ISSN:0975-9646

Arnab kumar Das / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (6) , 2015, 5552-5557

Mining Rare Association Rule

Arnab kumar Das

Asst Professor
Dept of Computer Application
GIMT, Guwahati-17, Assam, India

Abstract— **Pattern mining is a data mining method that involves finding existing patterns in data. The main objective is to uncover the hidden regularities in the data. Until now most of the research in pattern mining has been exclusively focused on frequent item set, however in some situation it is interesting to find patterns that are rare instead of frequent or patterns that reflect a negative co relation between items. These patterns are referred to as rare and negative patterns. For example in jewellery sales data, sales of diamond watches are rare however patterns involving the selling of diamond watches could be interesting. Rare items occur rarely but are of special interest as they may highlight exceptional behaviour in the data which is likely of interest. It tries to capture patterns that involve events that are unusual in the data set. This research proposes that in some domain such as medical databases, computer security, rare items are of higher importance than frequent one. This research investigates other kind of knowledge that is not usually discovered using frequent item-set mining i.e. rare item-sets and non-present item-sets. In order to mine rare and non-present item-sets, we methodologically extract the principles behind Apriori approach, and then generalize it in such a way it can not only be applied to mine frequent item-sets but also to mine other categories of patterns such as rare and non-present item-sets. Our goal is to apply the principle of the Apriori approach used for frequent item-set mining and apply it to rare and non-present item-set mining. The proposed approach is able to find the whole rare item set and at the same time pruning out the zero itemset.We will be using the concept of minimal rare item set (MRI) to generate strong Rare Association rules with high confidence and low support. The MRI will act as a generator, based on which valid rare association rule will be generated**

Keywords— **Rare item set, Minimal Rare itemset, Rare Association Rule.**

INTRODUCTION

Data mining techniques are used to process huge amounts of information in order to extract hidden knowledge to be directly interpreted or exploited to feed other processes. In many cases, data mining techniques are used to discover patterns that can be of interest to a specific domain of application. A pattern is a collection of events/features that occur together in a transaction database. As a matter of fact, different categories of patterns exist, such as sequences, item-sets, association rules and graph patterns, etc. The choice of a technique depends not only on the nature of input data but also depends on what we want to obtain and what view or part of the data we want to be described or represented in a more intelligible and concise way. To filter out the patterns, some criteria are used. The most known criteria are the support and the confidence. While the support represents the number of times a pattern occurs in the initial database, the confidence represents a proportion value that shows how frequently a part of the pattern, called premise, occurs among all the records containing the whole rule body. Herein, we classify pattern categories according to the specific use of the support threshold. In fact, by setting ranges for the support, one can obtain different categories of patterns. For example, if we set a minimum support threshold that a pattern has to satisfy to be considered as an interesting pattern, we obtain what is called frequent pattern. By setting a maximum support threshold we obtain another category of patterns called rare patterns. Whereas frequent patterns focus on mining patterns that appear more frequently in the database, rare patterns aim at discovering patterns that are less frequent. Rare patterns can be used in different domains such as biology, medicine and security etc.. For example, by analyzing clinical databases one can discover rare patterns that will help doctors to make decisions about the clinical care. In the security field, normal behavior is very frequent, whereas abnormal or suspicious behavior is less frequent. Considering a database where the behavior of people in sensitive places such as airports are recorded, if we model those behaviors, it is likely to find that normal behaviors can be represented by frequent patterns and suspicious behaviors by rare patterns.

BASIC CONCEPTS

An association rule is an expression of the form $P1 \rightarrow P2$, where $P1$ and $P2$ are arbitrary item sets ($P1, P2 \subseteq A$), $P1 \cap P2 = \emptyset$ and $P2 = \emptyset$. The left side, $P1$ is called *antecedent*, the right side, $P2$ is called *consequent*. The support of an association rule r: $P1 \rightarrow P2$ is defined as: $supp(r) = supp(P1 \cup P2)$. The *confidence* of an association rule r: $P1 \rightarrow P2$ is defined as the conditional probability that an object includes $P2$, given that it includes $P1$: $conf(r) = supp(P1 \cup P2)/supp(P1)$. An association rule r is called *confident*, if its confidence is not less than a given *minimum confidence* (denoted by *min_conf*), i.e. $conf(r) \geq min\_conf$. An association rule r with $conf(r) = 1.0$ (i.e. 100%) is an *exact* association rule; otherwise it is an *approximate* association rule.

