

Mining Association Rules Using Ontologies

P. Sarala^{#1}, S. Jayaprada^{*2}

[#]*Department Of Computer Science and Engineering*

V.R.Siddhartha Engineering College, Vijayawada, Andhra Pradesh, India

^{*}*Department Of Computer Science and Engineering*

V.R.Siddhartha Engineering College, Vijayawada, Andhra Pradesh, India

Abstract

Association rule mining is considered as one of the most important tasks in Knowledge Discovery in Databases. Among sets of items in transaction databases, it aims at discovering implicative tendencies that can be valuable information for the decision-maker. The rules generated by the existing methods are in more number. To reduce the number of rules several post processing methods and many techniques were developed but they are not effective. This paper aims to develop a new frame work called Mining Interest Rules Using Ontologies for extracting association rules based on user interest and also implementing a real time web semantic engine using an extended robust framework.

Keywords- *Association Rules, Association Rule Mining, Ontology, correlation measures, user constraints, Web Ontology Language.*

Introduction

Data mining is the nontrivial process of identifying valid, novel, potentially useful and ultimately understandable patterns from data. One important topic in data mining is concerned with the discovery of interesting association rules. Association rules mining allows nonsupervised discovery of implicative and interesting tendencies in databases. An association rule $a \rightarrow b$ implies the presence of the itemset b when an itemset a occurs in a database transaction. Apriori [3] is the first algorithm proposed to extract all association rules satisfying minimum thresholds of support and confidence. If the support threshold is low then we can extract more valuable information. But usually rules are high.

To reduce the number of rules several post processing methods were developed using nonredundant rules or pruning techniques such as pruning, summarizing, grouping or visualization based on statistical information in the database. Many methods such as Rule deductive method [4], Stream Mill Miner, Data Stream Management Systems [2], Partitioning around medoids clustering technique [5], Constraint-based Multi-level Association Rules with an ontology support [1] were developed but which are not effective.

This paper is implementing a new framework to reduce number of rules based on user interest. User constraints can be categorized as succinct constraints that can be pushed into the initial data selection process at the start of mining. Monotonic constraints can be checked and once satisfied not to do more constraint checking at their further pattern growth. Antimonotonic constraints can be pushed deep into the mining process to restrain pattern growth. Also instead of statistical measures we can use correlation measures such as lift, cosin and overall confidence and this work is extended to implement a real time web semantic engine using an Association Rule Interactive post-Processing using Schemas and Ontologies framework.

Methodology

Our proposed system 'Mining Interest Rules Using Ontologies' consists of the following steps:

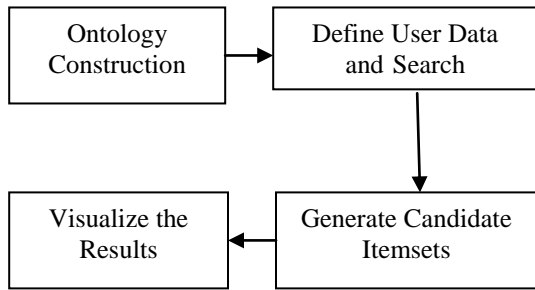


Figure1: Mining Interest Rules Using Ontologies

Input Ontology: Ontology construction for a supermarket dataset, to describe the ontology the framework uses the Web Semantic representation language, OWL.

The ontology representation of supermarket dataset which is given as input.

```

<Departments>
  <department_1>
    <Itemset_1>
      <I2>baby needs</I2>
      <I3>bread and cake</I3>
      <I4>baking needs</I4>
      <I5>juice-sat-cord-ms</I5>
      <I6>tea</I6>
      <I7>biscuits</I7>
      <I10>canned vegetables</I10>
      <I14>cleaners-polishers</I14>
      <I15>coffee</I15>
      <I16>sauces-gravy-pkle</I16>
      <I17>confectionary</I17>
      <I19>dishcloths-scour</I19>
      <I21>frozen foods</I21>
      <I22>razor blades</I22>
      <I29>party snack foods</I29>
      <I30>tissues-paper prd</I30>
      <I31>>wrapping</I31>
      <I39>mens toiletries</I39>
      <I48>cheese</I48>
      <I50>milk-cream</I50>
      <I53>margarine</I53>
      <I55>small goods</I55>
      <I68>fruit</I68>
      <I71>vegetables</I71>
    </Itemset_1>
    <Itemset_2>
      <I86>stationary</I86>
      <I92>bake off products</I92>
    </Itemset_2>
    <Itemset_3>
      <I68>fruit</I68>
      <I86>stationary</I86>
  </department_1>
  </Departments>
  
```

Figure2: Ontological representation of supermarket dataset

Define User Data and Search Ontology: User will specify the particular item on which he interested. Searching for the user specified (interest) item in ontology.

