A methodology for identifying interesting association rules by combining objective and subjective measures

Roberta Akemi Sinoara, Solange Oliveira Rezende

Computational Intelligence Laboratory
Computer Science Department
Institute of Mathematical and Computer Sciences
University of São Paulo - São Carlos Campus
P.O. box 668
São Carlos - SP - Brazil
{roberta, solange}@icmc.usp.br

Abstract

Evaluation measures, objective and subjective, are used to assist users in finding interesting association rules. Objective measures are more general, but they can be insufficient because they do not consider user's and domain features. However, getting user's knowledge and interest needed to calculate subjective measures can be a difficult task. Thus, this work presents a methodology to identify interesting association rules combining analysis with objective and subjective measures. This methodology aims to use the advantages of each kind of measure and to make user's participation easier. Objective measures are used to select some potentially interesting rules for the user's evaluation. Through this evaluation the user's subjectivity is obtained and used to calculate the subjective measures. Then, the subjective measures assist in identifying the interesting rules. In order to exemplify the methodology application, an experiment was carried out with a real database and the methodology showed to be feasible.

Keywords: Association Rules, Data Mining, Evaluation Measures.

1 Introduction

Interesting knowledge discovered in a data mining process can be used to assist users in some decision-making. However, identifying interesting knowledge is a difficulty in the data mining process. Many extraction algorithms generate a huge amount of patterns, which are in most cases not interesting to the user. This fact is frequent when the association technique is applied. The objective of this technique is to find out how a set of items in a database record is associated with other distinct set of items in the same record [1]. The extraction of association rules discovers all the associations present in a database record, which can lead to a huge number of rules and make the interesting knowledge identification difficult.

To address this problem in the analysis of patterns, many researches have focused on the development of methods to assist users in finding interesting knowledge [5, 11]. The application of knowledge evaluation measures is one of the most used techniques. Those measures are usually divided into objective (*data-driven*) and subjective (*user-driven*), considering what is necessary to calculate their values. Objective measures val-

ISSN: 1137-3601.

ues are based on the data and the rule structure, whereas user's knowledge and interest are considered, in some way, to calculate subjective measures.

Thus, objective measures have the advantage of being more general, because they do not depend on the domain and the user. Although they are efficient in many cases, the objective measures are not totally able to find rules that are really interesting to the user, due to the fact that the interest level of a rule can change depending on the user. A rule which is interesting in one evaluation might not be interesting in another moment, considering that the user's previous knowledge and goals have changed. These changes can be captured by subjective measures because their values are based on the user's knowledge and needs.

Nevertheless, in order to use subjective measures the user usually needs to provide his/her knowledge using a specific language or he/she needs to quantify imprecise concepts, e.g., indicating a numeric value to expressions like "low", "high", "big" and "small". This raises difficulties for the user to express his/her knowledge when providing it to a system which assists him/her in finding interesting rules.

In this context, this paper presents a methodology proposed to identify interesting association rules combining the application of objective and subjective measures. The aim of this methodology is to use the advantages of the different kinds of measures. Applying our methodology to rules from a real database, the user's knowledge and interest were obtained and rules of interest were identified.

This paper is organized as follow: Section 2 presents some works on objective and subjective measures. Section 3 describes the methodology to identify interesting association rules. An example of the application of our methodology is presented in Section 4 and finally Section 5 presents the conclusion of this work.

2 Related Work

Several researches have been done aiming to solve the rule interestingness problem. Some of these researches focus on objective aspects of interestingness evaluation and on the development of general techniques to select and order the rules according to their interestingness [5, 13]. Other researches also consider subjective aspects which are related to the user and the application domain [11, 7].

Considering only the data and the pattern structure, the interestingness can be estimated by statistical measures. Many measures, which can be calculated in a manner independent from the domain, have been developed and analyzed [3, 5, 2, 13]. In [8] a methodology is proposed to evaluate association rules based on these measures. This methodology aims to find a small set of potentially interesting rules to be delivered to the user. So a factor analysis is used to verify the similarity among the objective measures, and different graphs are used to verify the distribution of objective measure values in the rule set and to find the best cut point to filter these rules.

