

Interactive Post Mining Association Rules using Cost Complexity Pruning and Ontologies KDD

Ch. Raja Ramesh,
Scholar, K.L.University.
Associate Professor in
CSE, SVEC,
Tadepalligudem.

K V V Ramana
Assistant Professor in
CSE, SVEC
Tadepalligudem.

K. Raghava Rao,
Professor in CSE,
KL University, Guntur.

C V Sastry
Professor In CSE
RIT, Yanam.

ABSTRACT

In data mining, association rule mining is very strong and limited by the huge amount of delivered rules, because of these so many problems facing to implementation. To overcome these drawbacks, several methods were proposed in the literature such as item sets concise, redundancy reduction, and post processing. Based on statistical information by using these methods the extracted rules may not be useful for the user. Thus, it is crucial to help the decision-controller with an efficient post processing step in order to reduce the number of rules. In this paper we have implemented a new interactive approach using cost complexity pruning and filter discovered rules, by using this approach we have to reduce the tree size with minimum number of error of validation set.

Key Words: clustering, classification, and association rules, knowledge discover database

1. INTRODUCTION

Association rule mining, introduced in [1], is one of the most important tasks Knowledge Discovery in large Databases [2]. It aims to discover the implicative tendencies between the frequent data items in the databases that can be valuable information for the decision-controller.

An association rule is defined as the implication form of $X \rightarrow Y$, described by using two measures support and confidence, where X and Y are the sets of items and $X \cap Y = \emptyset$. Apriori [1] is the first algorithm proposed in the association rule mining field and many other algorithms were derived from it, it proposes 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; for instance, thousands of rules are extracted from a database of several dozens of attributes and several hundreds of transactions. Moreover, if we increase the support threshold, then the efficient algorithms are discovered rules are obviously less; these are interesting for the user. As a result, it is necessary to bring the support threshold low in order to extract valuable information. Unfortunately, the large volume of rules is generated to

become almost impossible to use when the number of rules are overpasses 100. Thus, it is crucial to help the decision-controller 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 we reintroduced to reduce the number of item sets by generating closed [4], maximal [5] or optimal item sets [6], and several algorithms to reduce the number of rules, using non redundant rules [7], [8], or pruning techniques [9]. On the other hand, post processing methods can improve the selection of discovered rules. Different complementary post processing methods may be used, like pruning, summarizing, grouping, or visualization [10]. 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.

1.1. Knowledge Representation

Knowledge-based systems have a computational model of some domain of interest in which symbols serve as surrogates for real world domain artifacts, such as physical objects, events, relationships, etc. The *domain of interest* can cover any part of the real world or any hypothetical system about which one desires to represent knowledge for computational purposes. A knowledge-based system maintains a *knowledge base* which stores the symbols of the computational model in form of statements about the domain, and it performs *reasoning* by manipulating these symbols. Applications can base their decisions on domain-relevant questions posed to a knowledge base.

1.2. Association Rule Notations And Definitions

The association rule mining task can be stated as follows: let $I = \{i_1, i_2, i_3, \dots, i_n\}$ be a set of literals, called items. Let $D = \{t_1, t_2, t_3, \dots, t_n\}$ be a set of transactions over I . A nonempty subset of I is called item set and is defined as $X = \{i_1, i_2, i_3, \dots, i_k\}$. In short, item set X can also be denoted as $X = i_1, i_2, i_3, \dots, i_n$. For an item set, the number of items

is called length of the item set and an item set of length k is referred to as k -itemset. Each transaction t_i contains an item set $i_1, i_2, i_3, \dots, i_k$ with a variable k number of items for each t_i

Definition 1. Let X is subset of I and T subset of D . We define the set of all transactions that contain the item set X as:

$$t : I \rightarrow D, t(X) = \{ t \in D \mid X \subseteq t \}$$

Similarly, we describe the item sets contained in all the transactions T by:

$$i : D \rightarrow I, i(T) = \{ x \in I \mid \forall t \in T, x \in t \}$$

Definition 2. An association rule is an implication $X \rightarrow Y$, where X and Y are two item sets and $X \cap Y = \emptyset$. The former, X , is called the antecedent of the rule, and the latter, Y , is called the consequent.

A rule $X \rightarrow Y$ is described using two important statistical factors:

The support of the rule, defined as $\text{supp}(X \rightarrow Y) = \text{supp}(X \cup Y) = |t(X \cup Y)|$, is the ratio of the number of transactions containing $X \cup Y$. If $\text{supp}(X \rightarrow Y) = s$, $s\%$ of transactions contains the item set $X \cup Y$

The confidence of the rule, defined as $\text{conf}(X \rightarrow Y) = \frac{\text{supp}(X \rightarrow Y)}{\text{supp}(X)} = \frac{\text{supp}(X \cup Y)}{\text{supp}(X)} = c$,

is the ratio ($c\%$) of the number of transactions that, containing X , also contains Y .

