Subjective Evaluation of Labeling Methods for Association Rule Clustering

Renan de Padua[1], Fabiano Fernandes dos Santos[1], Merley da Silva Conrado[1],
Veronica Oliveira de Carvalho[2], and Solange Oliveira Rezende[1]

[1] Instituto de Ciências Matemáticas e de Computação,
USP - Universidade de São Paulo, São Carlos, Brazil
{padua,fabianof,merleyc,solange}@icmc.usp.br
[2] Instituto de Geociências e Ciências Exatas,
UNESP - Univ Estadual Paulista, Rio Claro, Brazil
veronica@rc.unesp.br

Abstract. Among the post-processing association rule approaches, clustering is an interesting one. When an association rule set is clustered, the user is provided with an improved presentation of the mined patters. The domain to be explored is structured aiming to join association rules with similar knowledge. To take advantage of this organization, it is essential that good labels be assigned to the groups, in order to guide the user during the association rule exploration process. Few works have explored and proposed labeling methods for this context. Moreover, these methods have not been explored through subjective evaluations in order to measure their quality; usually, only objective evaluations are used. This paper subjectively evaluates five labeling methods used on association rule clustering. The evaluation aims to find out the methods that presents the best results based on the analysis of the domain experts. The experimental results demonstrate that there is a disagreement between objective and subjective evaluations as reported in other works from literature.

1 Introduction

Association rule mining (ARM), introduced in [1], is an important task of data mining. ARM aims to "find all co-occurrence relationships, called associations, among data items" [11].

Association rules have been successfully applied for decision support (such as the cross-marketing, attached mailing applications, catalog design, add-on sales, store layout, and customer segmentation based on buying patterns) [3], for applications of telecommunications alarm diagnosis and prediction [2], for inter-disciplinary domains beyond data mining (such as indexing and similarity search of complex structured data, spatio-temporal and multimedia data mining, stream data mining, web mining, software bug mining, and page-fetch prediction) [8], and for disease prediction [17].

When generating association rules, it is necessary to deal with a huge amount of rules since the number of rules grows exponentially with the number of items in

F. Castro, A. Gelbukh, and M. González (Eds.): MICAI 2013, Part II, LNAI 8266, pp. 289–300, 2013.

the data set [9]. Many algorithms have been developed to overcome the problem of dealing with these generated rules. These algorithms follow one of these post-processing approaches: *Q*uerying (*Q*), *E*valuation *M*easures (*EM*), *P*runing (*P*), *S*ummarizing (*S*), or *G*rouping (*G*) [5,22,14,10]. The algorithms that belong to the approaches of *Q*, *P*, and *S* aid the exploration process by *r*educing the *e*xploration *s*pace (*RES*); the ones that belong to *EM* approach explore the process by *d*irecting the *u*ser to what is *p*otentially *i*nteresting (*DUPI*); and, finally, the algorithms of *G* approach explore the process by *s*tructuring the *d*omain (*SD*).

Grouping is a relevant approach related to *SD*, since it organizes the rules in groups that contain, somehow, similar knowledge. These groups improve the presentation of the mined rules, providing the user a view of the domain to be explored [18,19]. A methodology was found in the literature for post-processing association rules that utilizes the grouping approach. This methodology, called PAR-COM [5], combines clustering and objective measures to direct the user to what is potentially interesting and, consequently, reduces the association rule exploration space. Thus, the user only needs to explore a small subset of the groups that contain the potentially interesting knowledge. However, it is essential that groups be represented by labels that may provide the user a view of the subjects contained in the exploration space, helping to guide its search.

Although some methods have been proposed to label document clusters in Text Mining (TM) and Information Retrieval (IR) [13,12,16], there are few researches in the literature that deal with selecting labels for association rule clustering. Padua et al. [15] and Carvalho et al. [4] assess some labeling methods using objective evaluations. Chang et al. [7] discuss about a disagreement between objective and subjective evaluation results in a topic extraction context. The latter found that some results of objective measures are not always a good predictor of human judgments regarding the terms selected as labels for the topic extraction task. The same problem is found here since the label selection task is similar to topic extraction and association rule clustering approaches.

