association rules mining that integrates shocking interestingness criterion during the process of building the model. A new interesting measure called shocking measure is introduced. One of the main features of the proposed approach is to capture the user background knowledge, which is monotonically augmented. The proposed approach is a self-upgrading filter that utilizes interestingness criterion to reflect the user subjectivity and extract patterns, incrementally, from datasets arrive at different points in time. The proposed algorithm makes use of interestingness measure as the basis of extracting interesting patterns. This important feature of the proposed algorithm is attractive and desirable in many real life applications as the volume of data keeps on growing and changing over the time and therefore the user background

knowledge is monotonically augmented. This changing environment updates the user understandability and comprehensibility about the domain.

## References

- [1] Agrawal R., Imielinski T., and Swami A., "Mining Association Rules between Sets of Items in Large Databases," in *Proceedings of the International Conference on Management of Data Washington, USA*, pp. 207-216, 1993.
- [2] Al-hegami A., "Pruning Based Interestingness of Mined Classification Patterns," *The International Arab Journal of Information Technology*, vol. 6, no. 4, pp. 336-343, 2009.
- [3] Bronchi F., Giannotti F., Mazzanti A., and Pedreschi D., "Adaptive Constraint Pushing in Frequent Pattern Mining," in *Proceedings of the 17<sup>th</sup> European Conference on Practice of Knowledge Discovery in Databases*, Berlin, pp. 47-58, 2003.
- [4] Bronchi F., Giannotti F., Mazzanti A., and Pedreschi D., "Examiner: Optimized Level-Wise Frequent Pattern Mining with Monotone Constraints," in *Proceedings of the 3<sup>rd</sup> IEEE International Conference on Data Mining*, pp. 11-18, 2003.
- [5] Bronchi F., Giannotti F., Mazzanti A., and Pedreschi D., "Exante: Anticipated Data Reduction in Constrained Pattern Mining," in *Proceedings of the 7<sup>th</sup> PAKDD*, Berlin, pp. 59-70, 2003.
- [6] Cheung D., Han J., Ng V., and Wong C., "Maintenance of Discovered Association Rules in Large Databases: An Incremental Updating Technique," in *Proceedings of the International Conference on Data Engineering*, LA, pp. 106-114, 1996.
- [7] Cheung D., Ng V., and Tam B., "Maintenance of Discovered Knowledge: A Case in Multi-Level Association Rules," in *Proceedings of the 2<sup>nd</sup> International Conference on Knowledge Discovery and Data Mining*, pp. 307-310, 1996.
- [8] Cheung D., Lee S., and Kao B., "A General Incremental Technique for Mining Discovered Association Rules," in *Proceedings of International Conference on Database System for Advanced Applications*, pp. 185-194, 1997.
- [9] Dunham M., *Data Mining: Introductory and Advanced Topics*, Pearson Education, Singapore, 2003.
- [10] Ganti V., Gehrke J., and Ramakrishnan R., "Demon: Mining and Monitoring Evolving Data," in *Proceedings of the 16<sup>th</sup> IEEE International Conference on Data Engineering*, USA, pp. 50-63, 2000.
- [11] Han J. and Kamber M., *Data Mining: Concepts and Techniques*, Kauffmann Publishers, 2001.



- [12] Hand D., Mannila H., and Smyth P., *Principles of Data Mining*, Prentice-Hall of India Private Limited, 2001.
- [13] Kaur H., Wasan S., Al-hegami A., and Bhatnagar V., "A Unified Approach for Discovery of Interesting Association Rules," in *Proceedings of Industrial Conference on Data Mining*, Berlin, pp. 53-63, 2006.
- [14] Klemetinen M., Mannila H., Ronkainen P., Toivonen H., and Verkamo A., "Finding Interesting Rules from Large Sets of Discovered Association Rules," in *Proceedings of the 3<sup>rd</sup> International Conference on Information and Knowledge Management*, Gaithersburg, pp. 1-9, 1994.
- [15] Lee S. and Cheung D., "Maintenance of Discovered Association Rules. When to Update?" in *Proceedings of Research Issues on Data Mining and Knowledge Discovery*, pp. 51-58, 1997.
- [16] Liu B., Hsu W., Chen S., and Ma Y., "Analyzing the Subjective Interestingness of Association Rules," *IEEE Intelligent Systems*, vol. 15, no. 5, pp. 47-55, 2000.
- [17] Psaila G., "Discovery of Association Rules Meta-Patterns," in *Proceedings of the 2<sup>nd</sup> International Conference on Data Warehousing and Knowledge Discovery*, pp. 219-228, 1999.
- [18] Yafi E., Alam A., and Biswas R., "Development of Subjective Measures of Interestingness: From Unexpectedness to Shocking," in *Proceedings of World Academy of Science, Engineering and Technology*, pp. 368-370, 2007.
- [19] Waikato, available at: <http://www.cs.waikato.ac.nz/ml/weka/>, last visited 2011.
- [20] Zaki M. and Hsiao C., "Charm: An Efficient Algorithm for Closed Itemset Mining," in *Proceedings of the 2<sup>nd</sup> SIAM International Conference on Data Mining*, USA, pp. 457-473, 2002.



**Ahmed Al-hegami** is an assistant professor at the Faculty of Computers and Information Technology, Sana'a University, Yemen. His research interests include artificial intelligence, machine learning, temporal databases, real time systems, data mining, and knowledge discovery in databases.



**Afshar Alam** is a professor and Head of the Department of Computer Science, Jamia Hamdard University, New Delhi, India.



**Ranjit Biswas** is a professor at ITM University. His area of specialisation is fuzzy Logic in computer engineering, rough theory and approximate reasoning, bag database, fuzzy database, parallel architecture, artificial intelligence, and fuzzy pattern recognition.



**Eiad Yafi** is a research scholar from Syria, working on his PhD in data mining at the Department of Computer Science, Jamia Hamdard University, India. He received his MSc of computer applications from Jamia Hamdard University in 2004.