1.3. User-Driven Association Rule Mining

Interestingness measures were proposed in order to discover only those association rules that are interesting according to these measures. They have been divided into objective measures and subjective measures. Objective measures depend only on data structure. Many survey papers summarize and compare the objective measure definitions and properties [13], [14]. Unfortunately, being restricted to data evaluation, the objective measures are not sufficient to reduce the number of extracted rules and to capture the interesting ones. Several approaches integrating user knowledge have been proposed. In addition, subjective measures were proposed to integrate explicitly the decision-maker knowledge and to offer a better selection of interesting association rules. Silbershatz and Tuzilin [3] proposed a classification of subjective measures in unexpectedness—a pattern is interesting if it is surprising to the user—and actionability—a pattern is interesting if it can help the user take some actions.

As early as 1994, in the KEFIR system [15], the key finding and deviation notions were suggested. Grouped in findings, deviations represent the difference between the actual and the expected values. KEFIR defines interestingness of a key finding in terms of the estimated benefits, and potential savings of taking corrective actions that restore the deviation back to its expected value. These

corrective actions are specified in advance by the domain expert for various classes of deviations. Later, Klemettinen et al. [16] proposed templates to describe the form of interesting rules (inclusive templates) and not interesting rules (restrictive templates). The idea of using templates for association rule extraction was reused in [17]. Other approaches proposed to use a rule-like formalism to express user expectations [3], [10], [18], and the discovered association rules are pruned/summarized by comparing them to user expectations. Imielinski et al. [19] proposed a query language for association rule pruning based on SQL, called M-SQL. It allows imposing constraints on the condition and/or the consequent of the association rules. In the same domain of query-based association rule pruning, but more constraints driven, Ng et al. [20] proposed architecture for exploratory mining of rules. The authors suggested a set of solutions for several problems: 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 [21] where the authors introduced domain knowledge in order to prune and summarize discovered rules. The first algorithm uses data taxonomy, defined by user, in order to describe the semantic distance between rules, and in order to group the rules. The second algorithm allows grouping the discovered rules that share at least one item in the antecedent and the consequent.

In 2007, a new methodology was proposed in [22] to prune and organize rules with the same consequent. The authors suggested transforming the database in an association rule base in order to extract second-level association rules. Called meta rules, the extracted rules $r_1 ! r_2$ express relations between the two association rules and help pruning/grouping discovered rules.

2. ONTOLOGIES IN DATA MINING

Ontology describes basic concepts in a domain and defines relations among them. Basic building blocks of ontology design include: classes or concepts, properties of each concept describing various features and attributes of the concept, restrictions on slots (facets).

In knowledge engineering and Semantic Web fields, ontologies have interested researchers since their first proposition in the philosophy branch by Aristotle. Ontologies have evolved over the years from controlled vocabularies to thesauri (glossaries), and later, to taxonomies [23].

In the early 1990s, ontology was defined by Gruber as a formal, explicit specification of a shared conceptualization [24]. By conceptualization, we understand here an abstract model of some phenomenon described by its important concepts. The formal notion denotes the idea that machines should be able to interpret 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 [25], 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 more artificial-intelligence-oriented definition. Thus, ontologies are described as (Meta) data schemas, providing a controlled vocabulary of concepts, each with an explicitly defined and machine processable semantics [11].

Depending on the granularity, four types of ontologies are proposed in the literature: upper (or top level) ontologies, domain ontologies, task ontologies, and application ontologies [25]. Top-level ontologies deal with general concepts; while the other three types deal with domains specific concepts.

Ontologies, introduced in data mining for the first time in early 2000, can be used in several ways [26]: 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) [26]. The authors proposed taxonomies of mined data (an is-a hierarchy) in order to generalize/specify rules. In [27], it is suggested that 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. From business understanding to deployment, the authors delivered a complete example of using ontologies in a cardiovascular risk domain. Related to Generalized Association Rules, the notion of rising was presented in [28]. Rising is the operation of generalizing rules (making rules more abstract) in order to increase support in keeping confidence high enough. This allows for strong rules to be discovered and also to obtain sufficient support for rules that, before raising, would not have minimum support due to the particular items they referred to. The difference with Generalized Association Rules is that this solution proposes to use a specific level for raising and mining.