An association rule r is called *frequent* if its support is not less than a given *minimum support* (denoted by *min_supp*), i.e. $supp(r) \geq min\_supp$. A frequent association rule is *valid* if it is confident, i.e. $supp(r) \geq min\_supp$ and $conf(r) \geq min\_conf$.

An association rule is called *rare* if its support is less than a given *minimum support*. A rare association rule r is *valid* if r is confident, i.e. $supp(r) < min\_supp$ and $conf(r) \geq min\_conf$.

Let us consider the standard pattern mining environment, i.e., a set of m *items* $I = \{i_1, i_2, \cdot \cdot \cdot, i_m\}$ and a *transaction database* (TDB) $D = \{t_1, t_2, \cdot \cdot \cdot, t_n\}$ on top of I. A subset X of I is referred to as *item set* whereby if |X| = k, then X is a k-item set. Moreover, a *transaction* t is made of an item set I and a unique identifier *tid*, typically, a natural number.

Support and confidence is the prime measure of interest for finding in regularities in the data that manifest in recurring patterns.

In this section we present an example of rare and non-present item-set mining. The input data is composed of a database of transactions, and each transaction is identified by an id and composed of a set of items. In the real world, a transaction may be seen as the basket bought by a customer during a determined period of time (day, week, month, etc.). Each basket is composed of a set of items that are purchased together. In Table 1 we represent an abstract database, denoted by D, where the alphabet letters are considered as items. Given the database of transactions such as presented in Table 1, our goal is to find two categories of sets of items, also called item-sets. The first category is composed of item-sets that are present in at most two transactions, and the second category is composed of item-sets that do not occur in any transaction and composed of a maximum number of items equal to the cardinality of the largest transaction. The number of times an item-set occurs in the database is called the item-set support. In our case the maximum support is equal to 3.

Id	Items
t1	a,b,c,d
t2	b,d
t3	a,b,c,e
t4	c,d,e
t5	a,b,c

Table 1

The set of all item-sets that can be generated from the transaction database is presented in *Figure 1* by a diagram of the subset lattice for five items with the associated frequencies in the database. In the lattice each level is composed of item-sets having the same length. The top element in the lattice is the empty set.

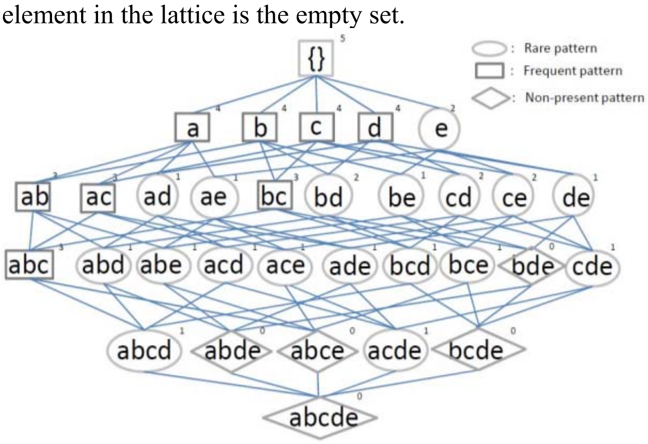

Figure 1: Lattice representing a hierarchically ordered space of item-sets and their frequencies. Frequent item-sets are square-shaped, rare item-sets are oval-shaped and non-present item-sets are diamond-shaped.