Here the user is searching for the item milk-cream on which he/she interest.

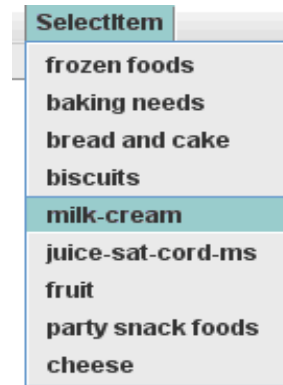


Figure3: List of items in dataset

Generate Candidate Itemsets: Generating candidate itemsets.

Visualize the Results: Visualization of the Results in the form of rules.

The rules/results are generated according to the user selected item ‘milk-cream’.

SelectItem
1. juice-sat-cord-ms=t milk-cream=t total=high 82 ==> bread and cake=t 80 <conf(0.98)> lift(1.33) lev(0.04) [19] convr(7.23)
3. frozen foods=t milk-cream=t total=high 98 ==> bread and cake=t 95 <conf(0.97)> lift(1.32) lev(0.05) [22] convr(6.48)
5. baking needs=t juice-sat-cord-ms=t milk-cream=t fruit=t 88 ==> bread and cake=t 85 <conf(0.97)> lift(1.31) lev(0.04) [20] convr(5.82)
7. biscuits=t frozen foods=t milk-cream=t total=high 86 ==> bread and cake=t 83 <conf(0.97)> lift(1.31) lev(0.04) [19] convr(5.68)
8. party snack foods=t milk-cream=t total=high 85 ==> bread and cake=t 82 <conf(0.96)> lift(1.31) lev(0.04) [19] convr(5.62)
16. juice-sat-cord-ms=t frozen foods=t milk-cream=t fruit=t 89 ==> bread and cake=t 85 <conf(0.96)> lift(1.3) lev(0.04) [19]
22. biscuits=t milk-cream=t total=high 102 ==> bread and cake=t 97 <conf(0.95)> lift(1.29) lev(0.04) [21] convr(4.5)
23. sauces-gravy-pkle=t milk-cream=t total=high 81 ==> bread and cake=t 77 <conf(0.95)> lift(1.29) lev(0.03) [17] convr(4.29)
25. milk-cream=t total=high 121 ==> bread and cake=t 115 <conf(0.95)> lift(1.29) lev(0.05) [26] convr(4.57)
27. baking needs=t juice-sat-cord-ms=t biscuits=t frozen foods=t milk-cream=t 80 ==> bread and cake=t 76 <conf(0.95)> lift(1.29) lev(0.03) [17] convr(4.23)
29. baking needs=t milk-cream=t total=high 92 ==> bread and cake=t 87 <conf(0.95)> lift(1.29) lev(0.04) [19] convr(4.06)
31. milk-cream=t fruit=t total=high 89 ==> bread and cake=t 84 <conf(0.94)> lift(1.28) lev(0.04) [18] convr(3.92)
32. juice-sat-cord-ms=t biscuits=t party snack foods=t milk-cream=t 87 ==> bread and cake=t 82 <conf(0.94)> lift(1.28) lev(0.04) [18] convr(3.84)
34. biscuits=t frozen foods=t party snack foods=t milk-cream=t 86 ==> bread and cake=t 81 <conf(0.94)> lift(1.28) lev(0.04) [17] convr(3.79)
35. baking needs=t biscuits=t frozen foods=t milk-cream=t fruit=t 86 ==> bread and cake=t 81 <conf(0.94)> lift(1.28) lev(0.04) [17] convr(3.79)
36. baking needs=t laundry needs=t milk-cream=t 85 ==> bread and cake=t 80 <conf(0.94)> lift(1.28) lev(0.04) [17] convr(3.75)
37. juice-sat-cord-ms=t tissues-paper prd=t milk-cream=t 84 ==> bread and cake=t 79 <conf(0.94)> lift(1.28) lev(0.03) [17] convr(3.7)
40. baking needs=t frozen foods=t milk-cream=t fruit=t 100 ==> bread and cake=t 94 <conf(0.94)> lift(1.28) lev(0.04) [20] convr(3.78)
43. baking needs=t juice-sat-cord-ms=t biscuits=t milk-cream=t 89 ==> bread and cake=t 83 <conf(0.94)> lift(1.28) lev(0.04) [20] convr(3.74)