Another contribution, very close to [26], [27], uses taxonomies to generalize and prune association rules. The authors developed an algorithm, called GART [43], which, having several taxonomies over attributes, uses iteratively each taxonomy in order to generalize rules, and then, prunes redundant rules at each step. A very recent

approach, [30], 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 difference with our approach is that, first; they apply constraints in the preprocessing task, whereas we work in the post processing task. The advantage of the pruning constraints is that it permits to exclude from the start the information that the

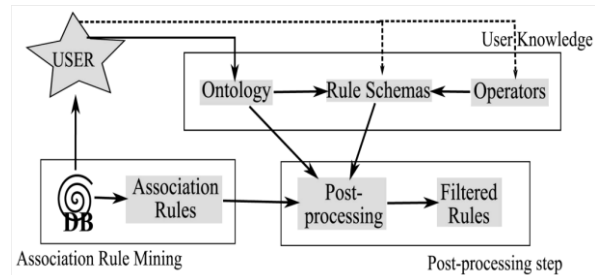


Fig 1: Description of the frame work

User is not interested in, thus permitting to apply the Apriori algorithm to this new database. Let us consider that the user is not sure about which items he/she should prune. In this case, he/she should create several pruning tests, and for each test, he/she will have to apply the Apriori algorithm whose execution time is very high. Second, they use SeRQL in order to express user knowledge, and we propose a more expressive and flexible language for user expectation representation, i.e., Rule Schemas.

The item-relatedness filter was proposed by Natarajan and Shekar [31]. 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 [32] and extended in [33] a novel technique called Knowledge Cohesion (KC). The proposed metric is composed of two new ones: Semantic Distance (SD) and Relevance assessment (RA). The SD one measures how close two items are semantically, using the ontology—each type of relation being weighted differently. The numerical value RA expresses the interest of the user for certain pairs of items in order to encourage the selection of rules containing those pairs. 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 item set selection, while we propose

pruning/filtering schemas based method of association rules.

Pruned subtree: For any tree T , T_2 is a pruned subtree of T if T_2 is a tree with the same root node as T and all nodes of T are also nodes of T_2 . Denote $T \setminus T_2$ if T_2 is a pruned subtree of T .

Given a tree T and a real number α , the cost-complexity risk of T with respect to α is

$$R_\alpha(T) = R(T) + \alpha |\tilde{T}|$$

Where \tilde{T} is the number of terminal nodes and $R(T)$ is the re-substitution risk estimate of T

To generate a sequence of pruned subtree in step 1, the cost – complexity pruning technique developed by Breiman et. At. 1984 is used. In generating the sequence of subtree, only the learning sample is used. Given any real value α_{min} ($\alpha_{min} = 0$ in any SPSS implementation) and an initial tree T_0 , there exists a sequence of real values $-\infty < \alpha_1 = \alpha_{min} < \alpha_1 < \alpha_2 < \alpha_3 \dots < +\infty$ and a sequence of pruned subtree $T_0 > T_1 > \dots T_k$ such that the smallest optimally pruned subtree of T_0 for a given α is

$$T_0(\alpha) = \begin{cases} T_0 & \alpha < \alpha_1 \\ T_0(\alpha_k) = T_k & \alpha_k \leq \alpha < \alpha_{k+1} \quad 1 \leq k < K, \\ T_0(\alpha_K) = T_K & \alpha_K \leq \alpha \end{cases}$$

Where

$$\alpha_{k+1} = \min_{t \in T_k} g_k(t), \quad T_{k+1} = \{t \in T_k$$

$$: g_k(s) > \alpha_{k+1} \text{ for all ancestors of } t\},$$

$$g_k(t) = \begin{cases} \frac{R(t) - R(T_{k,t})}{|\tilde{T}_{k,t}| - 1} & t \in T_k - \tilde{T}_k, \\ +\infty & t \in \tilde{T}_k \end{cases},$$

$\tilde{T}_{k,t}$ is the branch of T_k stemming from node t , and $R(t)$ is the re-substitution risk estimate of node t based on the learning sample.

3. FILTERS

In order to reduce the number of rules, three filters integrate the framework: operators applied over rule schemas, minimum improvement constraint filter [12], and item relatedness filter [31]. Minimum improvement constraint filter [12] (MICF) selects only those rules whose confidence is greater with minimp than the

confidence of any of its simplifications. We are taken the example from [34].

Example. Let us consider the following three association rules:

grape, pear \rightarrow *milk* (Confidence = 85%)

grape \rightarrow *milk* (Confidence = 90%)

pear \rightarrow *milk* (Confidence = 83%)

We can note that the last two rules are the simplifications of the first one. The theory of Bayardo et al. tells us that the first rule is interesting only if its confidence improves the confidence of all its simplifications. In our case, the first rule does not improve the confidence of 90 percent of its simplifications (the second rule), so it is not considered as an interesting rule, and it is not selected. The item-relatedness filter (IRF) was proposed by Shekar and Natarajan [31]. Starting from the idea that the discovered rules are generally obvious, they introduced the idea of relatedness between items measuring their semantic distance in 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.

In our approach we use the relatedness because users are interested to find association between itemset with different functionalities, coming from different domains. This measure is computed as the minimum distance between the condition items and the consequent items as presented hereafter. The distance between each pair of items from the condition and, respectively, the consequent is computed as the minimum path that connects the two items in the ontology, defined as $d(a,b)$. Thus, the item-relatedness (IR) for a rule is defined as the minimum of all the distance computed between the items in the condition and the consequent:

Table 1Example of Questions and meaning

Question number and description	
q1	Convenient transport
q2	City center access
q3	Shopping facilities
q11	Is it safe to take a walk in the district at night
q44	Apartment sound proofing
q47	Apartment ventilation
q48	Apartment standards
q70	Clarity of the documents from Nantes Habitat

$$RA_1: A \rightarrow B$$

$$IR(RA_1)MIN(d_{ij}(a_i, b_j), \forall a_i \in A \text{ and } b_j \in B$$

4. EXPERIMENTAL SETUP

The study is based on a questionnaires data base provided by the Nantes Habitat, dealing with customer satisfaction concerning accommodation. Since 2003, 50000 customers samples are available in database, out of 50000 we have taken 1500 sample, 67 questions are available in the questionnaires with following four possible answers: very satisfaction, quite satisfied, rather not satisfied and dissatisfied coded as {1,2,3,4}.

The above table 1 introduces a sample questions with the meaning of each instance. The item q1=1 that describes customer is very satisfied by the transport in his district. The most interesting rule, we fixed a minimum support 2% and a maximum support of 30% and a minimum confidence of 80% for the association rule mining process, we have to use apriori algorithm in order to extent association rules. For example, the following association rule describes the relationship between questions q2, q3, q47, and the question q70. Thus, if the customers are very satisfied by the access to the city center (q2), the shopping facilities (q3), and the apartment ventilation (q47), then they can be satisfied by the documents received from Nantes Habitat Agency (q70) with a confidence of 90.6%:

$$R1 : q2=1 \text{ } q3=1 \text{ } q47=1 \implies q70=1,$$

$$\text{Support}=13.8\% , \text{ confidence} = 90.6\%$$

5. CONCLUSION

By applying the a new interactive approach using cost complexity pruning and filter discovered rules we allowed to integration of domain expert knowledge in the post processing in order to reduce the number of rules to very less number, based on this the quality of the filtered rules was validated by the expert through the interactive process.

6. REFERENCES

- [1] R. Agrawal, T. Imielinski, and A. Swami, "Mining Association Rules between Sets of Items in Large Databases," Proc. ACM SIGMOD, pp. 207-216, 1993.
- [2] U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, *Advances in Knowledge Discovery and Data Mining*. AAAI/MIT Press, 1996.
- [3] A. Silberschatz and A. Tuzhilin, "What Makes Patterns Interesting in Knowledge Discovery Systems," IEEE Trans. Knowledge and Data Eng. vol. 8, no. 6, pp. 970-974, Dec. 1996.
- [4] M.J. Zaki and M. Ogihara, "Theoretical Foundations of Association Rules," Proc. Workshop Research Issues in Data Mining and Knowledge Discovery (DMKD '98), pp. 1-8, June 1998.
- [5] 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.
- [6] J. Li, "On Optimal Rule Discovery," IEEE Trans. Knowledge and Data Eng., vol. 18, no. 4, pp. 460-471, Apr. 2006.
- [7] M.J. Zaki, "Generating Non-Redundant Association Rules," Proc. Int'l Conf. Knowledge Discovery and Data Mining, pp. 34-43, 2000.
- [8] 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.
- [9] B. Baesens, S. Viaene, and J. Vanthienen, "Post-Processing of Association Rules," Proc. Workshop Post-Processing in Machine Learning and Data Mining: Interpretation, Visualization, Integration, and Related Topics with Sixth ACM SIGKDD, pp. 20-23, 2000.
- [10] B. Liu, W. Hsu, K. Wang, and S. Chen, "Visually Aided Exploration of Interesting Association Rules," Proc. Pacific-Asia Conf. Knowledge Discovery and Data Mining (PAKDD), pp. 380-389, 1999.
- [11] A. Maedche and S. Staab, "Ontology Learning for the Semantic Web," IEEE Intelligent Systems, vol. 16, no. 2, pp. 72-79, Mar. 2001.
- [12] R.J. Bayardo, Jr., R. Agrawal, and D. Gunopulos, "Constraint-Based Rule Mining in Large, Dense Databases," Proc. 15th Int'l Conf. Data Eng. (ICDE '99), pp. 188-197, 1999.
- [13] F. Guillet and H. Hamilton, *Quality Measures in Data Mining*. Springer, 2007.
- [14] P.-N. Tan, V. Kumar, and J. Srivastava, "Selecting the Right Objective Measure for Association Analysis," Information Systems, vol. 29, pp. 293-313, 2004.
- [15] G. Piatetsky-Shapiro and C.J. Matheus, "The Interestingness of Deviations," Proc. AAAI'94 Workshop Knowledge Discovery in Databases, pp. 25-36, 1994.
- [16] M. Klemettinen, H. Mannila, P. Ronkainen, H. Toivonen, and A.I. Verkamo, "Finding Interesting Rules from Large Sets of Discovered Association Rules," Proc. Int'l Conf. Information and Knowledge Management (CIKM),
- [17] E. Baralis and G. Psaila, "Designing Templates for Mining Association Rules," J. Intelligent Information Systems, vol. 9, pp. 7-32, 1997.
- [18] B. Padmanabhan and A. Tuzhilin, "Unexpectedness as a Measure of Interestingness in Knowledge Discovery," Proc. Workshop Information Technology and Systems (WITS), pp. 81-90, 1997.
- [19] 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, <http://www.aaai.org/Papers/KDD/1996/KDD96-042.pdf>, 1996.
- [20] 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.
- [21] 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.
- [22] A. Berrado and G.C. Runger, “Using Meta rules to Organize and Group Discovered Association Rules,” Data Mining and Knowledge Discovery, vol. 14, no. 3, pp. 409-431, 2007.
- [23] M. Us hold and M. Gru” ninger, “Ontologies: Principles, Methods, and Applications,” Knowledge Eng. Rev., vol. 11, pp. 93-155, 1996.
- [24] T.R. Gruber, “A Translation Approach to Portable Ontology Specifications,” Knowledge Acquisition, vol. 5, pp. 199-220, 1993.
- [25] N. Guarino, “Formal Ontology in Information Systems,” Proc. First Int’l Conf. Formal Ontology in Information Systems, pp. 3-15, 1998.
- [26] H. Nigro, S.G. Cisaro, and D. Xodo, Data Mining with Ontologies: Implementations, Findings and Frameworks. Idea Group, Inc., 2007.
- [27] V. Svatek and M. Tomeckova, “Roles of Medical Ontology in Association Mining Crisp-dm Cycle,” Proc. Workshop Knowledge Discovery and Ontologies in ECML/PKDD, 2004.
- [28] X. Zhou and J. Geller, “Raising, to Enhance Rule Mining in Web Marketing with the Use of an Ontology,” Data Mining with Ontologies: Implementations, Findings and Frameworks, pp. 18-36, Idea Group Reference, 2007.
- [29] M.A. Domingues and S.A. Rezende, “Using Taxonomies to Facilitate the Analysis of the Association Rules,” Proc. Second Int’l Workshop Knowledge Discovery and Ontologies, held with ECML/PKDD, pp. 59-66, 2005.
- [30] A. Bellandi, B. Furletti, V. Grossi, and A. Romei, “Ontology- Driven Association Rule Extraction: A Case Study,” Proc. Workshop Context and Ontologies: Representation and Reasoning, pp. 1-10, 2007.
- [31] R. Natarajan and B. Shekar, “A Relatedness-Based Data-Driven Approach to Determination of Interestingness of Association Rules,” Proc. 2005 ACM Symp. Applied Computing (SAC), pp. 551-552, 2005.
- [32] A.C.B. Garcia and A.S. Vivacqua, “Does Ontology Help Make Sense of a Complex World or Does It Create a Biased Interpretation?” Proc. Sensemaking Workshop in CHI ’08 Conf. Human Factors in Computing Systems, 2008.
- [33] A.C.B. Garcia, I. Ferraz, and A.S. Vivacqua, “From Data to Knowledge Mining,” Artificial Intelligence for Eng. Design, Analysis and Manufacturing, vol. 23, pp. 427-441, 2009.
- [34] Claudia Marinica and Fabrice Guillet “Knowledge-Based Interactive Postmining of Association Rules Using Ontologies” eVOL. 22, NO. 6, JUN.