

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2015

Ontology Based Post Mining of Association Rules

P Siva Bhavani¹, M N Hima Bindu², K.Syama SundaraRao³

Assistant Professor, Department of Computer Engineering, Potti Sriramulu Chalavadi MallikarjunaRao College of Engineering College, Vijayawada, India¹

Assistant Professor, Department of Computer Engineering, Potti Sriramulu Chalavadi MallikarjunaRao College of

Engineering College, Vijayawada, , India²

Assistant Professor, Department of Computer Science & Engineering, Prasad V Potluri Siddhartha Institute of

Technology, Vijayawada, India³

ABSTRACT: The Data Mining process enables the end users to analyze, understand and use the extracted knowledge in an intelligent system or to support in the decision-making processes Association refers to the data mining task of uncovering relationships among the data. In Data Mining, the usefulness of association rules is strongly limited by the huge amount of delivered rules. There are a number of ways to overcome this drawback that depends on the statistical information. But none of them guarantee that the rules extracted are interesting for user. Thus we introduced a new approach in which the user knowledge I taken into the consideration to extract the rules through user interactive system. We propose ontologies to in order to improve the integration of user knowledge in the postprocessing task. The user can even edit or validate ontologies. We propose the Rule Schema formalism extending the specification language for user expectations. The rules extracted are pruned and filtered. So that, voluminous set of rules were reduced to several dozens or less.

KEYWORDS: Association rules, knowledge discovery in databases, ontologies, pruning, filters.

I. INTRODUCTION

Association rule mining is considered as one of the most important tasks in Knowledge Discovery in Databases. It aims at discovering implicative tendencies that can be valuable information for the decision-maker. An association rule is an implication $X \rightarrow Y$, described by two interestingness measures support and Confidence, where X and Y are two item sets and $X \cap Y = \phi$. The former, X, is called the antecedent of the rule, and the latter, Y, is called the consequent. Apriori is the first algorithm proposed in the association rule mining field and many other algorithms were derived from it, to extract all association rules satisfying minimum thresholds of support and confidence. It is very well known that mining algorithms can discover a prohibitive amount of association rules. Valuable information is often represented by those rare low support and unexpected association rules which are surprising to the user. So, the more we increase the support threshold, the more efficient the algorithms are and the more the discovered rules are obvious, and hence, the less they are interesting for the user.

As a result, it is necessary to bring the support threshold low enough in order to extract valuable information. Unfortunately, the lower the support is, the larger the volume of rules becomes, making it intractable for a decision-maker to analyze the mining result. Experiments show that rules become almost impossible to use when the number of rules overpasses 100. Thus, it is crucial to help the decision-maker with an efficient technique for reducing the number of rules.

To overcome this drawback, several methods were proposed in the literature. On the one hand, different algorithms were introduced to reduce the number of itemsets by generating closed, maximal or optimal itemsets, and several



(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2015

algorithms to reduce the number of rules, using non-redundant rules, or pruning techniques. On the other hand, postprocessing methods can improve the selection of discovered rules. Different complementary postprocessing methods may be used, like pruning, summarizing, grouping, or visualization. Pruning consists in removing uninteresting or redundant rules. In summarizing, concise sets of rules are generated. Groups of rules are produced in the grouping process; and the visualization improves the readability of a large number of rules by using adapted graphical representations. The representation of user knowledge is an important issue. In the Semantic Web field, ontology is considered as the most appropriate representation to express the complexity of the user knowledge, and several specification languages were proposed.

This paper proposes a new interactive postprocessing approach, ARIPSO (Association Rule Interactive post-Processing using Schemas and Ontologies) to prune and filter discovered rules. First, we propose to use Domain Ontologies in order to strengthen the integration of user knowledge in the postprocessing task. Second, we introduce Rule Schema formalism by extending the specification language for user beliefs and expectations toward the use of ontology concepts. Furthermore, an interactive and iterative framework is designed to assist the user throughout the analyzing task. The interactivity of our approach relies on a set of rule mining operators defined over the Rule Schemas in order to describe the actions that the user can perform.

II. PROPOSED WORK

In the Semantic Web context, the number of available ontologies has been increasing covering a wide domain of applications. This could be a great advantage in an ontology-based user knowledge representation. One of our most important contributions relies on using ontologies as user background knowledge representation. General Impressions (GI), Reasonably Precise Concepts (RPC), and Precise Knowledge (PK)—by the use of ontology concepts.

Interestingness measures represent metrics in the process of capturing dependencies and implications between database items, and express the strength of the pattern association. Since frequent itemset generation is considered as an expensive operation, mining frequent closed itemsets was proposed in order to reduce the number of frequent itemsets. The CLOSET algorithm was proposed in [3] as a new efficient method for mining closed itemsets. CLOSET uses a novel frequent pattern tree (FP-tree) structure, which is a compressed representation of all the transactions in the database. Moreover, it uses a recursive divide-and-conquer and database projection approach to mine long patterns. Another solution for the reduction of the number of frequent itemsets is mining maximal frequent itemsets[1]. The authors proposed the MAFIA algorithm based on depth-first traversal and several pruning methods as Parent Equivalence Pruning (PEP), FHUT, HUTMFI, or Dynamic Recording. However, the main drawback of the methods extracting maximal frequent itemsets is the loss of information because the subset frequency is not available; thus, generating rules is not possible.

Zaki and Hsiao used frequent closed itemsets in the CHARM algorithm [4] in order to generate all frequentclosed itemsets. They used an itemset-tid set search tree and pursued with the aim of generating a small non-redundantrule set [5]. To this goal, the authors first found minimal generator for closed itemsets, and then, they generated non-redundant association rules using two closed itemsets. Pasquier et al. [2] proposed the Close algorithm in order to extract association rules. Close algorithm is based on a new mining method: pruning of the closed set lattice (closed itemset lattice) in order to extract frequent closed itemsets. Association rules are generated starting from frequent itemsetsgenerated from frequent closed itemsets. Nevertheless, Zaki and Hsiao [4] proved that their algorithm CHARM outperforms CLOSET, Close, and Mafia algorithms.

Li [6] proposed optimal rules sets, defined with respect to interestingness metric. An optimal rule set contains all rules except those with no greater interestingness than one of its more general rules. A set of reduction techniques for redundant rules was proposed and implemented in [6]. The developed techniques are based on the generalization/specification of the antecedent/consequent of the rules and they are divided in methods for multiantecedent rules and multi consequent rules Interestingness measures were proposed in order to discover only those association rules that are interesting



(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2015

according to these measures. They have been divided into objective measures and subjective measures. Objective measures depend only on data structure. Subjective measures were proposed to integrate explicitly the decision-maker knowledge and to offer a better selection of interesting association rules. Subjective measures are classified as unexpectedness—a pattern is interesting if it is surprising to the user—and actionability—a patternis interesting if it can help the user take some actions

Imielinski et al. [7] proposed a query language for association rule pruning based on SQL, called M-SQL. Itallows imposing constraints on the condition and/or the consequent of the association rules. Ng et al. [8] proposed an architecture for exploratory mining of rules. the lack of user exploration and control, the rigid notion of relationship, and the lack of focus. In order to overcome these problems, Ng et al. proposed a new query language called Constrained Association Query and they pointed out the importance of user feedback and user flexibility in choosing interestingness metrics.

Another related approach was proposed by An et al. in [9] where the authors introduced domain knowledge in order to prune and summarize discovered rules. a new methodology was proposed in [10] toprune and organize rules with the same consequent. The authors suggested transforming the database in an associationrule base in order to extract second-level association rules. Called metarules, the extracted rules $r_1 \rightarrow r_2$ express relations between the two association rules and help pruning/grouping discovered rules.

In the early 1990s, an ontology was defined by Gruber as a formal, explicit specification of a shared conceptualization [11]. The formal notion denotes the idea that machines should be able to interpret an ontology. Moreover, explicit refers to the transparent definition of ontology elements. Finally, shared outlines that an ontology brings together some knowledge common to a certain group, and not individual knowledge.