SelectItem
94. biscuits=t,confectionary=t,milk-cream=t 93 ==> bread and cake=t 86 <conf:(0.92)> lift:(1.26) lev:(0.04) [17] convr:(3.08)
96. biscuits=t,party snack foods=t,milk-cream=t,fruit=t 93 ==> bread and cake=t 86 <conf:(0.92)> lift:(1.26) lev:(0.04) [17] convr:(3.08)
98. sauces-gravy-pkls=t,frozen foods=t,milk-cream=t 105 ==> bread and cake=t 97 <conf:(0.92)> lift:(1.26) lev:(0.04) [19] convr:(3.09)
99. confectionary=t,milk-cream=t 118 ==> bread and cake=t 109 <conf:(0.92)> lift:(1.26) lev:(0.04) [22] convr:(3.12)
103. laundry needs=t,milk-cream=t 103 ==> bread and cake=t 95 <conf:(0.92)> lift:(1.25) lev:(0.04) [19] convr:(3.03)
106. biscuits=t,cheese=t,milk-cream=t 102 ==> bread and cake=t 84 <conf:(0.92)> lift:(1.25) lev:(0.04) [18] convr:(3)
108. juice-sat-cord-ms=t,milk-cream=t,margarine=t 100 ==> bread and cake=t 92 <conf:(0.92)> lift:(1.25) lev:(0.04) [18] convr:(2.94)
110. cheese=t,milk-cream=t,fruit=t 100 ==> bread and cake=t 92 <conf:(0.92)> lift:(1.25) lev:(0.04) [18] convr:(2.94)
112. breakfast food=t,frozen foods=t,milk-cream=t 87 ==> bread and cake=t 80 <conf:(0.92)> lift:(1.25) lev:(0.03) [16] convr:(2.88)
115. baking needs=t,cheese=t,milk-cream=t 99 ==> bread and cake=t 91 <conf:(0.92)> lift:(1.25) lev:(0.04) [18] convr:(2.91)
120. biscuits=t,sauces-gravy-pkls=t,frozen foods=t,milk-cream=t 86 ==> bread and cake=t 79 <conf:(0.92)> lift:(1.25) lev:(0.03) [15] convr:(2.84)
123. biscuits=t,frozen foods=t,milk-cream=t,fruit=t,vegetables=t 86 ==> bread and cake=t 79 <conf:(0.92)> lift:(1.25) lev:(0.03) [15] convr:(2.84)
124. baking needs=t,frozen foods=t,milk-cream=t 135 ==> bread and cake=t 124 <conf:(0.92)> lift:(1.25) lev:(0.05) [24] convr:(2.98)
126. cheese=t,milk-cream=t,vegetables=t 98 ==> bread and cake=t 90 <conf:(0.92)> lift:(1.25) lev:(0.04) [17] convr:(2.88)
128. biscuits=t,frozen foods=t,milk-cream=t 146 ==> bread and cake=t 134 <conf:(0.92)> lift:(1.25) lev:(0.05) [26] convr:(2.97)
129. juice-sat-cord-ms=t,cheese=t,milk-cream=t 85 ==> bread and cake=t 78 <conf:(0.92)> lift:(1.25) lev:(0.03) [15] convr:(2.81)
130. biscuits=t,soft drinks=t,milk-cream=t 85 ==> bread and cake=t 78 <conf:(0.92)> lift:(1.25) lev:(0.03) [15] convr:(2.81)
132. baking needs=t,frozen foods=t,milk-cream=t,margarine=t 85 ==> bread and cake=t 78 <conf:(0.92)> lift:(1.25) lev:(0.03) [15] convr:(2.81)
134. biscuits=t,pet foods=t,milk-cream=t 97 ==> bread and cake=t 89 <conf:(0.92)> lift:(1.25) lev:(0.04) [17] convr:(2.85)
135. juice-sat-cord-ms=t,frozen foods=t,milk-cream=t 120 ==> bread and cake=t 110 <conf:(0.92)> lift:(1.25) lev:(0.04) [21] convr:(2.89)
136. baking needs=t,soft drinks=t,milk-cream=t 84 ==> bread and cake=t 77 <conf:(0.92)> lift:(1.25) lev:(0.03) [15] convr:(2.78)
137. confectionary=t,milk-cream=t,fruit=t 84 ==> bread and cake=t 77 <conf:(0.92)> lift:(1.25) lev:(0.03) [15] convr:(2.78)
140. baking needs=t,frozen foods=t,milk-cream=t,vegetables=t 95 ==> bread and cake=t 87 <conf:(0.92)> lift:(1.25) lev:(0.03) [17] convr:(2.79)
141. biscuits=t,milk-cream=t,department1 37=t 83 ==> bread and cake=t 76 <conf:(0.92)> lift:(1.25) lev:(0.03) [14] convr:(2.74)
143. soft drinks=t,milk-cream=t,fruit=t 83 ==> bread and cake=t 76 <conf:(0.92)> lift:(1.25) lev:(0.03) [14] convr:(2.74)
145. baking needs=t,biscuits=t,sauces-gravy-pkls=t,milk-cream=t 83 ==> bread and cake=t 76 <conf:(0.92)> lift:(1.25) lev:(0.03) [14] convr:(2.74)
148. cheese=t,milk-cream=t,fruit=t,vegetables=t 83 ==> bread and cake=t 76 <conf:(0.92)> lift:(1.25) lev:(0.03) [14] convr:(2.74)