However, the interestingness of a rule is associated with the user's belief, knowledge and needs. Methods have been proposed to obtain and use the user's knowledge, experience and interest in search of interesting rules [4, 11]. The user's subjectivity can be obtained through: the specification of a domain taxonomy [10], the specification of rules using specific languages [7, 6], or the evaluation of some rules, directly done by the user [9].

An approach to identify interesting association rules proposed in [10] introduces the idea of *item-relatedness* to measure the interestingness among pairs of items of an association rule. In order to calculate this value, the user has to express his/her view of the domain through a fuzzy taxonomy, where items with similar functionalities are represented by close nodes.

In the technique presented in [7], the user provides his/her previous knowledge using a specific language. This language allows the specification of knowledge of different degrees of preciseness, called: general impressions, reasonably precise concepts, and precise knowledge. The user's specifications are matched against the generated rules. Then, four subjective measures are calculated: conforming, unexpected condition, unexpected consequent, and both-side unexpected. Using this technique it is possible to find the rules that conform to the user's knowledge or the rules that are, in some way, unexpected.

A different approach is proposed in [9], where the user directly evaluates some rules. This approach aims to reduce the original rule set by eliminating rules that are considered not interesting for the user. Therefore the rules are grouped in families, according to some features. Then, the user classifies some rules considering his/her interest and the rule validity. After a rule is evaluated, other rules are eliminated depending on how the rule was classified.

In the presented methodology the user's knowledge and interest are also obtained through the evaluation of some rules, aiming to make the user's participation easier. The rules to be evaluated are selected by applying the objective measures and then with these evaluated rules, the subjective measures are calculated.

3 Methodology to identify interesting association rules

The methodology to identify interesting association rules, presented in this paper, combines the application of evaluation measures, objective and subjective measures, using the advantages of each kind of measure. In this way, first the objective measures are used, aiming to filter the rule set, which usually has a large number of rules. Then, the subjective measures are used at the end of the process, to assist the user in analyzing the rules according to his/her knowledge and goals.

Figure 3 shows the methodology, which is divided into four phases: objective analysis, user's knowledge and interest extraction, evaluation processing, and subjective analysis.

3.1 Objective Analysis

The aim of the objective analysis phase is to filter the rule set generated by an extraction algorithm and select a rule subset to be evaluated by the user. Therefore the analysis stream proposed by Melanda [8] is used. This analysis stream uses rule set querying and objective evaluation measures.

The rule set querying is used if the user wishes to analyze rules that contain certain items, so a focus rule set is defined by the queries. If the user is not interested in specific items, the focus rule set is formed by the whole rule set. After defining the focus rule set the analysis using objective measures can start. For each measure an analysis can be done graphically of its value distribution in the focus rule set using scatter charts, bar charts, and Pareto's analysis graphs. After the analysis of the distribution of objective measures values a cut point to filter the focus rule set for each measure is set to select a subset of the rules. The union or the intersection of the subsets defined by each measure forms the subset of potentially interesting rules (PIR). After presenting this subset to the user, the next phase of the methodology starts with the user's knowledge and interest extraction.

3.2 User's Knowledge and Interest Extraction

The user's knowledge and interest extraction can be seen as an interview with the user, who answers some questions to evaluate the rules from the PIR subset. To optimize the evaluation, the rules are ordered according to *itemset* length. For each rule from the PIR subset, the user has to indicate one or more of the evaluation options. Each evaluation option classifies the knowledge represented by the rule. The options are: unexpected knowledge, useful knowledge, obvious knowledge, previous knowledge, and irrelevant knowledge considering the analysis goals.

A rule evaluated as an unexpected knowledge represents a novelty, something that the user has never thought about or something that contradicts his/her previous knowledge. The useful knowledge option indicates that the rule represents a knowledge which can be used to assist the user in some decision-making. A rule evaluated as a previous knowledge represents a user's knowledge which was formed by his/her past experiences. So, the previous knowledge can be different for each user. The obvious knowledge option indicates solid domain knowledge. An obvious knowledge is also a previous knowledge, though the obvious knowledge is not usually different for each user. A rule might be evaluated as an irrelevant knowledge if it represents a knowledge that is not important or is not necessary according to the data mining goals.