Considering that, we use a subjective methodology to evaluate label sets obtained by labeling methods for association rule clustering. For that, this paper presents an adapted version of the subjective evaluation methodology proposed in [7] (details in Section 3). The evaluation was applied in five labeling methods for association rule clustering in order to identify which one obtains suitable label sets according to the

The proposal of an evaluation methodology adapted from [7] is introduced and adjusted for an environment that considers clusters of association rules obtained from structured data. Specifically, the proposed evaluation methodology is based on a task named *word intrusion*. The word intrusion task, proposed in [7], consists of identifying a spurious word inserted into a set of words[1] that represent the extracted topic. The *word intrusion* task was initially proposed to evaluate whether an extracted topic has human-identifiable semantic coherence.

[1] In this work, a set of words represents the labels of a group.

This paper is organized as follows. The labeling methods used in this work are presented in Section 2. The subjective evaluation methodology is described in Section 3. The configuration of the experiments are introduced in Section 4 and the results are discussed in Section 5, arguing about the differences obtained between the subjective and the objective analysis. Finally, the related works are presented in Section 6 and conclusions in Section 7.

2 Labeling Methods

Although the organization of association rules through clustering provides some important clues for user, the exploration task remains a challenge since there is no explicit information about the subject of each cluster. Even in a small data set, it is not easy to define a main idea that links the association rules in each cluster. However, the cluster may be represented by a set of meaningful labels. Therefore, it is important to find a good set of labels for each cluster. In this paper, five labeling methods (LM) for association rule clustering, briefly described below, were selected and implemented to be subjectively evaluated. These labeling methods were the ones indicated by [4,6,15] as good solutions to this kind of problem (for details, please, see the references).

LM-M (*L*abeling *M*ethod *M*edoid) selects as labels of each cluster the items in the rule that is more similar to all the other rules in the same cluster (the cluster's medoid). The method computes the accumulated similarity of each rule considering its similarity with respect to all the other rules; then, the one with the highest value is selected. Therefore, the labels of each cluster are built by the items that appear in the cluster's medoid.

LM-T (*L*abeling *M*ethod *T*ransaction) builds the clusters labels by selecting the items in the rule that covers the largest number of transactions. A rule covers a transaction t if all the rule items are contained in t. Therefore, this method counts the number of transactions each rule covers and selects the rule that covers the largest number of transactions. In the end, the rule items are considered as the clusters labels.

In **LM-S** (*L*abeling *M*ethod *S*ahar due to its reference to [19]), a simplified version of the process described in [19], the labels of each cluster are built as follows: (i) considering a set $I = \{i_1, ..., i_m\}$ containing all the distinct cluster items, a set $R = \{r_1, ..., r_n\}$ containing all the possible relationships $a \Rightarrow c$, where $a, c \in I$ – each one of these relationships represents a rule pattern; (ii) the number of rules that each pattern $r_i \in R$ covers is computed (N_c); a pattern $a \Rightarrow c$ covers a rule $A \Rightarrow C$ if $a \in A$ and $c \in C$; (iii) the pattern with the highest cover is selected; in the event of a tie all tied pattern are selected; (iv) all the selected patterns compose a set $P \subseteq R$; (v) in the end, all the distinct items in P compose the labels.

In **LM-PU** (*L*abeling *M*ethod *P*opescul and *U*ngar due to its reference to [16]), the labels of each cluster are built by the N items in the cluster that present the best tradeoff between frequency and predictiveness; formally we have: $f(i_n|C_n) * \frac{f(i_n|C_n)}{f(i_n)}$. The $f(i_n|C_n)$ measure computes the frequency f of each item

i_n in its cluster C_n. The $\frac{f(i_n|C_n)}{f(i_n)}$ measure computes the frequency f of each item i_n in its cluster C_n divided by the item frequency in all the clusters. The i_n items are all the distinct items that are present in the rules of the cluster. Each time an item i_n occurs in a rule its frequency is incremented by one. Therefore, the labels are built by the N items that are more frequent in their own cluster.