Figure4: Association Rules for 'Milk-Cream' item

It is producing the results based on user selected/interested item so the rules generated by this approach is reducing the number of association rules compared with existing approach.

The above approach is used for handling single dataset, so a real time web semantic engine is implementing using an Association Rule Interactive post-Processing using Schemas and Ontologies.

The framework implementation consists of following steps:

Step1: we propose to integrate user knowledge in association rule mining using two different types of formalism: ontologies and rule schemas.

Step2: we propose to use ontologies in order to improve the integration of user knowledge in the post processing task

Step3: we propose the Rule Schema formalism extending the specification language proposed by Liu et al. for user expectations.

Step4: Furthermore, an interactive framework is designed to assist the user throughout the analyzing task.

Interactive post mining Process Description:

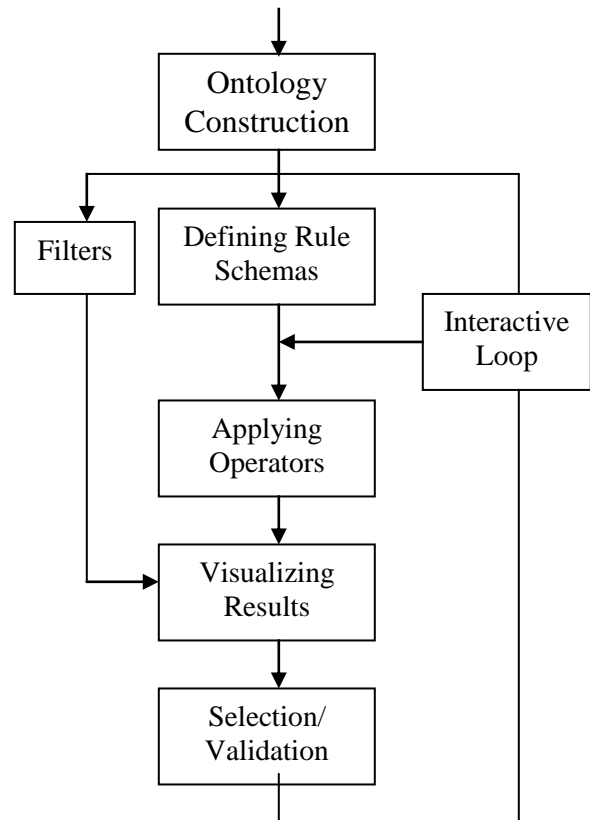


Figure5: Interactive post mining Process

Step1: Ontology will be developing by the user on database items. To describe the ontology the framework uses the Web Semantic representation language, OWL-DL.

Step2: The user expresses his/her local goals and expectations concerning the association rules that he/she wants to find in terms of rule schemas.

Step3: Applying the operators over the rule schemas created. The operators are pruning and filtering. The filters are minimum improvement constraint filter, and item relatedness filter.

Step4: Visualizing the filtered association rules by the user.

Step5: The user validates gained results.

Step6: Filters can be applied over rules whenever the user wants to reducing the number of rules.

Step7: The interactive loop permits to the user to revise the information that he/she proposed. So the user can return to step 2 in order to modify the rule schemas, or can return to step 3 in order to change the operators.

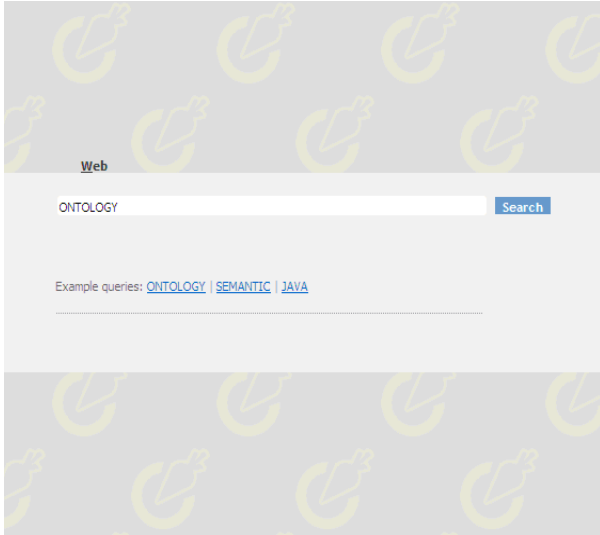


Figure 6: 'Ontology' Searching

Here the user wants to search 'ontology' then it will collect all url's related to that word and constructs ontology for those and then we are applying filters which are specified in the framework. Then it will produce filtered rules (filtered links) which is shown above.

For visualizing the results two ways of representations using here are 'Tree' and 'Visualization'. The rules generation is also done by placing user interest. That means the user will selects his/her interest item then rules will be generates according to the selected item.

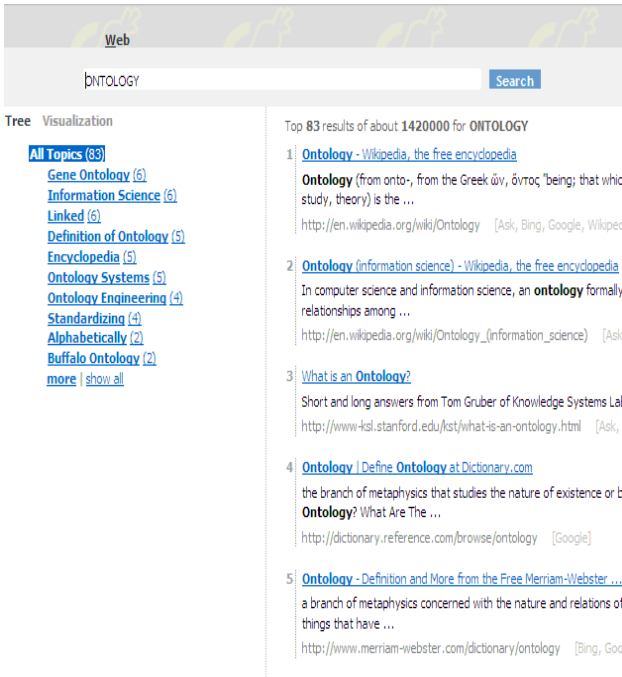


Figure 7: Representation of Results using Tree



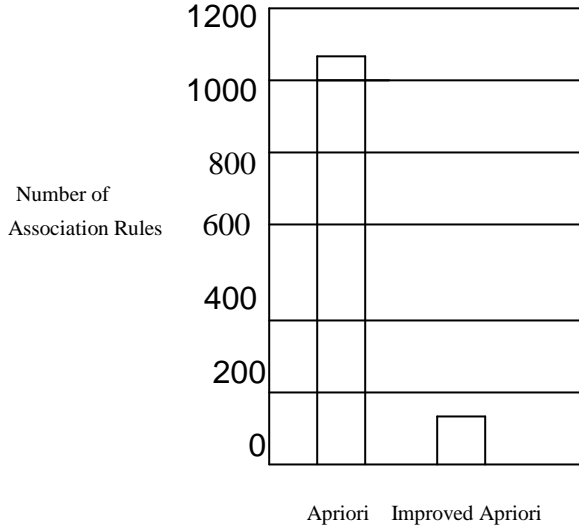
Figure8: Representation of Results using Visualization



Figure9: Rules for user selected item 'Encyclopedia'

Results

The results of the implementation are defined as follows:

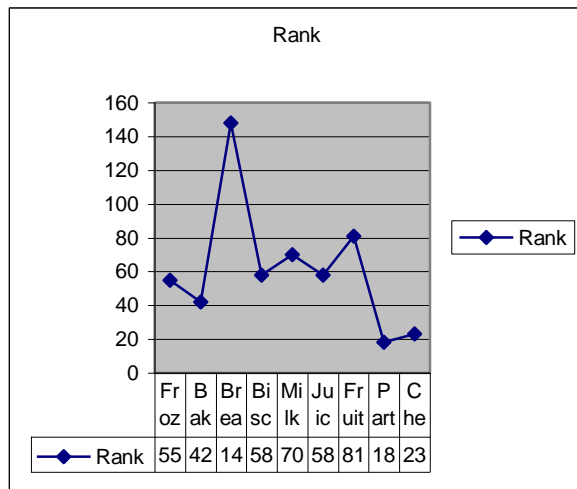


This graph is presenting the comparative results between Apriori and Improved Apriori.

-Apriori algorithm implemented by using support and confidence measures so it is generating 1048 association rules.

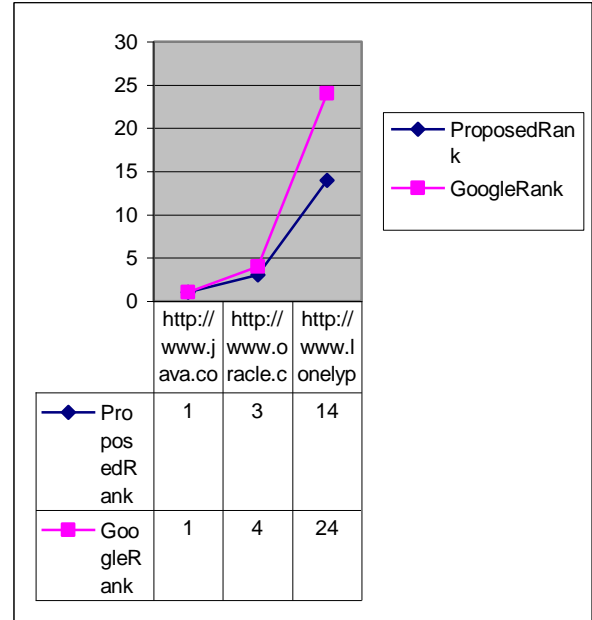
-Improved Apriori implemented by using lift, confidence, leverage, conviction measures so it is generating 148 association rules.

-When compared with the Apriori, Improved Apriori is reducing the association rules from 1048 to 148.



This graph is showing the results of number of association rules provided for each item based on user interest.

Real-time analysis: Proposed approach gives better rank results



This graph is presenting the comparative results between google search and our approach (using Association Rule Interactive Post mining using Ontologies and Rule schemas).

While we are searching for oracle then the related link <http://www.oracle.com> will be provided in 4th position by google search and in 3rd position by our approach.

We want search for lonely planet then the related link <http://www.lonelyplanet.com> will be provided in 24th position by google search and in 14th position by our approach.

Conclusion

The post-processing of association rules is improved by using a framework in order to generate association mining rules of user interest and also implemented a real time web semantic engine using an extended robust framework called Association Rule Interactive Post mining Using Schemas and Ontologies.

References

- [1] 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.
- [2] Hetal Thakkar, Barzan Mozafari, Carlo Zaniolo. Continuous Post-Mining of Association Rules in a Data Stream Management System. Chapter VII in *Post-Mining of Association Rules: Techniques for Effective Knowledge Extraction*, Yanchang Zhao; Chengqi Zhang; and Longbing Cao (eds.), ISBN: 978-1-60566-404-0.
- [3]. Srikant, R., Agrawal, R.: Mining generalized association rules. In: *VLDB 1995: Proceedings of the 21th International Conference on Very Large Data Bases*, pp. 407–419. Morgan Kaufmann Publishers Inc., San Francisco (1995)
- [4] Wenxiang Dou, Jinglu Hu, Gengfeng Wu, "Interesting Rules Mining with Deductive Method", ICROS-SICE International Joint Conference 2009.
- [5] Xuan-Hiep Huynh, Fabrice Guillet and Henri Briand, "Extracting representative measures for the post-processing of association rules", 2006.