Each lower level k contains all of the item-sets of length k, also denoted k-item-sets and the last level contains an item-set composed of all items (i.e. *a, b, c, d, e*). Lines between nodes represent a subset relationship between item-sets. For each item-set we compute its support. For example, the item-set composed of the items b and d denoted by <*bd*> have 2 as support, and we denote it by <*bd*,2>. In fact, the item-set *bd* is present in the transactions *t1*, *t2*. The set of rare item-sets we are looking for are those in the lattice with support greater than 0 and less than 3. The rest of item-sets are either frequent (having support greater or equal to 3) or non-present (support=0). In Figure 1 rare item-sets are drawn with ovals, frequent item-sets are drawn with rectangles where non-present item-sets are drawn with diamonds. Thus, after counting the frequencies of each item-set, we obtain the following set of rare item-sets <*e*,2>, <*ad*,1>, <*ae*,1>, <*bd*,2>, <*be*,1>, <*cd*,2>,<*cd*,2>, <*de*,1>, <a*bd*,1>, <*abe*,1>, <acd,1>, <ace,1>, <ade,2>, <*bcd*,1>, <bce,1>, <*cde*,1>, <a*bcd*,1>, <acde,1>. The set of non-present item-sets is composed of the following elements <abde, 0>, <*abce*,0>, <*bcde*,0>, <abcde,0>.

Definition 1: An item set is a maximal frequent item set (MFI) if it is frequent but all its proper supersets are rare.

Definition 2: An item set is a minimal rare item set (MRI) if it is rare but all its proper subsets are frequent.

Definition 3: An item set X is a (minimal or key) generator if it has no proper subset with the same support (For all Y which is a subset of X, supp(X) < supp(Y)).

Definition 4: A minimal zero generators (mZG) is a zero item set whose proper subsets are all non-zero item sets.

RELATED WORK

The previous methods to mine item-sets can be divided into two categories, namely frequent item-set mining and rare item-set mining techniques. Whereas the problem of frequent item-set mining have been widely studied, the problem of rare item-set mining has just started to spark the researchers interest and non-present item-sets have not yet been addressed as independent mining task.

Frequent Item-set Mining

Item-set space has two properties, a monotone property and an anti-monotone property [12][2] [20]. Every non-empty subset of a frequent item-set is a frequent item-set, and every superset of a non-frequent item-set is non-frequent. Based on those properties, the algorithm Apriori [2] was developed to efficiently mine frequent item-sets. Later on, new algorithms have been proposed to mine frequent item-sets such as Eclat [24], FP-Growth [11] and TM [19]. A recent survey on frequent item-sets mining techniques is presented in [23]. The Apriori approach was successfully used to generate frequent item-sets contained in a transaction database [2] [3] [17]. Apriori approach exploits the monotonicity property of the support of item-sets. Apriori-based algorithms perform top-down breadth-first search through the space of all item-sets. In the first pass, the support of each individual item is counted, and the frequent ones are inserted to the frequent item-set set of level 1. In each subsequent pass, the frequent item-sets determined in the previous pass are used to generate new item-sets called candidate item-sets. The support of each candidate item-set is counted, and the frequent ones are determined. This process continues until no new frequent item-sets are found.

Rare and Non-Present Item-sets

Recently, a work presented in [21] proposes an approach to mine rare item-sets that is based on the Apriori algorithm used to mine frequent item-sets. The main idea consists at traversing the item-set space by the Apriori algorithm used to mine frequent item-sets and collect at each level the item-sets that are usually pruned out in the original algorithm and are used as seed for a second algorithm in order to mine the remaining rare item-sets. Another algorithm, called MINIT, is proposed in [10] to mine only minimal infrequent item-sets. A minimal infrequent item-set is an infrequent item-set that do not have a subset of items which forms an infrequent item-set. In other words, an infrequent item-set is minimal if all its proper subsets are frequent. It is noteworthy that the output of this algorithm may be used to mine all rare item-sets. More recently in [1] a framework is proposed that absorbs the spirit of the Apriori approach to tackle the problem of rare item-set mining. In this work, we started by factorizing the key elements of Apriori approaches such as the traversal of the item-set space, the pruning principle, the combination of item-sets at a level to generate new item-set in the next level, and then proposed an Apriori generalized framework that abstracts those elements. However, in the International Journal on Soft Computing, Artificial Intelligence and Applications proposed approach, no distinction is made between rare and non-present patterns. A performance comparison of our algorithm with the algorithm presented in [21] is also presented.

Intrusion Detection

Intrusion detection methods are of two types: anomaly detection and misuse (signature) detection. While anomaly detection techniques focus on the detection of user behaviour that is considered as abnormal [5] [22] [7], signature detection focuses on the identification of a behaviour that is similar to known cases that are considered as intrusions. A model is generally is used to represent the known intrusions [13] [16]. The main drawback of model based intrusion detection is the new attacks may be missed if not always kept up-to-date. In order to overcome this limitation, data mining and machine learning approaches are is the current trend to detect intrusion [14] [6] [4] [8] [25].

<u>Rare item set Geneartion Algorithm</u>

Apriori-Rare[20]: This algorithm generates the frequent as well as rare itemsets. It is a slight modification of Apriori algorithm to generate frequent and rare itemsets.

In Apriori algorithm the rare items are pruned out but here the rare items are also collected along with the frequent item set based on the minimum support value. However, it fails to find all the rare itemsets.

Apriori-Inverse[10]: This algorithm determines only the sporadic rules using one Minsup value and one Maxsup value. The sporadic rules have the property that they fall below user define Maxsup but above the Minconf value. The main advantage of Apriori-Inverse is that it can find the sporadic itemsets much more quickly than apriori. However, a major limitation is that it is incapable of finding all the rare itemsets.

MRG-Exp[26]: This algorithm takes the dataset and min support as input and produces the minimal rare generator as output. This algorithm uses the concept of predecessor support of an item set (i) which is the minimum of the supports of all (i – 1)-long subset of i. The predecessor support is then compared to the actual support of a candidate. If both values are different then the candidate is a true generator. Moreover, depending on its min support, it is either a frequent generator or a minimal rare generator, i.e., an MRI.The main disadvantage of this algorithm is that it is unable to produce the all rare item set or also does not prune out the zero item set.

ARIMA (Another Rare Item set Miner Algorithm)[26]: This algorithm takes the dataset and minimal rare item set(MRI) as input and produces all non zero rare item set plus minimal zero generator. For this it initially calls the Apriori-Rare or MRG-Exp algorithm that generates the Minimal Rare Item sets. It first consider the smallest minimal rare generator and from this its supersets are generated. A dataset scan is made to check the presence of zero generators and is copied to the minimal zero generator (mRG) set which ultimately reduce the search space. The main advantage of ARIMA is that it can find all rare item set . However, it is dependent on two the algorithms MRG-Exp and Apriori-Rare for the minimal rare item set.

DISCUSSION AND MOTIVATION

After studying and analyzing the above mentioned rare item set generation algorithm it is observed that all the above mentioned algorithm is having some limitations. The algorithm Apriori-Rare, Apriori- Inverse and MRG-exp are unable to generate all rare item set. Although the algorithm ARIMA is capable of generating all rare item set it is dependent on either Apriori-Rare or MRG-Exp. Moreover the rule generated from this item set is not all interesting. To overcome the limitations faced by the above mentioned algorithm we propose an algorithm that takes the min support and dataset as input and produces valid rare association rule as output. In this approach the candidate set is generated not only from the frequent set but also from the rare item set. The Proposed algorithm is discussed below.

PROPOSED APPROACH FOR FINDING RARE ASSOCIATION RULE

<u>Proposed algorithm</u>

Description: Finding Rare Association Rule
Input: Dataset and minimum support
Output: Minimal Rare item set (MRI) and valid Rare Association Rule

1) CG1(1 item set)
2) Support Count(CG1)
3) Zi= item set having support count equal to zero

4) F_i= item set having support count larger than min support
5) R_j= item set having support count smaller than min support
6) While(F_i and R_j is not Null)
7) GenerateCandidate(F_i,Rj)
8) R_{all}→ Combine all rare item set.
9) For all $R_j \in R_{all}$
10) R_{isub}=all subset R_S of R_j
11) If for all $R_S \in R_{isub}$
12) $R_S \in F_{i\ Then}$
13) $MRI_= R_i$
14) For each MRI
15) MRI_{Clo}→Closure of the MRI
16) For each MRI_{Clo}
17) Generate Association rule of the form MRI→(MRI_{Clo} - MRI)

DESCRIPTION OF THE ALGORITHM

The proposed algorithm takes the dataset and minimum support as input and produces minimal rare item set and valid rare association rule as output.

In the initial step it makes a scan of the dataset to find the support count of the items and divide the items in the following category as follows

Zero items:→ Items having support count zero

Frequent items:→ Items having support count greater or equal to minimum support

Rare items:→Items having support count less than minimum support

In the next step candidate items are generated not only from the frequent items but also from the rare items. That is both frequent and rare items are combined to generate the candidate item list. Again a database scan is made to count the support count of the candidate item set and divide it into zero item set, frequent item set and rare item set as before. This procedure is repeated until no more candidate item can be generated.

In the next step all rare item sets generated at each iterartion are combined to find the whole rare item sets. Now from this whole rare item set the minimal rare item sets are found. An item set is called minimal rare item set if it is rare but all its proper subsets are frequent. This is done to remove the sporadic rules that may be generated if we consider all rare item set.

After finding the minimal rare item set we find the closure of the item set to find their equivalence class.

The closure of an itemset is obtained by first applying ∂ and then f:

C (I) = f(∂ (I)), where ∂ (I) and f(T) are defined as follows

For an itemset I, denote by ∂ (I) the set of transactions in which all items in I are present, that is,

∂ (I) = {t ϵ db/ for all i ϵ I, i ϵ t}

For a set of transactions T let f (T) denote the set of items that are present in all transactions in T, that is,

f(T) = {i ϵ R / for all t ϵ T, i ϵ t}

i. e to obtain the closure of an itemset we first find the transactions in which all items in I are present, and then see whether there are any more items that are common to all these transactions. If any item is common then add it to the closure of that itemset.

After finding the closure the association rule is generated as follows

R: MRI→(MRI_{Clo} - MRI)

Where MRI is a minimal rare item set and MRI_{Clo} is the closure of the MRI and conf(R) =1.0.

It is observed that all association generated using this approach is valid rare association rule.

EXECUTION OF THE PROPOSED ALGORITHM ON DATASET D

Min support=3

Tid	Itemset
1	A,B,D,E
2	A,C
3	A,B,C,E
4	B,C,E
5	A,B,C,E

Fig: Dataset D

CG1	Supp
A	4
B	4
C	4
D	1
E	4

FG1	Supp
A	4
B	4
C	4
E	4

RG1	Supp
D	1

ZG1
Nil

CG2	Supp
AB	3
AC	3
AE	3
AD	1
BC	3
BE	4
BD	1
CE	3
CD	0
DE	1

FG2	Supp
AB	3
AC	3
AE	3
BC	3
BE	4
CE	3

RG2	Supp
AD	1
BD	1
DE	1

ZG2
CD

CG3	Supp
ABC	2
ABE	3
ABD	1
ACE	2
ACD	0
ADE	1
BCE	3
BCD	0
BDE	1

FG3	SUPP
ABE	3
BCE	3

RG3	Supp
ABC	2
ABD	1
ACE	2
ADE	1
BDE	1

ZG3
ACD
BCD

CG4	SUPP
ABCE	2
ABCD	0
ABDE	1

FG4	SUPP
NIL	

RG4	SUPP
ABCE	2
ABDE	1

ZG4
ABCD

Item Set	Supp	Subset	Frequent or Rare	Minimal rare itemset
D	1	Empty set	Frequent	Yes
AD	1	A	Frequent	No
		D	Rare	
BD	1	B	Frequent	No
		D	Rare	
DE	1	D	Rare	No
		E	Frequent	
ABC	2	AB	Frequent	Yes
		AC	Frequent	
		BC	Frequent	
ABD	1	AB	Frequent	No
		AD	Rare	
		BD	Rare	
ACE	2	AC	Frequent	Yes
		AE	Frequent	
		CE	Frequent	
ADE	1	AD	Rare	**No**
		AE	Frequent	
		DE	Rare	
BDE	1	BD	Rare	No
		DE	Rare	
		BE	Frequent	
ABCE	2	ABC	Rare	No
		ACE	Rare	
		BCE	Frequent	
		ABE	Frequent	
ABDE	1	ABD	Rare	No
		ADE	Rare	
		BDE	Rare	
		ABE	Frequent	

Fig: Finding minimal rare item set (MRI) from RI_{all}(All rare item set)

MRI= {D, ABC, ACE}

MRI	CLOSURE
D	ABDE
ABC	ABCE
ACE	ABCE

Fig: Closure of the MRI

From the closure the Association rule is generated as follows

MRI →(MRI_{Clo} - MRI)

Thus we get

D→ABDE-D

D→ABE, CONFIDENCE=SUP (ABDE)/SUP (D) =1/1=1

Similarly we get

ABC→E, CONFIDENCE=SUP (ABCE)/SUP (ABC) =2/2=1

ACE→B, CONFIDENCE=SUP (ACEB)/SUP (ACE) =2/2=1

OBSERVATION

It is observed that the proposed algorithm is able to find out the whole rare item set. Moreover it is able to prune out the zero item set. Also the rule generated from this algorithm are all valid rules with confidence of 100%.Another advantage of this algorithm is that it is not dependent on any other algorithm.

CONCLUSION

Mining rare patterns is a challenging endeavor because there is an enormous number of such patterns that can be derived from a given data set. The key issues in mining infrequent patterns is to how to identify interesting infrequent patterns and how to effectively discover them in large data set.

In this research we first highlight the importance of rare item set in some application. Then we proposes some methods for finding the rare item set and finally we try to find interesting patterns from those item set. We have used the basic principle of Apriori Approach to find out the rare item set in a level wise manner from the minimal rare item set (mRI). We will also examine the benefit of depth first search method for finding the rare item set.Finally we will look for a compact representation of rare item set family.

REFERENCES

[1] Adda, M., Wu, L., & Feng, Y. (2007). Rare item-set mining. *Proceedings of the Sixth International Conference on Machine Learning and Applications* (pp. 73–80). Washington, DC, USA: IEEE Computer Society.

[2] Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules in Large Databases. *Proceedings of the 20th International Conference on Very Large Data Bases* (pp. 487–499). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

[3] Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. *Proceedings of the 1993 ACM SIGMOD international conference on Management of data* (pp. 207–216). Washington, D.C., United States: ACM.

[4] Amor, N., Benferhat, S., & Elouedi, Z. (2004). Naive Bayes vs decision trees in intrusion detection systems. *ACM Symposium on Applied Computing* (pp. 420–424). New York: ACM Press.

[5] Barbara, D., Couto, J., Jajodia, S., & Wu, N. (2001). Adam: a testbed for exploring the use of data mining in intrusion detection. *Special Section on Data Mining for Intrusion Detection and Threat Analysis* , 15–24.

[6] Chan, P., Mahoney, M., & Arshad, M. (2003). *A machine learning approach to anomalydetection.* Florida Institute of Technology.

[7] Denning, D. (1987). An intrusion-detection model. *IEEE Transactions on Software Engineering* , 222–232.

[8] DongJun, Z., Guojun, M., & Xindong, W. (2008). An Intrusion Detection Model Based on Mining Data Streams. *Proceedings of The 2008 International Conference on Data Mining, DMIN* (pp. 398–403). Las Vegas: CSREA Press.

[9] Gou, M., Norihiro, S., & Ryuichi, Y. (2002). A framework for dynamic evidence based medicine using data mining. *Proceedings of the 15th IEEE Symposium on Computer-Based Medical Systems* (p. 117). Maribor, Slovenia: IEEE.

[10] Haglin, D. J., & Manning, A. M. (2007). On minimal infrequent item-set mining. *Proceedings of the 2007 International Conference on Data Mining, DMIN 2007* (pp. 141–147). Las Vegas, Nevada, USA: CSREA Press.

[11] Han, J., Pei, J., & Yin, Y. (2000). Mining frequent patterns without candidate generation. *In Proceedings of the 2000 ACM SIGMOD international conference on Management of data* (pp. 1–12). New York, NY, USA: CM Press.

[12] Iwanuma, K., Takano, Y., & Nabeshima, H. (2004). On anti-monotone frequency measures for extracting sequential patterns from a single very-long data sequence. *IEEE Conference on Cybernetics and Intelligent Systems*, (pp. 213–217).

[13] Kumar, S., & Spafford, E. (1994). *An application of pattern matching in intrusion detection.* Technical Report 94–013, Department of Computer Sciences, Purdue University.

[14] Lane, T., & Brodley, C. (1997). An application of machine learning to anomaly detection. *Proceedings of the 20th NIST-NCSC National Information Systems Security Conference*, (pp. 366–380).

[15] Lee, W. (1998). Data mining approaches for intrusion detection. *In Proceedings of the Seventh USENIX Security Symposium* (pp. 6–6). San Antonio, Texas: USENIX Association Berkeley, CA, USA.

[16] Mansfield, G., Ohta, K., Takei, Y., Kato, N., & Nemoto, Y. (2000). Towards trapping wily intruders in the large. *Computer Networks* , 659–670.

[17] Pasquier, N., Bastide, Y., Taouil, R., & Lakhal, L. (1999). Efficient Mining Of Association Rules Using Closed Item-set Lattices. *Information Systems*, 25–46.

[18] Prati, R. C., Monard, M. C., André, C. P., & de Carvalho, L. F. (2004). A method for refining knowledge rules using exceptions, SADIO. *Electronic Journal of Informatics and Operations Research* , 53–65.

[19] Song, M., & Sanguthevar, R. (2006). A transaction mapping algorithm for frequent item-sets mining. *IEEE Transactions on Knowledge and Data Engineering* , 472–481.

[20] [20] Srikant, R., & Agrawal, R. (1997). Mining Generalized Association Rules. *Future Generation Computer Systems* , 161–180.

[21] Szathmary, L., Maumus, S., Petronin, P., Toussaint, Y., & Napoli, A. (2006). Vers l'extraction de motifs rares. *EGC: Extraction et Gestion de Connaissances*, (pp. 499–510).

[22] Tan, K., Killourhy, K., & Maxion, R. (2002). Undermining an Anomaly-Based Intrusion Detection System Using Common Exploits. *Recent Advances in Intrusion Detection* , 54–73.

[23] Tiwari, A., Gupta, R., & Agrawal, D. (2010). A survey on frequent pattern mining: Current status and challenging issues. *Inform. Technol* , 1278–1293.

[24] Zaki, M. J., Parthasarathy, S., Ogihara, M., & Li, W. (1997). New algorithms for fast discovery of association rules. *In Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining*, (pp. 283–296).

[25] Zhan, J., & Leshan Teachers Coll., L. (2008). Intrusion Detection System Based on Data Mining. *First International Workshop on Knowledge Discovery and Data Mining, 2008. WKDD 2008*, (pp. 23–24). Adelaide, SA.

[26] Szathmary, L., Napoli, A., Valtchev, P.: Towards Rare Itemset Mining. In: Proceedings of the 19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI '07). Volume 1., Patras, Greece (Oct 2007) 305–312