In **GLM** (*G*enetic *L*abeling *M*ethod [15]), the labels of each cluster are chosen by optimizing two measures, Precision and Repetition Frequency [4]. The Precision measure (P) computes the number of rules the labels cover in their own cluster and divides by the number of rules in that cluster. A rule is covered by the clusters labels if one or more items in the labels are part of the rule. The Repetition Frequency (RF) measures how different the labels are among the clusters by counting the number of items that are repeated in different cluster labels and dividing this value by the number of distinct items in all the labels. Values near 0 show that there are few repetitions among the labels while values near 1 indicate that the labels are very similar. It was considered, during the optimization process, $RF = 1.0 - RF$ so that both P and RF were maximized. GLM is a genetic algorithm approach that aims to ensure a good tradeoff between P and RF. The fitness function of an individual is defined by $Fitness(I) = (P+RF) - \left(\frac{Max(P,RF)}{Min(P,RF)} * 10^{-5}\right)$, where 10^{-5} indicates the minimal possible value the measures may get, (P + RF) show how good are the measures according to the labels and $\left(\frac{Max(P,RF)}{Min(P,RF)} * 10^{-5}\right)$ the penalty proportional to the distance between P and RF. Initially, the method randomly selects the labels of each cluster among the rule items in each cluster. Thereafter, the population of labels undergoes crossover until it reaches a given number of generations.

3 Item Intrusion Task: The Subjective Evaluation Methodology

In order to subjectively evaluate the labels obtained by each labeling method described in Section 2, we implemented the *item intrusion task*, which was adapted from [7]. As proposed in [7], the users' task is to find the item, among a set of items, that is out of place or does not relate with the others, i.e., *the intruder*. The methodology works as follows:

Step A. The n best items and the m worst items in each cluster are selected according to each labeling method described in Section 2. n and m are numbers to be chosen. The n best items in each cluster represents the labels of the clusters. The m worst items, called here as intruders, are the m least items to be selected as labels and, also, the ones that appear in some of the other clusters. The last condition ensures an item will not be selected as a bad item due solely to its rarity. The process is illustrated in Figure 1. In this example, considering $n = 5$, the best ranked items, that compose the labels of the cluster k, are "Item 1", "Item 2", "Item 3", "Item 4", and "Item 5". On the other hand, considering $m = 5$, the worst ranked items that also occur in the other clusters are “Item j”, “Item j-2”, “Item j-4”, “Item j-6”, and “Item j-8”. The underlined and emphasized items are considered to be present only in the current cluster. The rank is based on the criteria used on a given LM.

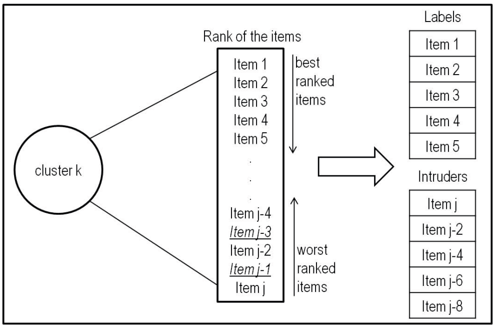
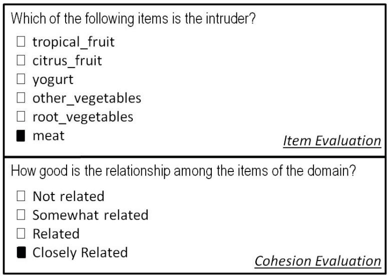


Fig. 1. The selection process of the n best items and the m worst items



Which of the following items is the intruder?

□ tropical_fruit
□ citrus_fruit
□ yogurt
□ other_vegetables
□ root_vegetables
■ meat

Item Evaluation

How good is the relationship among the items of the domain?

□ Not related
□ Somewhat related
□ Related
■ Closely Related

Cohesion Evaluation

Fig. 2. Illustrative example of the evaluation process

Step B. A intruder i is picked up at random from a pool built with the m items identified in [Step A].

Step C. The n items plus the intruder i are shuffled and presented to the user, as in Figure 2. In this example, the user is asked to identify the intruder, i.e., the item that should not be selected as label. It is expected that the user chooses the item i. The authors of [7] claim that when the set of labels minus the intruder i makes sense together, then the subject should easily identify the intruder.

Step D. Finally, as seen in Figure 2, the user is also asked about the cohesion of the set of labels minus the intruder i. The cohesion in this work is evaluated through four available options: “not related”, “somewhat related”, “related”, and “closely related”. The options are associated with values ranging from 1 to 4 – 1 regarding the “not related” option until 4 regarding the “closely related” option. The cohesion value may be used to determine if a labeling method is selecting the labels in a random way. When combined with the results of [Step C], the cohesion value improves the evaluation regarding the coherence of the set of labels.

It is important to mention that the methodology here presented is primarily an adaptation of the one presented in [7] for an environment that considers clusters of association rules obtained from structured data. The major difference is [Step D]. Therefore, the main contribution of this work is the subjective evaluation of labeling methods for association rule clustering. However, the presented methodology is also another contribution, since it discusses a standardized assessment process to the context of association rule clustering.