Several other definitions are proposed in the literature. For instance, in [12], an ontology is viewed as a logical theory accounting for the intended meaning of a formal vocabulary, and later, in 2001, Maedche and Staab proposed a moreartificial-intelligence-oriented definition. Thus, ontologies are described as data schemas, providing a controlled vocabulary of concepts, each with an explicitly defined and machine processable semantics [13].

Ontologies, introduced in data mining for the first time in early 2000, can be used in several ways [14]: Domain and Background Knowledge Ontologies, Ontologies for Data Mining Process, or Metadata Ontologies. Background Knowledge Ontologies organize domain knowledge and play important roles at several levels of the knowledge discovery process. Ontologies for Data Mining Process codify mining process description and choose the most appropriate task according to the given problem; while Metadata Ontologies describe the construction process of items. In this paper, we focus on Domain and Background Knowledge Ontologies. The first idea of using Domain Ontologies was introduced by Srikant and Agrawal with the concept of Generalized Association Rules (GAR) [15]. The authors proposed taxonomies of mined data (an is-a hierarchy) in order to generalize/specify rules. In [16], it is suggested that an ontology of background knowledge can benefit all the phases of a KDD cycle described in CRISP-DM. The role of ontologies is based on the given mining task and method, and on data characteristics. Related to Generalized Association Rules, the notion of raising was presented in [17]. Raising is the operation of generalizing rules (making rules more abstract) in order to increase support in keeping confidence high enough. Thisallows for strong rules to be discovered and also to obtain sufficient support for rules that, before raising, would nothave minimum support due to the particular items they referred to. The difference with Generalized AssociationRules is that this solution proposes to use a specific level for raising and mining. The authors developed an algorithm, called GART [18], which, having several taxonomies over attributes, uses iterativelyeach taxonomy in order to generalize rules, and then, prunes redundant rules at each step.

The Algorithm GART

We analyzed the structure of the association rules generated by algorithms that do not use taxonomies. The results of the analysis show us that it is possible to generalize association rules using taxonomies. In Fig. 1 we show how the association rules can be generalized.



(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2015

First we changed the items t-shirt and short of the rules short & slipper) cap, sandal & short) cap, sandal & t-shirt) cap and slipper & t-shirt) cap by the item light clothes (which represents a generalization). This change generated two rules light clothes & slipper) cap and two rules light clothes & sandal) cap. Next, we pruned the repeated generalized rules, maintaining only the two rules: light clothes & slipper) cap and light clothes & sandal) cap.

The two rules generated by the Step 1 (Fig. 1) were generalized again. We changed the items slipper and sandal by the item light shoes (which represented another generalization) generating two rules light clothes & light shoes) cap. Then we pruned the repeated generalized rules again, maintaining only one generalized association rule: light clothes & light shoes) cap. Due to the possibility of generalization of the association rules (Fig. 1), we propose an algorithm to generalize association rules. The proposed algorithm is illustrated in Fig. 2. We called the proposed algorithm of GART (Generalization of Association Rules using Taxonomies).



Figure 1. Generalization of association rules using two taxonomies.



Figure 2. The proposed algorithm to generalize association rules.

The proposed algorithm just generalizes one side of the association rules -LHS or RHS (after to look to a small set of rules without taxonomies, the user decides which side will be generalized). First, we grouped the rules in subsets that present equal antecedents or consequents. If the algorithm were used to generalize the left hand side of the rules (LHS), the subsets would be generated using the equals consequents (RHS). If the algorithm were used to generalize the right hand side of the rules (RHS), the subsets would be generated using the equals consequents using the equal antecedents (LHS). Next, we used the taxonomies to generalize each subset (as illustrated in Fig. 1). In the final algorithm we stored the rules in a set of generalized association rules.

In the final algorithm, we also calculated the Contingency Table for each generalized association rules to get more information about the rules. The Contingency Table of a rule represents the coverage of the rule with respect to the database used in its mining with the calculation of the Contingency Tablewe finished the algorithm.