When a rule is classified as an irrelevant knowledge, the user must indicate the items that make it irrelevant. So the elimination of that rule and other irrelevant ones is possible. This elimination is part of the evaluation processing and is done to avoid the presentation of other irrelevant rules,

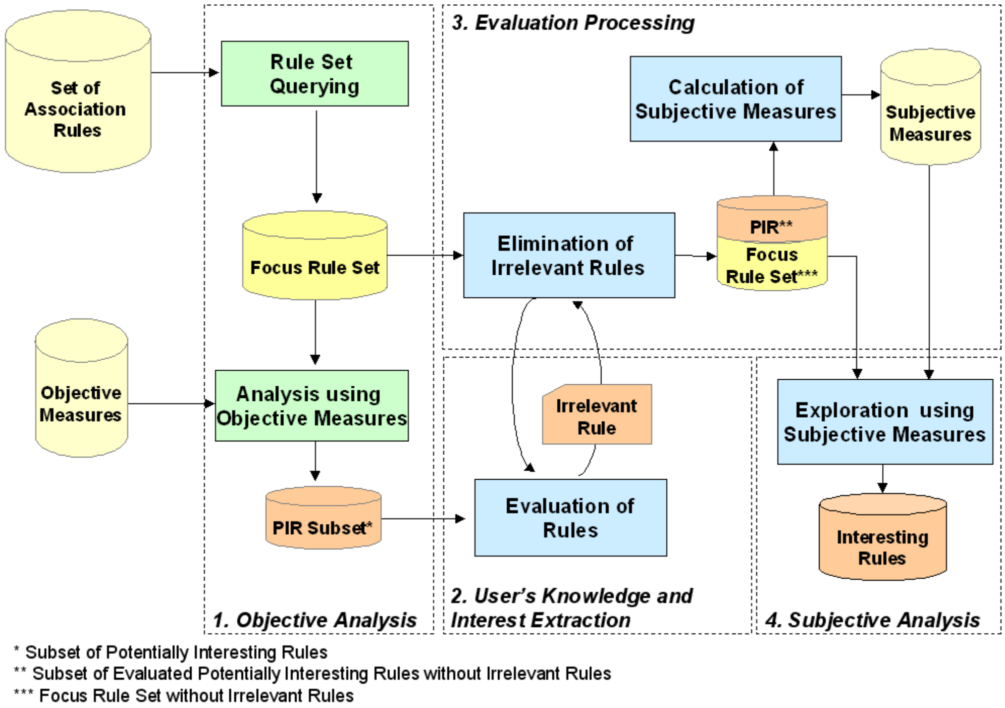


Figure 1: Methodology to interesting association rules identification

whose items have been evaluated before as irrelevant by the user. After the evaluation of every rule in the PIR subset, the subjective measures are calculated, also in the evaluation processing.

3.3 Evaluation Processing

After the user's evaluation for each rule from PIR subset, the evaluation processing starts in the focus rule set defined in the objective analysis. This processing consists of eliminating irrelevant rules and calculating subjective measures.

Every time a rule is classified as irrelevant knowledge by the user, all rules from the focus rule set which represent irrelevant knowledge according to the items indicated by the user during the evaluation are eliminated. First, the rules that contain the items in the same rule side (antecedent or consequent) are eliminated. After that, it is checked whether there are rules in the focus rule set which contain those items in different sides. When a rule is found, the user is asked about its irrelevance.

After the elimination, the subjective measures are calculated for the rules from the focus rule set that were not eliminated and do not belong to the PIR subset. These subjective measures defined by Liu et al. [7] are: conforming, unexpected condition, unexpected consequent, and both-side unexpected. In order to calculate these measures the rules evaluated by the user, except for the ones evaluated as irrelevant knowledge, are used as a reasonably precise concept provided by the user.

At the end of the processing, the user can analyze the focus rule set using the subjective measures, in the next phase of the methodology.

3.4 Subjective Analysis

The user can explore the resultant focus rule set using the subjective measures as a guide by accessing the rules according to each measure and considering each evaluated rule. This exploration should be carried out according to the goals of the user during the analysis. For example, if the user wishes to confirm his/her previous knowledge, he/she can use the conforming measure and list the rules that conforms to the rules that had been evaluated as obvious or previous knowledge; but if his/her aim is to find new knowledge, he/she can use the measures unexpected antecedent, unexpected consequent, and both-side unexpected, and list the rules that are contrary to an evaluated rule.

By browsing the focus rule set, the user identifies his/her rules of interest. Thus, at the end of the analysis, the user will have a set of the rules which were considered interesting.

3.5 Considerations about the Methodology

Our methodology combines the application of objective and subjective measures. The combined use of these measures can be useful for identifying interesting association rules, considering that the objective measures can filter the rule set and decrease the number of rules to be presented and evaluated by the user. Through the evaluation of these rules, it is possible to obtain the user's knowledge and interest needed to calculate the subjective measures. The subjective measures can be used to rank the rules according to the user's interest facilitating the identification of interesting rules.

As the methodology uses the analysis stream proposed by Melanda [8] combined with the subjective measures defined by Liu et al. [7], it reaches two aspects that make an association rule interesting: the presence of interesting items and the presence of interesting association among the items. The rules that contain items which are of interest to the users can be selected through the rule set querying during the objective analysis. The rules that contain interesting associations can be identified using the subjective measures.

Another aspect of our methodology that must be considered is the way the user's subjectivity is obtained. The user's knowledge and interest are obtained through the evaluation of some rules. Thus, some frequent limitations of subjective analysis are suppressed. As the user does not have to provide his/her knowledge explicitly to calculate the subjective measures, he/she does not need to use a specific language and to quantify his/her imprecise concepts.

In the user's knowledge and interest extraction it is also possible to identify the knowledge that is irrelevant to the user. This identification is very important, because when a rule is evaluated as irrelevant knowledge other rules are also eliminated from the focus rule set. The elimination of irrelevant rules assist in the identification of interesting rules during the subjective analysis regarding that the amount of rules to be analyzed decreases. However, if all the rules from PIR subset were classified as irrelevant, it will not be possible to calculate the subjective measures and another objective analysis must be done to define a new PIR subset for user's evaluation.

4 Application of the Methodology: An Example

This section describes an experiment carried out with a real database, in order to exemplify the application of our methodology. The entire data mining process was done. A domain specialist participated in the process, especially in the post-processing of the rules, which includes the user's knowledge and interest extraction (the second phase of our methodology).

The goal of this data mining process was to find associations among the features of certain urban sectors which influence the life quality of the residents. The database used in this experiment is about urban life quality (ULQ) and is composed of 120 examples, each one representing an urban sector of the city of São Carlos, SP, Brazil. All of its 19 attributes are continuous and do not have missing values. In order to generate the association rules, the database attributes were discretized under the user's supervision.

The association rules were generated by the *Apriori* algorithm[1], with minimum support and minimum confidence equal to 20% and 50%, respectively. The maximum number of items per rule was set at 3, aiming to improve the rules's comprehension. 4122 rules were obtained.

In the post-processing of these rules, our methodology was applied, using the ARInE [8] and RulEE-SEAR [12] tools. The description of the activities are divided into: ULQ rules' objective analysis, related to the first phase of the methodology; and ULQ rules' subjective exploration, related to the user's knowledge and interest extraction, evaluation processing, and subjective analysis.

[1] http://fuzzy.cs.uni-magdeburg.de/∼borgelt/apriori.html

4.1 ULQ Rules' Objective Analysis

In the objective analysis of the ULQ rules, the focus rule set was defined, the similarity among the objective measures was analyzed, and the PIR subset was formed.

The focus rule set was formed by all the 4122 extracted rules considering that the user does not have interest in a specific item.

In order to verify the similarity among the 31 objective measures used by the ARInE tool, the factor analysis was applied and with the result of this analysis, the measures were then divided into four groups, according to the behavior of each measure in the rule set.

One measure of each group was selected to filter the focus rule set so that the PIR subset could be formed. The selected measures were: support, confidence, lift, and relative specificity. These measures were selected because they belong to only one group and because their meaning has an easy interpretation.

The measures support and confidence had already been used to generate the rules. Therefore, they are not suitable to filter the rules. Every rule from focus rule set has acceptable values for these measures, according to the goals of the data mining process. To verify the value distribution of the measures lift and relative specificity in the focus rule set, scatter charts were generated in the ARInE tool. Figure 4.1 presents these charts. In each chart, for each rule from the focus rule set, identified by its number in the rule set (x axis) there is a number (y axis) referring to the measure value.

The measure lift indicates the dependency level between the antecedent and the consequent of the rule. For rules with lift value greater than 1, the antecedent positively influences the consequent. These rules can be considered interesting. The chart of Figure 4.1(a) presents the distribution of the lift values in the focus rule set. It can be observed that few rules have high values for the measure lift. So the rules from the focus rule set with lift value greater than 2 were selected. Table 4.1 shows the three rules that were found and their values for the measures support (Sup), confidence (Conf), lift, and relative specificity (RSpec).

The measure relative specificity indicates the rule specificity gain when the consequent is added. So the interesting rules might have high values of this measure. Figure 4.1(b) shows the distribution of the relative specificity values in the focus rule set. As it was observed that few rules have higher values (values close to 1), the rules having relative specificity value greater than 0.9 were selected. The 17 selected rules and their values for the four measures are presented in Table 4.1.

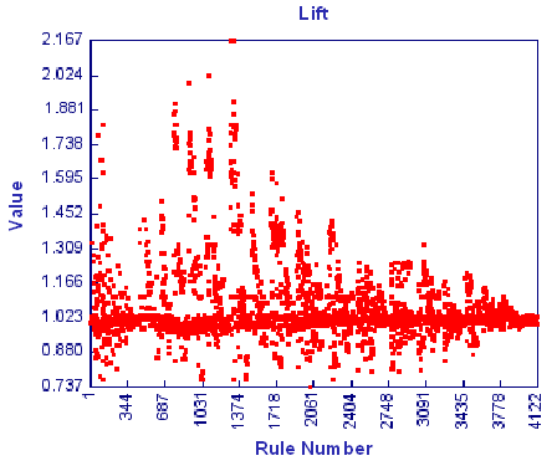
Analyzing Tables 4.1 and 4.1, it was observed that the 20 selected rules have acceptable values for the four considered measures . Every rule has support > 0.2, confidence > 0.5, lift > 1 and relative specificity > 0. Thus, all of them were considered potentially interesting and were used to form the PIR subset.

4.2 ULQ Rules' Subjective Exploration

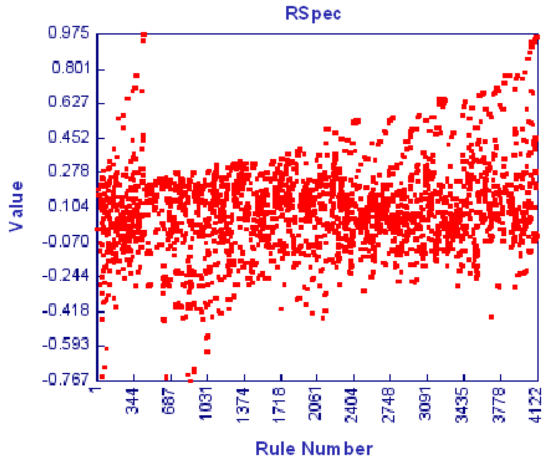
Having defined the focus rule set and the PIR subset, a subjective exploration process was done using the RulEE-SEAR tool. The subjective exploration consisted of the last three phases of the methodology: user's knowledge and interest extraction, evaluation processing, and subjective analysis.

In the user's knowledge and interest extraction, the user evaluated seven rules from the PIR subset. One rule was evaluated as unexpected knowledge, two as useful knowledge, one as obvious knowledge, and three as previous knowledge. The three rules classified as unexpected or useful knowledge were also considered interesting by the user. And the four classified as obvious or previous knowledge were also classified as irrelevant knowledge. Most of the rules from PIR subset were not evaluated because they contained items considered as irrelevant during the evaluation. So these rules were eliminated during the evaluation without presenting them to the user.

In the evaluation processing the irrelevant rules were eliminated during the evaluation, and the subjective measures were calculated. For each rule evaluated as irrelevant knowledge, the rules from the focus rule set that contained the irrelevant items were eliminated. From the four rules classified as irrelevant, 272 other rules were eliminated, including 13 rules from PIR subset that had never been presented to the user during the evaluation. After that, the four subjective measures were calculated for each rule from the focus rule set, that was not eliminated and does not be-



(a) Lift



(b) Relative specificity

Figure 2: Scatter charts of the values of the measures lift and relative specificity

Rule	Sup	Conf	Lift	RSpec
[R01] kilometers=(1.596667-3.193333], rotas=(-inf-35] ⇒ acessibi=(0.713-0.807]	0.2417	0.6744	2.0228	0.1833
[R02] cmd_dom=(5.25-7.5], acessibi=(0.807-inf) ⇒ kilometers=(-inf-1.596667]	0.2167	0.7223	2.1666	0.1750
[R03] acessibi=(0.807-inf), porpopal=(80.503333-inf) ⇒ kilometers=(-inf-1.596667]	0.3250	0.7222	2.1669	0.2625

Table 1: Rules selected with the condition lift > 2

long to the PIR subset. This calculation is based on the three rules that were evaluated and were not eliminated.

Using the subjective measures, the user analyzed the focus rule set. In this subjective analysis, other five interesting rules, that did not belong to the PIR subset, were identified. Table 4.2 presents these rules.

4.3 Considerations about the Example

This experiment shows that the application of the methodology is feasible. The use of the objective measures reduced the number of rules to be evaluated by the user. And the analysis using the subjective measures allows the identification of interesting rules, which were not found in the objective analysis.

During the evaluation, the user noticed that some database attributes were correlated. The rules which contained these attributes were classified as irrelevant. The identification of four irrelevant rules led to the elimination of 276 rules from the focus rule set. In order to verify the validity of the elimination, the rules from PIR subset that had been automatically eliminated were presented to the user. Analyzing the 20 rules from PIR subset, the user concluded that those rules represented only six different kinds of knowledge. All of them were evaluated during the user's knowledge and interest extraction. So, the elimination of irrelevant rules did not miss any kind of knowledge.

It can be noticed that the three rules considered interesting in the evaluation are the same three rules that were selected using the measure lift in the objective analysis. And all of the rules, that were selected using the measure relative specificity, were classified as irrelevant. Moreover, the values of relative specificity for the interesting rules, which were identified in the subjective analysis phase (Table 4.2), are relatively low whereas the values of lift are closer to the cut point set in the objective analysis. Therefore, in this experiment the measure lift represented better the user's interest as it did not point out rules with correlated attributes.

Rule	Sup	Conf	Lift	RSpec
[R04] agua_p=(91.719078-inf) ⇒ lixo_p=(87.578616-inf)	0.9417	1.0000	1.0169	0.9416
[R05] porpdimp=(-inf-1.30002] ⇒ dpi_p=(-inf-1.388889]	0.9750	1,0000	1.0256	0.9750
[R06] dpi_p=(-inf-1.388889] ⇒ porpdimp=(-inf-1.30002]	0.9750	1.0000	1.0256	0.9750
[R07] agua_p=(91.719078-inf), porpdimp=(-inf-1.30002] ⇒ porpdp=(97.018349-inf)	0.9167	1.0000	1.0344	0.9167
[R08] agua_p=(91.719078-inf), dpi_p=(-inf-1.388889] ⇒ porpdp=(97.018349-inf)	0.9167	1.0000	1.0344	0.9167
[R09] agua_p=(91.719078-inf), porpdp=(97.018349-inf) ⇒ lixo_p=(87.578616-inf)	0.9333	1.0000	1.0170	0.9333
[R10] agua_p=(91.719078-inf), porpdimp=(-inf-1.30002] ⇒ dpi_p=(-inf-1.388889]	0.9167	1.0000	1.0256	0.9167
[R11] agua_p=(91.719078-inf), dpi_p=(-inf-1.388889] ⇒ porpdimp=(-inf-1.30002]	0.9167	1.0000	1.0256	0.9167
[R12] agua_p=(91.719078-inf), porpdimp=(-inf-1.30002] ⇒ lixo_p=(87.578616-inf)	0.9167	1.0000	1.0169	0.9166
[R13] agua_p=(91.719078-inf), dpi_p=(-inf-1.388889] ⇒ lixo_p=(87.578616-inf)	0.9167	1.0000	1.0169	0.9166
[R14] agua_p=(91.719078-inf), sinstsan=(-inf-6.333333] ⇒ lixo_p=(87.578616-inf)	0.9333	1.0000	1.0170	0.9333
[R15] porpdp=(97.018349-inf), porpdimp=(-inf-1.30002] ⇒ dpi_p=(-inf-1.388889]	0.9500	1.0000	1.0256	0.9500
[R16] porpdp=(97.018349-inf), dpi_p=(-inf-1.388889] ⇒ porpdimp=(-inf-1.30002]	0.9500	1.0000	1.0256	0.9500
[R17] porpdimp=(-inf-1.30002], lixo_p=(87.578616-inf) ⇒ dpi_p=(-inf-1.388889]	0.9583	1.0000	1.0256	0.9583
[R18] dpi_p=(-inf-1.388889], lixo_p=(87.578616-inf) ⇒ porpdimp=(-inf-1.30002]	0.9583	1.0000	1.0256	0.9583
[R19] porpdimp=(-inf-1.30002], sinstsan=(-inf-6.333333] ⇒ dpi_p=(-inf-1.388889]	0.9583	1.0000	1.0256	0.9583
[R20] dpi_p=(-inf-1.388889], sinstsan=(-inf-6.333333] ⇒ porpdimp=(-inf-1.30002]	0.9583	1.0000	1.0256	0.9583

Table 2: Rules selected with the condition relative specificity $>$ 0.9

Rule	Sup	Conf	Lift	RSpec
cmd_dom=(5.25-7.5] ⇒ kilometers=(-inf-1.596667]	0.2167	0.5417	1.6249	0.1250
kilometers=(1.596667-3.193333], perocppr=(62.253333-79.886667] ⇒ acessibi=(0.713-0.807]	0.2500	0.6000	1.8000	0.1667
kilometers=(1.596667-3.193333], rendchef=(-inf-261494] ⇒ acessibi=(0.713-0.807]	0.2583	0.5345	1.6035	0.1458
kilometers=(1.596667-3.193333], rotas=(-inf-35] ⇒ cmd_dom=(-inf-5.25]	0.2667	0.7441	1.4402	0.1686
kilometers=(1.596667-3.193333], rotas=(-inf-35] ⇒ denscons=(-inf-266705.546667]	0.2667	0.7441	1.1448	0.0964

Table 3: Interesting rules identified in the subjective analysis phase

5 Conclusion

This paper presents a methodology proposed to identify interesting association rules through the combination of objective and subjective measures. The main aim of the combination is to make the user's participation easier. As the objective measures are more general, they are used as filters to select a subset of potentially interesting rules. These rules are presented to the user, who evaluates them according to his/her knowledge and interest. Through this evaluation the user's subjectivity is obtained. This process may not be difficult for the user, considering that he/she does not need to provide explicitly his/her knowledge. He/she just needs to evaluate the rules by choosing one or more options that indicate the kind of knowledge the rule represents. Based on the evaluated rules the subjective measures conforming, unexpected condition, unexpected consequent, and both-side unexpected can be calculated. So, these measures are used in the last phase of our methodology to guide the user in the analysis of the rules.