4 Experiments

Experiments were carried out to subjectively evaluate the five labeling methods applied in the context of association rule clustering (see Section 2). Thereby, this section is divided in three parts: one related to the criteria used to evaluate the obtained results, one that describes the data sets used and, finally, the one that discuss the experimental setup.

4.1 Evaluation Criteria

Based on the methodology described in Section 3, the evaluation of the results was done considering two different aspects: (i) percentage of correct answers (PCA) and (ii) cohesion (Co).

In the first aspect, the quality of the labels is measured by the rate of correct answers given by the subjects regarding the intruder selection method [Step C]. In high quality label sets the relationship of selected items presents a good summary of the cluster content. Also, the intruder item can be easily identified in the label sets with high quality. Thus, PCA metric may be expressed by $PCA = \frac{\#\ of\ correct\ answers}{\#\ of\ clusters}$. In this case, the metric checks how many times the user identifies the intruder i in each of the clusters, related to a given association rule clustering, and divides it by the total number of clusters in the clustering. A $PCA = \frac{5}{10} = 50\%$ indicates that the clustering has 10 clusters and in 5 of them the user identified, among the labels, the intruder item.

In the second aspect, the cohesion metric aims to measure how correlated the labels are without the intruder i in the subject opinion. However, in this case, a mean of the cohesions is obtained. Thus, this metric may be expressed by $Co = \frac{\sum_{i=1}^{\#\ of\ clusters} Co_i}{\#\ of\ clusters}$. In this case, the metric sums the cohesions assigned to each of the clusters by the user, related to a given association rule clustering, and divides it by the total number of clusters in the clustering. A $Co = \frac{4+3+2+4+2}{5} = 3.0$ indicates that the clustering has 5 clusters and, in average, a cohesion of 3.0.

4.2 Data Sets

Four *d*ata *s*ets (DS) were considered to run the experiments: Adult (48842;115), Income (6876;50), Groceries (9835;169), and Sup (1716;1939). The numbers in parenthesis indicate, respectively, the number of transactions and the number of distinct items in each data set. The first three are available through the package "arules"[2]. The last one was donated by a supermarket located in São Carlos city, Brazil. All the transactions in Adult and Income contain the same number of items (named here as *s*tandardized *d*ata *s*ets (SDS)), different from Groceries and Sup (named here as *n*on-*s*tandardized *d*ata *s*ets (NSDS)), whereupon each transaction contains a distinct number of items. Thus, the experiments considered different data types. The rules, in each data set, were mined using

[2] `http://cran.r-project.org/web/packages/arules/index.html`.

an *Apriori* implementation[3] with a minimum of 2 and a maximum of 5 items per rule. From the Adult set 6508 rules were extracted using a minimum support (min-sup) of 10% and a minimum confidence (min-conf) of 50%; Income 3714 rules with min-sup=17%, min-conf=50%; Groceries 2050 rules with min-sup=0.5%, min-conf=0.5%; Sup 7588 rules with min-sup=0.7%, min-conf=0.5%. The parameters were set experimentally.

4.3 Experimental Setup

Initially, in order to apply the five labeling methods described in Section 2 to perform the subjective analysis, the rule sets, related to each data set, were clustered. The clusterings were obtained using the Ward Link algorithm [21] and, as similarity measure, J-RT (see Section 6). This configuration, clustering algorithm + similarity measure, was selected because, as shown by [6], it obtains the best results when the clustering is done in the post-processing phase. Each one of the dendrograms were cut with a threshold of 0.2[4] (the value was obtained experimentally). Based on these cuts, 12 groups were obtained for the Adult data set, 5 for Income, 33 for Groceries, and 10 for Sup.

After clustering the rule sets, all the five labeling methods were applied to each clustered set. The methodology described in Section 3 was then executed in each considered configuration, i.e., labeling method + data set. In total, 20 experiments were done (20 = 5 LM × 4 DS). The values of $n = 5$ and $m = 5$ were considered to apply the subjective evaluation (see Section 3).

Once subjective evaluations are often expensive due to the large amount of data available and to the limited available time of the evaluators, a sampling of clusters was considered in each data set in all the LM. Thus, in this work, we randomly selected 20% of the clusters in each clustered rule set. Therefore, 2 groups were considered for Adult in each LM, 1 for Income, 6 for Groceries, and 2 for Sup. These selected groups, for each data set, in each LM, were the ones presented to the users as shown in Figure 2.

All the users assessed the same clusters, and also the items in the groups, in the same sequence, because as the users' interaction with the system improves, he/she may change the choices and, consequently, the results. The sequence of presentation was set randomly. However, each user evaluated only one data set of each type, i.e., one group of users evaluated the results related to Adult and Sup data sets and the other the ones related to Groceries and Income. This split was done aiming to lower the number of questions each subject should answer and to force each subject to evaluate both a standardized and a non-standardized data set.

Finally, it is important to mention that a warm-up step was considered during the experiments. In this phase, two more groups were selected to be initially presented to the users (so in Adult, for example, 4 groups were selected in the

[3] `http://www.borgelt.net/apriori.html` [Christian Borgelt's Web Page].

[4] Considering dendrograms with maximum height of 1. The root node is close or equal to 1.

total). These 2 additional groups were used as a training phase, aiming to introduce the validation process to the subjects. In this stage, the user learns how the environment works and how to interact with it. In this phase, right and wrong answers may not represent the user knowledge. These initial groups were not considered in the final results.

The Groceries and Sup data sets were evaluated by 5 users, each one having a good knowledge of the domain. The Adult and Income data sets were also evaluated by 5 users, but in this case the users did not have a good knowledge of the domain (related to customer's profile).

5 Results and Discussion

As mentioned before, to analyze the obtained results the metrics PCA and Co (Section 4) were used. The results are presented in the Tables 1 and 2 for Groceries and Sup data sets and in the Tables 3 and 4 for Adult and Income data sets. The results express the averages of PCA and Co obtained from the assessments of all the users in the analyzed clusters in each data type.

The highest values in each of the tables, regarding each one of the metrics, are marked with ▲ in each considered data set. For the Groceries data set, for example, the best value for PCA is the one related to GLM (51.43% (Table 1)); for Co also the one related to GLM (3.43 (Table 2)). It may be noted that, regarding the NSDS (Tables 1 and 2):

PCA **aspect:** according to the results obtained by [15], LM-PU is the more suitable method to be used according to objective evaluations; however, in this subjective evaluation:

- GLM is the more suitable LM to be used according to the user views;
- LM-M obtained good results;
- LM-PU and LM-S present the worst results.

Thus, there is a disagreement between the two evaluations, as noted by [7] for the topic extraction task.

Co **aspect:** following the same reasoning, it may also be observed the same disagreement regarding this other aspect:

- although GLM presents the highest value in only one data set (Groceries), in the other data set the values are more closely, indicating that GLM is also the more suitable LM to be used according to the user views considering this criterion;
- the values for Groceries have a low variation among the LM, except for GLM that presents a high value – mean of 3.26;
- the values for Sup also have a low variation among the LM, except for LM-PU that presents the worse value – mean of 3.20.

The analysis of the tables also shows interesting behaviors. In cases of high Co and low PCA values, such as in LM-S on Sup data set (3.50 x 0%), it is understandable that the method failed to distinguish the groups, i.e., the labels of the clusters were not specific enough to describe their own groups and, consequently, to distinguish each one from the others. On the other hand, in cases of low Co and high PCA values, such as in LM-PU on Sup data set (2.70 x 40%), it is understandable that the method obtained less correlated labels, but it was successful on selecting labels that best represent each group, making it easier for the subjects to find the intruder. Finally, Figure 2 shows an example of an answer given by an user regarding a group related to the Groceries data set. In this case, considering this unique group, $PCA = 1$ and $Co = 4$.

On the other hand, regarding the SDS, the users did not have a good knowledge of the domain (related to customer profile) and, therefore, the results were much worse. It may be noted that (Tables 3 and 4):

PCA **aspect:** according to the results obtained by [15], GLM is the more suitable method to be used according to objective evaluations; however, in this subjective evaluation:

- the number of low PCA values is high, being some of them 0%. Thus, although LM-T presents the highest value in only one data set (Income), it is the only method that has an uniform behavior in both data sets. Therefore, it may be assumed that LM-T is the more suitable LM to be used according to the user's views.

Thus, there is a disagreement between the two evaluations, as noted by [7] for the topic extraction task.

Co **aspect:** following the same reasoning, it may also be observed the same disagreement regarding this other aspect:

- Both data sets had high Co values according to the subject's opinion, contrasting to the low PCA values.

Table 1. PCA results related to NSDS

Data set	LM-M	LM-T	LM-PU	LM-S	GLM
Groceries	46.67%	20%	11.43%	11.43%	51.43%▲
Sup	50%	50%	40%	0%	70%▲

Table 2. Co results related to NSDS

Data set	LM-M	LM-T	LM-PU	LM-S	GLM
Groceries	3.23	3.23	3.29	3.14	3.43▲
Sup	3.30	3.30	2.70	3.50▲	3.20

Table 3. *PCA* results related to SDS

Data set	LM-M	LM-T	LM-PU	LM-S	GLM
Adult	30%	40%	50%▲	30%	10%
Income	0%	40%▲	0%	0%	20%

Table 4. *Co* results related to SDS

Data set	LM-M	LM-T	LM-PU	LM-S	GLM
Adult	3.0	3.1▲	3.0	2.9	3.0
Income	3.6	3.8▲	3.6	3.4	3.8▲

6 Related Works

Since this paper aims to evaluate labeling methods for association rule clustering, in this section, we briefly review some papers related to association rule clustering, mentioning their labeling methods and the methodologies used to evaluate the methods.

The aim of the clustering approach in the post-processing phase is to improve and organize the presentation of the obtained association rules. The result of this process is a structured view of domain to be explored. In [20] the authors propose a similarity measure based on transactions and apply a density cluster algorithm to group the association rules. They also present an evaluation in a small set of the rules to motivate the research with association rule clustering approach. The approach proposed in [10] explores the lexical features of the rules, rather than their statistical properties, for structuring the rule space. They explored hierarchical cluster algorithms in the evaluations using Jaccard as the similarity measure. In [10], the Jaccard value between two rules r and s, expressed by J-RI(r,s)=$\frac{|\{items\ in\ r\}\cap\{items\ in\ s\}|}{|\{items\ in\ r\}\cup\{items\ in\ s\}|}$, is calculated considering the items the rules share. The authors of [18] compare two kinds of clustering methods, partitional and hierarchical, also using Jaccard as the similarity measure. However, in this case, the Jaccard value between two rules r and s is expressed by J-RT(r,s)=$\frac{|\{t\ matched\ by\ r\}\cap\{t\ matched\ by\ s\}|}{|\{t\ matched\ by\ r\}\cup\{t\ matched\ by\ s\}|}$, where t is the common transactions the rules match. In [10] and [18], the labels of each group are compound of the items that are presented in the rule that is more similar to all the other rules in the group (the medoid of the group). Toivonen et al. [20] do not mention how the labels are found, but provide some clues that the labels represent the most frequent and distinct items in the group.

In all the cases, the authors are mainly concerned with the domain organization and do not present a deeper evaluation of the labels. In [15], an interesting labeling method based on genetic algorithm is proposed. A comparison of label methods for association rule clustering is presented by [4]. The authors of [4] evaluate the method ideas presented in [20,10,18]. They also use an adaptation of an idea presented by [19] and the method proposed in [16] applied in the context of the document cluster. More detailed results may be found in [6]. Finally, to objectively evaluate labeling methods for association rule clustering, [4] propose two measures, P and RF, the ones used by [15] and, therefore, described in Section 2.

7 Conclusions

In this paper, we performed a subjective evaluation of labeling methods used for association rule clustering. The results show that the GLM method is the most suitable method to be used for NSDS according to the user's views. On the other hand, LM-T is the most suitable method to be used for SDS according to the user's views. However, in the last case, the results obtained from the evaluation suggests that it is more difficult to identify the intruder in standardized data sets. This result may be explained due to the fact that the users did not have a good knowledge of the domain (related to customer profile) but more experiments are necessary to identify the causes of this behavior.

Considering all the results, we may affirm that a good objective evaluation does not imply in a good subjective evaluation. As noted by [7] for the topic extraction task, there is, with a certain frequency, a high disagreement between the two kinds of evaluation. The results also indicate that the organization of the data set has a high impact on the quality of the results.

As future works, we intend to evaluate hybrid labeling methods that combine objective and subjective aspects.

Acknowledgements. Grants 2009/16142-3, 2010/07879-0, 2011/19850-9, 2013/12383-0, São Paulo Research Foundation (FAPESP) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (Grants DS-6345378/D, DS-8434242/M).

References

1. