Key correlation mining by simultaneous monotone and anti-monotone constraints checking

Souad Bouasker

University of Tunis ElManar
Faculty of Sciences of Tunis
LIPAH-LR 11ES14,2092, Tunis, Tunisia
bouasker.souad@gmail.com

Sadok Ben Yahia

University of Tunis ElManar
Faculty of Sciences of Tunis
LIPAH-LR 11ES14 2092, Tunis, Tunisia
sadok.benyahia@fst.rnu.tn

Institut Telecom, Telecom SudParis
umr 5157, CNRS SAMOVAR, France

ABSTRACT

Correlated pattern mining has become increasingly an important task in data mining and knowledge discovery. In practice, the exploitation of correlated patterns is hampered by the high number of the generated patterns. Thus, the integration of the constraint of frequency with the constraint of correlation has been proved to be very interesting by mining *Frequent correlated patterns* [2, 14] and *Rare correlated patterns* [4, 3]. In this situation, the main task concerns the manipulation of the constraints of correlation and of frequency. One way to deal with this issue is to mine all the correlated patterns and then to filter by the constraint of frequency. However, this filtering is done as a post-processing phase and it suffers from the important number of patterns and loses the opportunity to exploit the selectivity power of both constraints.

In this paper, we introduce an approach that puts the focus on mining rare correlated patterns according to the *bond* measure. We were based on the simultaneous integration of the anti-monotone constraint of correlation and the monotone constraint of rarity during the mining process. Our experimental studies shows an important benefit when early pushing the constraints of distinct types. We also flag out better performances than the GMJP approach [3] which also dealt with both types of constraints.

Categories and Subject Descriptors

H.4 [**Interesting Pattern Extraction**]: Constraint Data Mining

General Terms

Algorithms, Experimentation, Verification

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
SAC'15 April 13-17, 2015, Salamanca, Spain.
Copyright 2015 ACM 978-1-4503-3196-8/15/04$15.00.
http://dx.doi.org/10.1145/2695664.2695802

Keywords

Correlated pattern, *bond* measure, monotone, anti-monotone, constraint

1. INTRODUCTION AND MOTIVATIONS

Correlated item set mining is at the core of numerous data mining tasks. The enormous research efforts dedicated to this topic have led to a variety of sophisticated approaches [1, 10, 11]. Thus, a variety of correlation measures were proposed and studied. In this regard, frequent correlated itemset mining was then shown to be an attractive research field. Since its inception, this key task grasped the interest of many researchers since it meets the needs of experts in several application fields [2], such as market basket study [5]. Nevertheless, the application of these correlated frequent patterns may not be a very attractive solution for some other applications, *e.g.*, detection of rare diseases from medical data, intrusion detection, to cite but a few [13]. Recently, the authors in [3] proposed a generic approach, called GMJP, managing with both frequent correlated and rare correlated patterns according to the *bond* measure.

In this work, we propose a significant improvement of the performance of the GMJP approach. Indeed, we study the challenging problem of constraints integration from another perspective. We are mainly based on a simultaneous pushing of both constraints in an early stages of the mining process.

The remainder of the paper is organised as follows. The problem statement is presented in Section 2. Section 3 presents the related works. Section 4 is dedicated to presenting the relevant types of constraints monotone, and anti-monotone besides to the properties of the set of the desired mined patterns. In Section 5, we present our new CORI approach based on a depth traversal of the itemsets lattice. In addition, we pay attention to the way to push simultaneously both monotone and anti-monotone constraints. Our experimental results are discussed in Section 6. Finally, the last Section concludes the paper and points out avenues of future work.

2. PROBLEM STATEMENT

The problem of mining association rules was initially applied to market basket analysis [5]. This problem consists in finding interesting associations among items and itemsets from transactional dataset. Formally, we define a dataset.

Table 1: An example of a dataset $\mathcal{D}$.

	A	B	C	D	E
1	×		×	×	
2		×	×		×
3	×	×	×		×
4		×			×
5	×	×	×		×

DEFINITION 1. (***Dataset***) *A dataset is a triplet $\mathcal{D} = (\mathcal{T}, \mathcal{I}, \mathcal{R})$ where $\mathcal{T}$ and $\mathcal{I}$ are, respectively, a finite set of transactions and items, and $\mathcal{R} \subseteq \mathcal{T} \times \mathcal{I}$ is a binary relation between the transaction set and the item set. A couple $(t, i) \in \mathcal{R}$ denotes that the transaction $t \in \mathcal{T}$ contains the item $i \in \mathcal{I}$.*

In this work, we are mainly interested in itemsets as a class of patterns. The two main kinds of support a pattern can have are defined as follows, for any non-empty pattern I:
- **Conjunctive support**: $Supp(\wedge I) = |\{t \in \mathcal{T} \mid (\forall\, i \in I, (t,i) \in \mathcal{R})\}|$
- **Disjunctive support**: $Supp(\vee I) = |\{t \in \mathcal{T} \mid (\exists\, i \in I, (t,i) \in \mathcal{R})\}|$ An itemset I is frequent if its support Supp(I) is above a user-defined minimum support threshold *minsupp*, otherwise the itemset I is said to be *infrequent* or rare. The task of generating association rules from frequent patterns suffers from the sheer number of those rules. Thus, with a very large number of rules, the exploitation of the conveyed knowledge becomes a difficult task. To solve this problem, different measures of correlation have been proposed to filter the extracted patterns and thus reduce the discovered rules. In our work, we are interested in the *bond* measure.

The *bond* correlation measure [9] is mathematically equivalent to *Jaccard*. It was also used in many works under various names like *extended Jaccard measure*, *coherence*, *Tanimoto coefficient*. It was redefined in [2] as:

$$bond(I) = \frac{Supp(\wedge I)}{Supp(\vee I)}$$

The set of correlated patterns associated with the *bond* measure is defined as follows.

DEFINITION 2. (***Correlated patterns***) *Considering a minimum correlation threshold minbond, the set $\mathcal{CP}$ of correlated patterns is equal to: $\mathcal{CP} = \{I \subseteq \mathcal{I} |\ bond(I) \geq minbond\}$* [1].

The *bond* measure takes its values within the interval $[0, 1]$. While considering the universe of a pattern I, *i.e.*, the set of transactions containing a non empty subset of I, the *bond* measure represents the simultaneous occurrence rate of the items of the pattern I in its universe. Thus, the more the items of I are dependent on each other, (*i.e.* strongly correlated), the higher the value of the *bond* measure is, since $Supp(\wedge I)$ would be closer to $Supp(\vee I)$.

3. RELATED WORK

The problem of mining correlated patterns was thoroughly studied in various works in the literature. The *bond* measure was studied in [8], where the authors proposed an apriori-like algorithm for mining classification rules. Moreover, the authors in [10] proposed a generic approach for correlated patterns mining. Indeed, the *bond* correlation measure and eleven other correlation measures were used, all of them fulfil the anti-monotonicity property. Correlated patterns mining was then shown to be more complex and more informative than frequent patterns mining [10]. Also, in [12], a study of different properties of interesting measures was carried out in order to suggest a set of the most adequate properties to consider while mining rare associations rules. However, it is important to highlight that the extraction of rare correlated patterns was not carried out in [10] nor in [12].

Recently, the authors in [4] proposed a concise exact representation of the set of rare correlated patterns. As we also previously mentioned, in [3] the authors proposed the GMJP approach allowing the extraction of both frequent correlated patterns, rare correlated patterns and their associated concise representations. Another correlation measure, *All-Confidence*, was also studied in [7] where the authors proposed an algorithm allowing the extraction of frequent correlated patterns. The authors dressed the problem of the *minsupp* threshold fixing. In fact, a very low *minsupp* value, allow to extract long itemsets. In turn, this can lead to a combinatorial explosion of the number of the generated correlated frequent itemsets.

4. THE CHALLENGE

We are motivated in this work, to build a smart link between correlation mining and frequency mining. Our idea is based on the treatment of the *bond* correlation measure. Our choice of this measure is justified by the theoretical framework presented in [9], in addition to the structural study that was done in [10].

While integrating the frequency constraint with the correlation constraint, we can distinguish between two sets of correlated patterns, which are the "Frequent correlated patterns" set and the "Rare correlated patterns" set. In the following, we present the category of constraints fulfilled by these two previous sets.

4.1 Categories of constraints

We begin by defining *monotone* and *anti-monotone* constraints.

DEFINITION 3. (***Anti-Monotone constraint***) *Let Amc be a constraint,*

• *Amc is anti-monotone : if $\forall\ I \subseteq \mathcal{I}$, $\forall\ I_1 \subseteq I$: I fulfils Amc $\Rightarrow$ I_1 fulfils Amc*

DEFINITION 4. (***Monotone constraint***) *Let Mc be a constraint,*

• *Mc is monotone : if $\forall\ I \subseteq \mathcal{I}$, $\forall\ I_1 \supseteq I$: I fulfils Mc $\Rightarrow$ I_1 fulfils Mc*

The constraint of rarity is a monotone constraint, *i.e.*, $\forall\ I, I_1 \subseteq \mathcal{I}$, if $I_1 \supseteq I$ and $Supp\ (\wedge I) < minsupp$, then $Supp\ (\wedge I_1) < minsupp$ since $Supp(\wedge I_1) \leq Supp\ (\wedge I)$. Thus, it induces an *order filter* [6] on the set of all the subsets of $\mathcal{I}$, $\mathcal{P}(\mathcal{I})$. Worth of mention that the frequency constraint is an anti-monotone constraint and it induces an *order ideal* [6] in the itemset lattice. Furthermore, the correlation constraint induced by the *bond* measure fulfils an anti-monotone constraint. Thus, it induces an *order ideal* in the itemset lattice.

[1] We refer in the rest of the paper to the minimum support threshold by *minsupp* and to the minimum correlation threshold by *minbond*.

In what follows, we carry out a detailed study of the two sets of frequent correlated and rare correlated patterns.

4.2 Properties of the correlated patterns

We start by defining the set of frequent correlated patterns:

DEFINITION 5. (***Frequent correlated patterns***) *[2] The set $\mathcal{FCP}$ of frequent correlated patterns is equal to: $\mathcal{FCP} = \{I \subseteq \mathcal{I} \mid Supp(\wedge I) \geq minsupp \text{ and } bond(I) \geq minbond\}$.*

DEFINITION 6. (***Rare correlated patterns***) *[4] The $\mathcal{RCP}$ set of rare correlated patterns is equal to: $\mathcal{RCP} = \{I \subseteq \mathcal{I} \mid Supp(\wedge I) < minsupp \text{ and } bond(I) \geq minbond\}$.*

In fact, the $\mathcal{FCP}$ set of frequent correlated patterns corresponds to the conjunction of two anti-monotonic constraints which are the correlation constraint and the frequency constraint. Thus, the $\mathcal{FCP}$ respect an anti-monotone constraint and induces an order ideal in the itemset lattice. However, the $\mathcal{RCP}$ set of rare correlated patterns corresponds to the conjunction of two constraints of distinct types, which are the anti-monotone constraint of correlation and the monotone constraint of rarity. Thus, the $\mathcal{RCP}$ do not fulfil an anti-monotone constraint nor a monotone one. Therefore, the localization of these rare correlated patterns is much more complex than the localization of theories corresponding to constraints of the same nature

The $\mathcal{RCP}$ set is depicted by Figure 1. The support shown at the top left of each frame represents the conjunctive support. As shown in Figure 1, the rare correlated patterns are localized below the border shown in red of the anti-monotone constraint of correlation, and over the border shown in black of the monotone constraint of rarity. We present, in the next section, our new CORI algorithm [2].

5. OUR CONTRIBUTION

We introduce in this section the CORI approach which allows the extraction of the $\mathcal{RCP}$ set of rare correlated patterns. This latter results from the conjunction of the anti-monotone constraint of correlation and the monotone constraint of rarity. Thus, the important question is how to efficiently integrate these two types of constraints within the mining process ? Our strategy proposed within the CORI algorithm consists in three main steps:

1. **Step 1: Reduction of the initial dataset (*cf.* Line 1 of Algorithm 1)**
   Initially, the dataset is scanned once to build the new transformed dataset and to construct an in-memory structure. In fact, we assign to each item a bitset, each column of this bitset indicates the presence or the absence of the item in a specified transaction. For example, if the third column of this list contains 0, then the item I is not present in the third transaction. Our choice of bitsets is owe to the fact that each a bitset is a container that can store a huge number of bits while optimizing the memory consumption [3] (For example, 32 elements are stored in a memory block of 4 bytes). Each block of memory is treated in just one `CPU` operation by a 32 bits processor.

[2]CORI stands for **Co**rrelated **R**are **I**temset.

Algorithm 1: CORI

Data:
1. A dataset $\mathcal{D}$.
2. A minimal correlation threshold *minbond* of the anti-monotone constraint.
3. A minimal conjunctive support threshold *minsupp* of the monotone constraint.

Result: The $\mathcal{RCP}$ set of rare correlated patterns.

1 **begin**
1. Scan the dataset $\mathcal{D}$ once to build the transformed dataset $\mathcal{D}*$
2. Initialization of the tree-data structure
   (a) Computing the conjunctive support of the items and sorting them in an ascendant order of their support value.
   (b) Rare items are printed in the output result set.
   (c) The sorted items are added to the tree structure.
3. Recursive processing of each item in order to extract the rare correlated itemsets
4. Memory liberation
5. **return** $\mathcal{RCP}$

2 **end**

The examples of the bitsets associated with item `A` and to item `C` are illustrated in Figure 2 while the new transformed dataset associated with the dataset $\mathcal{D}$ (*cf.* Table 1), is given in Table 2.

2. **Step 2: Initialization of the tree-data structure**
   We begin by computing the conjunctive support of the items using their associated bitsets (*cf.* Figure 2 where BSVector stands for bitset vector). The items are then sorted in an ascending order of their support. All the items are evaluated according to the monotone constraint of rarity: The rare items are printed to the output set while frequent ones are not pruned but they are maintained. All the items are added to the nodes of the first level of our tree structure.

3. **Step 3: Solving recursively the mining problem**
   This third step is the main core of our algorithm. Our idea consists in dividing our mining problem into subproblems. For each item `I`, a sub-tree is constructed and a depth-first traversal is therefore performed. The candidates of size n are generated by building intersection of itemsets of size $n-1$. They are then evaluated as follows:
   - **Evaluation of the monotone constraint of rarity**: if the candidate is rare, then its correlation value is computed. We have to distinguish two possible cases: **The anti-monotone constraint of correlation is satisfied**, then it is added to the result set. If the candidate is not correlated, then all its superset will not be correlated according to the monotonicity property of the non-correlation constraint. In this case, all the associated sub-tree is pruned.
   - If the candidate does not fulfil the monotone constraint of rarity, then it wont be pruned. In fact, a

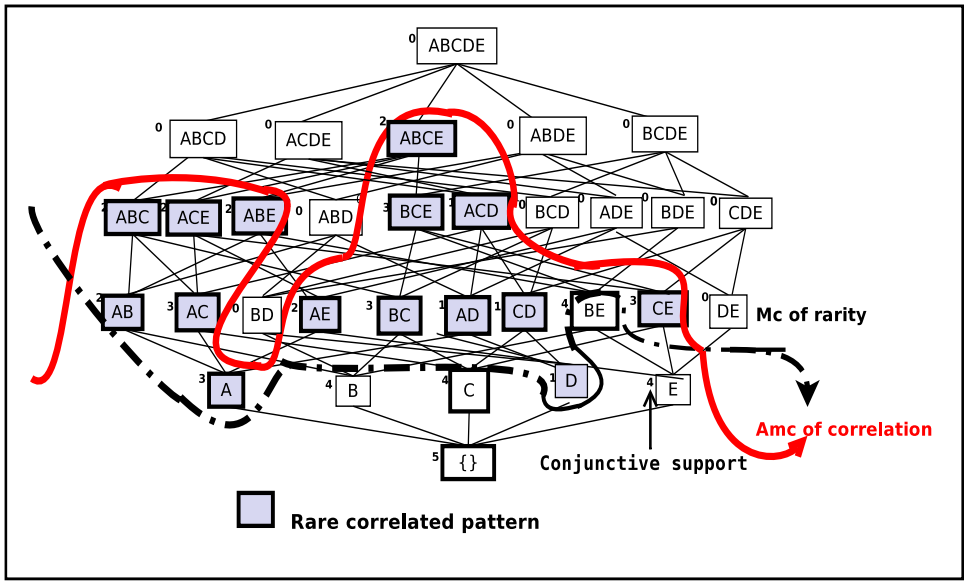

Figure 1: Localization of the $\mathcal{RCP}$ set of rare correlated patterns for the context $\mathcal{D}$ for *minsupp* = 4 and *minbond* = 0.2.

Table 2: The transformed dataset $\mathcal{D}*$ associated with the dataset $\mathcal{D}$.

	T1	T2	T3	T4	T5
A	1	0	1	0	1
B	0	1	1	1	1
C	1	1	1	0	1
D	1	0	0	0	0
E	0	1	1	1	1

frequent pattern can have rare supersets. In this case, the candidate is maintained and we continue to develop its sub-tree.

This generation and evaluation process is continued, the dedicated procedure is recursively called while there is a number of candidates to be generated. Finally, the used memory is liberated.

The example, shown in Figure 3, illustrates how CORI works. The tree is made from five sorted items, D, A, B, C, and E, with their supports 1, 3, 4, 4 and 4 respectively. We begin by the node containing the item with the lowest support: which is D. The anti-monotone constraint 'Amc' is having the *bond* value ≥ 0.20 and the monotone constraint 'Mc' is having the conjunctive support < 4. Intersecting D with A, B, C, and E produces DA, DB, DC and DE. The candidates DB and DE are pruned since they have null support. Thus, their supersets will have null support values. The candidates DA and DC are rare correlated. Thus, they both pass constraints Amc and Mc. Their intersection produces DAC with a support value equal to 1 and bond value equal to $\frac{1}{4}$, which also fulfils both constraints. The rare correlated itemsets are added to the output set and the D-sub-tree is deleted from our tree. The A-sub-tree, B-sub-tree, C-sub-tree and E-sub-tree are built

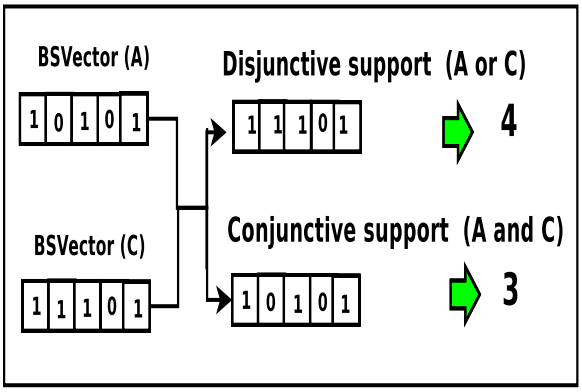

Figure 2: An example of conjunctive and disjunctive supports computing using the bitsets.

successively. The same process is recursively repeated until there's no more candidates to be generated and evaluated. Finally, the obtained result set is composed by: $(D, 1, \frac{1}{1})$, $(A, 3, \frac{3}{3})$, $(AB, 2, \frac{2}{5})$, $(AC, 3, \frac{3}{4})$, $(AE, 2, \frac{2}{5})$, $(BC, 3, \frac{3}{5})$, $(CE, 3, \frac{3}{5})$, $(DA, 1, \frac{1}{3})$, $(DC, 1, \frac{1}{4})$, $(ABC, 2, \frac{2}{5})$, $(ABE, 2, \frac{2}{5})$, $(ACE, 3, \frac{3}{4})$, $(BCE, 3, \frac{3}{5})$, $(DAC, 1, \frac{1}{4})$, and $(ABCE, 2, \frac{2}{5})$.

6. PERFORMANCE EVALUATION

To evaluate our CORI algorithm, we conducted a set of experiments with dense and sparse datasets of varying size [3]. We compared our algorithm with GMJP [3] while run-

[3]Available at *http://fimi.cs.helsinki.fi/data* and at

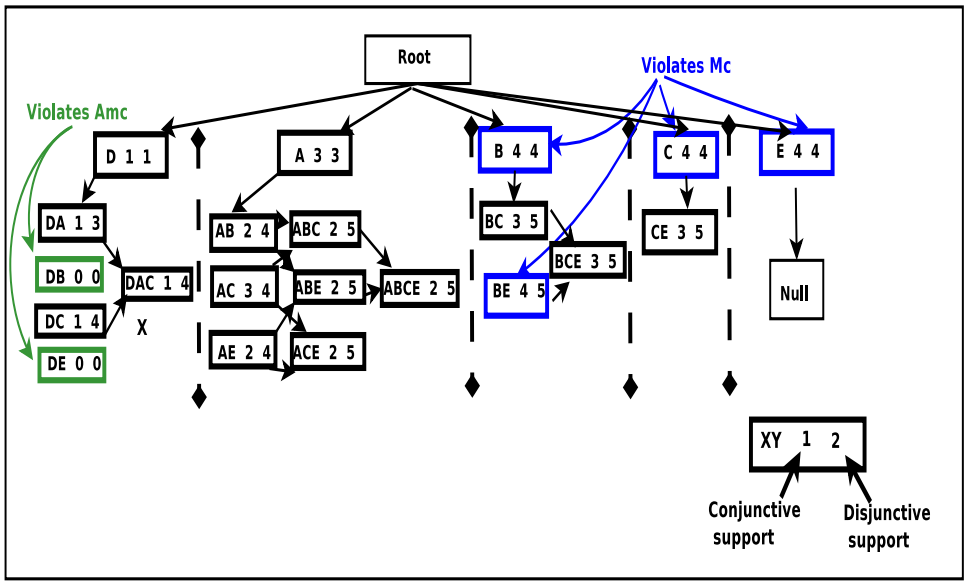

Figure 3: Building the tree while pushing Amc and Mc.

ning to extract the $\mathcal{RCP}$ set. Our experiments were carried out on a PC equipped with a 2.40 GHz Intel Core TM *i*3 processor and 2.92 GB of main memory, running the Linux Ubuntu 10.04. Scalability is an important criteria for constrained itemset mining approaches. Our CORI algorithm demonstrates good scalability versus increasing datasets size according to two dimensions: the number of transactions $|\mathcal{T}|$ and the number of items $|\mathcal{I}|$. While GMJP reached a point where it consumed little meadows seven times more CPU-time than CORI. Table 3 reported our results, running times were averaged over 5 executions. As example, while testing the MUSHROOM dataset containing 8, 124 transactions, CORI finishes in 0.432 seconds while GMJP needs 11 seconds. As another experiment example, we tested also on ACCIDENTS with 340, 183 transactions and 468 items. CORI finishes in 47.617 seconds while GMJP needs 793 seconds. GMJP finished the MUSHROOM dataset with about 8K transactions in 20 seconds while CORI finished in 0.278*s*, 52.591*s* and 47.617*s* for the 8K, 100K and 340K transactions datasets, respectively.

We highlight that the obtained results, for the different datasets, are very encouraging. Our designed CORI algorithm outperformed GMJP in the different evaluated bases. This is dedicated to the efficient integration of the monotone and anti-monotone constraints in an early stages of the mining process.

Our aim in our second series of experiments is to evaluate the impact of varying the thresholds of both constraints Amc and Mc. Table 4 presents the results obtained when varying the correlation threshold *minbond* for a fixed *minsupp* threshold. In this experiment, we found that for the MUSHROOM dataset, the *minbond* threshold was chosen to be increasingly selective, from 0.2 to the highest value, equal to 1. This variation affects very slightly the CPU-time and the memory consumption, while the size of the output set drastically decreases from 54, 395 to 126 itemsets. For the CHESS dataset, the size of $\mathcal{RCP}$ set and the CPU-time are very sensitive to the *minbond* variation. For example, a slight variation of *minbond* from 0.40 to 0.45 induces an important decrease of the $\mathcal{RCP}$ set from 5167, 090 to 1560, 073 itemsets. The CPU-time is also lowered from 40.124 to 0.451 seconds when *minbond* decrease from 0.4 to 0.5. Whereas, for the T40I10D100K dataset, the variation of *minsupp* from 2% to 15% induces an increase in the CPU-time of the second and third steps from 60.09 to 79.49 seconds. The size of the output result also increase from 341 to 932 itemsets.

http://archive.ics.uci.edu/ml.

7. CONCLUSION

We proposed, in this paper, the CORI algorithm to mine the set of rare correlated patterns according to the *bond* measure. The key core of our work was the intelligent integration of the two paradigms of monotonicity and anti-monotonicity. The CORI algorithm, it uses a tree structure with a specific node structure and an optimized construction and pruning strategies. Two main features constitute the thrust of CORI algorithm: (*i*) only one scan of the database is performed to build the new transformed dataset. This latter helps to optimize the time and the space needed for supports computing; (*ii*) it offers a resolution of the problem of handling both rarity and correlation constraints. An important perspective for future work is the application of our approach in a real context.

8. REFERENCES

[1] M. Barsky, S. Kim, T. Weninger, and J. Han. Mining flipping correlations from large datasets with taxonomies. In *Proceedings of the 38th International Conference on Very Large Databases - VLDB 2012, Istanbul, Turkey*, pages 370–381, 2012.

[2] N. Ben Younes, T. Hamrouni, and S. Ben Yahia. Bridging conjunctive and disjunctive search spaces for mining a new concise and exact representation of

Table 3: Performance comparison of our Cori approach and the Gmjp approach on UCI benchmarks (time in seconds).

Dataset	Average *minsupp*	Average *minbond*	Average Time Gmjp	Average Time Cori
Mushroom	58%	0.30	20	**0.278**
	40%	0.57	11	**0.432**
T10I4D100K	5%	0.20	163	**52.591**
T40I10D100K	8.2%	0.50	491	**68.491**
Accidents	7.8%	0.50	793	**47.617**

Table 4: Impact of the thresholds variation.

Dataset	*minsupp*	*minbond*	$\lvert\mathcal{RCP}\rvert$	CPU Time Step 1	CPU Time Steps 2 and 3	Avg. Memory Consumption (Ko)
Mushroom	40%	0.2	54, 395	0.20	**0.977**	18, 590
		1	126	0.20	**0.198**	
Chess	50%	0.40	5167, 090	0.068	**40.124**	
		0.45	1560, 073	0.068	**12.127**	
		0.50	162	0.068	**0.451**	13, 556
		0.60	40	0.068	**0.073**	
		1	38	0.068	**0.054**	
T40I10D100K	2%	0.50	341	16	**60.09**	
	11%		889	16	**63.028**	131,516
	15%		932	16	**79.49**	

correlated patterns. In *Proceedings of the 13th International Conference Discovery Science (DS 2010), LNCS, volume 6332, Springer-Verlag, Canberra, Australia*, pages 189–204, 2010.

[3] S. Bouasker and S. Ben Yahia. Inferring knowledge from concise representations of both frequent and rare jaccard itemsets. In *Proceedings of the 24th International Conference on Databases and Expert Systems Applications (DEXA 2013), Prague, Czech Republic*, 2013.

[4] S. Bouasker, T. Hamrouni, and S. Ben Yahia. New exact concise representation of rare correlated patterns: Application to intrusion detection. In *Proceedings of the 16th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2012), Kuala Lumpur, Malaysia*, pages 61–72, 2012.

[5] S. Brin, R. Motwani, and C. Silverstein. Beyond market baskets: generalizing association rules to correlations. In *Proceedings of the ACM-SIGMOD International Conference on Management of Data (SIGMOD 1997), Washington D. C., USA*, pages 265–276, 1997.

[6] B. Ganter and R. Wille. *Formal Concept Analysis*. Springer, 1999.

[7] R. U. Kiran and M. Kitsuregawa. Mining correlated patterns with multiple minimum all-confidence thresholds. In *Proceedings of the 17th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2013)*, pages 295–306, 2013.

[8] Y. Le Bras, P. Lenca, and S. Lallich. Mining classification rules without support: an anti-monotone property of jaccard measure. In *Proceedings of the 14th international conference on Discovery science - DS 2011, Espoo, Finland*, pages 179–193, 2011.

[9] E. Omiecinski. Alternative interest measures for mining associations in databases. *IEEE Transactions on Knowledge and Data Engineering*, 15(1):57–69, 2003.

[10] M. Segond and C. Borgelt. Item set mining based on cover similarity. In *Proceedings of the 15th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2011), Shenzhen, China*, pages 493–505, 2011.

[11] A. Soulet, C. Raissi, M. Plantevit, and B. Crémilleux. Mining dominant patterns in the sky. In *Proceedings of the 11th IEEE International Conference on Data Mining - ICDM 2011, Vancouver, Canada*, pages 655–664, 2011.

[12] A. Surana, R. U. Kiran, and P. K. Reddy. Selecting a right interestingness measure for rare association rules. In *Proceedings of the 16th International Conference on Management of Data (COMAD 2010), Nagpur, India*, pages 115–124, 2010.

[13] L. Szathmary, P. Valtchev, and A. Napoli. Generating rare association rules using the minimal rare itemsets family. *International Journal of Software and Informatics*, 4(3):219–238, 2010.

[14] G. Webb and J. Vreeken. Efficient discovery of the most interesting associations. *ACM Transactions on Knowledge Discovery from Data,*, 2013.