In order to exemplify the application of our methodology, an experiment was carried out with the participation of a specialist in urban life quality. This experiment shows that with the application of our methodology it is possible to identify interesting association rules by using the advantages of the objective and subjective measures.

In our future work, other experiments using our methodology with different real databases and specialists in other domains will be carried out.

Acknowledgements

We would like to thank Edson Melanda and Magaly Fujimoto for contributing to this work. We also thank Anandsing Dwarkasing for revising this paper. This work was partially supported by CNPq, Institute Factory of the Millennium (IFM).

References

[1] Rakesh Agrawal and Ramakrishnan Srikant. Fast algorithms for mining association rules. In Jorge B. Bocca, Matthias Jarke, and Carlo Zaniolo, editors, *Proceedings of the 20th International Conference on Very Large Data Bases, VLDB'94*, pages 487–499, 1994.

[2] D. R. Carvalho, A. A. Freitas, and N. F. F. Ebecken. A critical review of rule surprisingness measures. In *Proceedings of the Fourth International Conference on Data Mining - Data Mining IV*, pages 545–556. WIT Press, 2003.

[3] Alex Alves Freitas. On objective measures of rule surprisingness. In J. M. Zytkow and M. Quafafou, editors, *Proceedings of the 2nd European Conference on the Principles of Data Mining and Knowledge Discovery*, volume 1510 of *Lecture Notes in Artificial Intelligence*, pages 1–9, 1998.

[4] Mika Klemettinen, Heikki Mannila, Pirjo Ronkainen, Hannu Toivonen, and A. Inkeri Verkamo. Finding interesting rules from large sets of discovered association rules. In Nabil R. Adam, Bharat K. Bhargava, and Yelena Yesha, editors, *Proceedings of the Third International Conference on Information and Knowledge Management CIKM'94*, pages 401–407. ACM Press, 1994.

[5] N. Lavrač, P. Flach, and R. Zupan. Rule evaluation measures: A unifying view. In S. Dzeroski and P. Flach, editors, *Proceedings of the 9th International Workshop on Inductive Logic Programming ILP'99*, volume 1634 of *Lecture Notes in Artificial Intelligence*, pages 174–185. Springer-Verlag, 1999.

[6] Bing Liu and Wynne Hsu. Post-analysis of learned rules. In *Proceedings of the 13th National Conference on Artificial Intelligence AAAI'96*, pages 828–834, 1996.

[7] Bing Liu, Wynne Hsu, Shu Chen, and Yiming Ma. Analyzing the subjective interestingness of association rules. *IEEE Intelligent Systems & their Applications*, 15(5):47–55, 2000.

[8] Edson Augusto Melanda. *Post-processing of Association Rules*. PhD thesis, Instituto de Ciências Matemáticas e de Computação – USP – São Carlos, 2004. (In Portuguese).

[9] S. Sahar. Interestingness via what is not interesting. In *Proceedings of the Fifth International Conference on Knowledge Discovery and Data Mining*, pages 332–336. ACM Press, 1999.

[10] B. Shekar and Rajesh Natarajan. A framework for evaluating knowledge-based interestingness of association rules. *Fuzzy Optimization and Decision Making*, 3(2):157–185, 2004.

[11] A. Silberschatz and Alexander Tuzhilin. What makes patterns interesting in knowledge discovery systems. *IEEE Transactions on Knowledge and Data Engineering*, 8(6):970–974, 1996.

[12] R. A. Sinoara. Identification of interesting association rules through analysis using objective and subjective measures. Master's thesis, Instituto de Ciências Matemáticas e de Computação – USP – São Carlos, 2006. (In Portuguese).

[13] P. N. Tan, V. Kumar, and J. Srivastava. Selecting the right objective measure for association analysis. *Information Systems*, 29(4):293–313, 2004.