(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2015

A very recent approach, [19], uses ontologies in a preprocessing step. Several domain-specific and user-defined constraints are introduced and grouped into two types: pruning constraints, meant to filter uninteresting items, and abstraction constraints permitting the generalization of items toward ontology concepts. The data set is first preprocessed according to the constraints extracted from the ontology, and then, the data mining step takes place.

The item-relatedness filter was proposed by Natarajan and Shekar [20]. Starting from the idea that the discovered rules are generally obvious, they introduced the idea of relatedness between the items measuring their similarity according to item taxonomies. This measure computes the relatedness of all the couples of rule items. We can notice that we can compute the relatedness for the items of the condition or/and the consequent, or between the condition and the consequent of the rule.

While Natarajan and Shekar measure the item-relatedness of an association rule, Garcia et al. developed in [21] and extended in [22] a novel technique called Knowledge Cohesion (KC). The proposed metric is composed of twonew ones: Semantic Distance (SD) and Relevance Assessment (RA). The SD one measures how close two items aresemantically, using the ontology—each type of relation being weighted differently. The numerical value RAexpresses the interest of the user for certain pairs of items in order to encourage the selection of rules containing thosepairs. In this paper, the ontology is used only for the SD computation, differing from our approach which uses ontologies for Rule Schemas definition. Moreover, the authors propose a metric-based approach for itemset selection, while we propose a pruning/filtering schemas-based method of association rules.

Framework Description:





The proposed approach is composed of two main parts. First, the knowledge base allows formalizing user knowledge and goals. Domain knowledge offers a general view over user knowledge in database domain, and user expectations express the prior user knowledge over the discovered rules. Second, the postprocessing task consists in applying iteratively a set of filters over extracted rules in order to extract interesting rules: minimum improvement constraint filter, item-relatedness filter, rule schema filters/pruning. The novelty of this approach resides in supervising theknowledge discovery process using two different conceptual structures for user knowledge representation: one or several ontologies and several rule schemas generalizing general impressions, and proposing an iterative process.

Interactive Postmining Process



Figure 4. Interactive process description



(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2015

Taking into account his/her feedbacks, the user is able to revise his/her expectations in function of intermediate results. Several steps are suggested to the user in the framework as follows

1. ontology construction—starting from the database, and eventually, from existing ontologies, the user develops an ontology on database items;

2. defining Rule Schemas (as GIs and RPCs)—the user expresses his/her local goals and expectations concerning the association rules that he/she wants to find;

3. choosing the right operators to be applied over the rule schemas created, and then, applying the operators;

4. visualizing the results—the filtered association rules are proposed to the user;

5. selection/validation—starting from these preliminary results, the user can validate the results or he/she can revise his/her information;

6. We propose to the user two filters already existing in the literature and detailed in Section 5.5. These two filters can be applied over rules whenever the user needs them with the main goal of reducing the number of rules; and

7. The interactive loop permits to the user to revise the information that he/she proposed. Thus, he/she can return to step 2 in order to modify the rule schemas, or he/she can return to step 3 in order to change the operators. Moreover, in the interactive loop, the user could decide to apply one of the two predefined filters discussed in step 6

III. EXPERIMENTAL RESULTS

To describe the ontology, we propose to use the Web Semantic representation language, OWL-DL Based on description logics, OWL-DL language permits, along with the ontological structure, to create restriction concepts using necessary and sufficient conditions over other concepts. Also, we use the Protégé [50] software to edit the ontology and validate it.

The Jambalaya [51] environment was used for ontology graph exploration. During several exchanges with the Nantes Habitat expert, she developed an ontology composed of two main parts, a sample being presented in Fig. 6. The ontology has seven depth levels, a total of 130 concepts among which 113 are primitive concepts and 17 are restriction concepts. Concerning siblings, the concepts have a mean of six child concepts, with a maximum of 13 child concepts. Moreover, two data properties are introduced.

The first part of the ontology is a database item organization with the root defined by the Attribute concept, grouping subsumed concepts. The items are organized among the question topic in the Nantes Habitat questionnaire. The second hierarchy Topics regroups all restriction concepts created by the expert using necessary and sufficient conditions over primitive concepts. Moreover, the subsumption relation (\leq) is completed by the relation *hasAnswer* associating the Attribute concepts to an integer from f1; 2; 3; 4g, simulating the relation attribute value in the database.



Figure 5. Ontology structure visualized with Jambalaya Protégé plug-in.



(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2015

For instance, let us consider the restriction concept SatisfactionDistrict. In natural language, it expresses the satisfaction answers of clients in the questions concerning the district. In other words, an item is instantiated by the SatisfactionDistrict concept if it represents a question between q1 and q14, subsumed by the District concept, witha satisfied answer (1 or 2).

The SatisfactionDistrict restriction concept is described using description logicslanguage by: SatisfactionDistrict = District $\Pi \exists$ has Answer:1 OR hasAnswer:2

 ▼ ■ Topic ■ DissatisfactionCalmDistrict ■ DissatisfactionPrice 	▲ _	CalmDistrict (hasAnswer has 3) or (hasAnswer has 4)		
 SatisfactionConfortApartment SatisfactionDistrict 				

Figure 6. Restriction concept construction using necessary and sufficient conditions in Protégé

Ontology-Database Mapping

As a part of rule schemas, ontology concepts are mapped to database items. Thus, several connections between ontology and database can be designed. Due to implementation requirements, the ontology and the database are mappedthrough instances. The ontology-database connection is made manually by the expert. a manually connection could be very time-consuming. That is why integrating an automatic ontology construction plug-in in our tool is one of our principal perspectives. Thus, using the simplest ontology-database mapping, the expert directly connected one instance of the ontology to an item (semantically, the nearest one).

A second type of connection implies connecting concepts of the Topic hierarchy to the database. Let us consider the restriction concept DissastisfactionCalmDistrict (Fig.6). In natural language, it is defined as all the concepts, subsumed by CalmDistrict and with a dissatisfied answer. The DissastisfactionCalmDistrict restriction concept isdescribed by the expert using description logics language by:

DissastisfactionCalmDistrict = CalmDistrict $\Pi \exists$ hasAnswer:3 OR hasAnswer:4:

TABLE-1	Pruning	Rule	Schemas
---------	---------	------	---------

Rule Schema
$RS_5 :< EntryHall \rightarrow CloseSurrounding >$
$RS_6 :< Stairwell \rightarrow EntryHall >$
$RS_7 :< CloseSurrounding > \rightarrow < EntryHall >$
$RS_8 :< EntryHall \rightarrow Stairwell >$
$RS_9 :< CommonAreas \rightarrow GarbageRoom >$
$RS_{10} :< Technical Maintenance \rightarrow Technical Maintenance >$

TABLE-2 Filtering Rule Schemas

Rule Schema	Operator
$RS_1 :< Dissatisfaction Price >$	$C(RS_1)$
$RS_2 :< DissatisfactionCalmDistrict >$	$C(RS_2)$
$RS_3 :< DissasisfactionPrice$	C
DissatisfactionCommonAreas >	(RS_3)
$RS_4 :< DissasisfactionPrice \rightarrow$	U_p/E
DissatisfactionCommonAreas >	(RS_4)

TABLE-3 Pruning Rate for Each Filter Combination



(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2015

Nb	MICF	IRF	PRS	Rule number
1				358,072 (100%)
2	Х			27,602 (7.7%)
3		Х		103,891 (29%)
4			X	207,196 (57%)
5	Х	Х		16,473 (4.6%)
6	Х		X	21,822 (7.7%)
7		Х	X	73,091 (20%)
8	Х	X	X	13,382 (3.7%)

TABLE-4 Notation Meaning

Nb	The <i>id</i> of the filter combination
MICF	Minimum improvement constraint filter
IRF	Item-relatedness filter
PRS	Pruning with Rule Schemas
Rule number	The number of rules remaining
	after filter application

Results:

This example proposes to present the efficiency of our new approach concerning the reduction of the number of rules. To this end, we propose to the expert to test the four filters, the pruning filters-MICF, IRF, and pruning rule schemas, the selection filters—rule schema filters. The expert could use each filter separately and in several combinations in order to compare the results and validate them.

Hence, the expert proposed a set of pruning rule schemas (Table 1) and a set of filtering rule schemas (Table 2). She constructed these rule schemas during several meetings of testing the new tool and analyzing generated results. At the beginning, the expert is faced to the whole set of 358,072 association rules extracted. In a first attempt, we focus on pruning filters. If the MICF is applied, all the specialized rules not improving confidence are pruned. In Table 3, we can see that the MICF prunes 92.3 percent of rules, being a very efficient filter for redundancy pruning. In addition, IRF prunes 71 percent of rules—these rules implying items close semantically. The third pruning filter, Pruning Rule Schemas, prunes 43 percent of rules. We propose to compare the three pruning filters and the combinations of the pruning filters, as presented in Table 3. The first column is the reference for our experiments. The rates of number of rules remaining after the three filters are used separately are presented in columns 2, 3, and 4. We can note that the MICF filter is the most discriminatory, pruning 92.3 percent of rules, comparing to other two ones pruning 71 percent and, respectively, 43 percent of rules. We can also note that combining the first two filters, MICF and IRF, the pruning is more powerful than combining the first one with the third one. Nevertheless, applying the three filters over the set of the association rules implies a rule reduction of 96.3 percent.



(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2015

Nb	$C(RS_1)$	$C(RS_2)$	$C(RS_3)$	$U_p(RS_4)$	$E(RS_4)$
1	185	1,008	96	1399	98
	(100%)	(100%)	(100%)	(100%)	(100%)
2	92	361	50	462	48
	(49%)	(35%)	(52%)	(33%)	(48.9%)
3	39	162	39	401	3
	(21%)	(16%)	(40%)	(28.7%)	(3.1%)
4	107	472	20	238	96
	(57%)	(46%)	(20%)	(17%)	(6.9%)
5	28	77	28	187	3
	(15%)	(7.6%)	(29%)	(13.4%)	(3.1%)
6	56	231	14	154	48
	(30%)	(22%)	(14%)	(11%)	(48.9%)
7	3	3	3	24	3
	(1.6%)	(0.2%)	(3.1%)	(1.7%)	(3.1%)
8	3	3	3	11	3
	(1.6%)	(0.2%)	(3.1%)	(0.8%)	(3.1%)

TABLE-5 Rates for Rule Schema Filters Applied after the Other Three Filter Combinations

However, applying the most reducing combination number 8 (Table 3), the expert should analyze 13,382 rules which is impossible manually. Thus, other filters should be applied. The expert was interested in the dissatisfaction phenomena, presented by answers 3 and 4 in the questionnaire. The expert is interested in applying all the rule schemas with the corresponding operator (Table 2) for each combination of the first three filters presented in Table 3. Table 5 presents the number of rules filtered by each rule schema.

In Table 5, the first column Nb represents the identification each filter combination as denoted in Table 4. We can note that the rule schema filters are very efficient. Moreover, studying the dissatisfaction of the clients improves the filtering power of the rule schemas. Let us consider the second rule schema. Applied over the initial set of 358,072 association rules with the conforming operator, it filters 1,008 rules representing 0.28 percent of the complete set. But it is obvious that it is very difficult for an expert to analyze a set of rules of the order of thousands of rules. Thus, we can note the importance of the pruning filters, the set of rules extracted in each case having less than 500 rules. We can also note that the IRF filter is more powerful than the other pruning filters, and the combination of two filters at the same time gives remarkable results: on the fifth line, combining MICF with IRF reduces the number of rules to 77 rules; combining IRF with pruning using Rule Schemas the set of rules is reduced to three rules; and. We can also note that in the last two rows, the filters have the same results. We can explain this by the fact that we are working on an incomplete set of rules because of the maximum support threshold that we impose in the mining process. It is very important to note that the quality of the selected rules was certified by the Nantes Habitat expert.



(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2015



Figure 7. Description of the interactive process during the experiment

IV. CONCLUSION

This paper discusses the problem of selecting interesting association rules throughout huge volumes of discovered rules. First, we propose to integrate user knowledge in association rule mining using two different types of formalism: ontologies and rule schemas. We even given the user to edit or validate ontologies. Second, a set of operators, applicable over the rule schemas, is proposed in order to guide the user throughout the post processing step. Thus, several types of actions, as pruning and filtering, are available to the user. Finally, the interactivity of our ARIPSO framework, relying on the set of rule mining operators, assists the user throughout the analyzing task and permits him/her an easier selection of interesting rules by reiterating the process of filtering rules.

By applying our new approach over a voluminous questionnaire database, we allowed the integration of domain expert knowledge in the post processing step in order to reduce the number of rules to several dozens or less. Moreover, the quality of the filtered rules was validated by the expert throughout the interactive process.

REFERENCES

[1] D. Burdick, M. Calimlim, J. Flannick, J. Gehrke, and T. Yiu, "Mafia: A Maximal Frequent Itemset Algorithm," IEEE Trans.Knowledge and Data Eng., vol. 17, no. 11, pp. 1490-1504, Nov. 2005.

[2] N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal, "Efficient Mining of Association Rules Using Closed Itemset Lattices," Information Systems, vol. 24, pp. 25-46, 1999.

[3] J. Pei, J. Han, and R. Mao, "Closet: An Efficient Algorithm for Mining Frequent Closed Itemsets," Proc. ACM SIGMOD Workshop Research Issues in Data Mining and Knowledge Discovery, pp. 21-30,2000.

[4] M.J. Zaki and C.J. Hsiao, "Charm: An Efficient Algorithm for Closed Itemset Mining," Proc. Second SIAM Int'l Conf. Data Mining, pp. 34-43, 2002.

[5] M.J. Zaki, "Generating Non-Redundant Association Rules," Proc. Int'l Conf. Knowledge Discovery and Data Mining, pp. 34-43, 2000.

[6] M.Z. Ashrafi, D. Taniar, and K. Smith, "Redundant Association Rules Reduction Techniques," AI 2005: Advances in Artificial Intelligence – Proc 18th Australian Joint Conf. Artificial Intelligence, pp. 254-263, 2005.

[7] T. Imielinski, A. Virmani, and A. Abdulghani, "Datamine: Application Programming Interface and Query Language for Database Mining," Proc. Int'l Conf. Knowledge Discovery and Data Mining (KDD), pp. 256-262, <u>http://www.aaai.org/Papers/KDD/</u> 1996/KDD96-042.pdf, 1996.

[8] R.T. Ng, L.V.S. Lakshmanan, J. Han, and A. Pang, "Exploratory Mining and Pruning Optimizations of Constrained Associations Rules," Proc. ACM SIGMOD Int'l Conf. Management of Data, vol. 27, pp. 13-24, 1998.

[9] A. An, S. Khan, and X. Huang, "Objective and Subjective Algorithms for Grouping Association Rules," Proc. Third IEEE Int'l Conf. Data Mining (ICDM '03), pp. 477-480, 2003.

[10] A. Berrado and G.C. Runger, "Using Metarules to Organize and Group Discovered Association Rules," Data Mining and Knowledge Discovery, vol. 14, no. 3, pp. 409-431, 2007.

[11] T.R. Gruber, "A Translation Approach to Portable Ontology Specifications," Knowledge Acquisition, vol. 5, pp. 199-220, 1993.

- [12] N. Guarino, "Formal Ontology in Information Systems," Proc. First Int'l Conf. Formal Ontology in Information Systems, pp. 3-15, 1998
- [13] A. Maedche and S. Staab, "Ontology Learning for the Semantic Web," IEEE Intelligent Systems, vol. 16, no. 2, pp. 72-79, Mar. 2001.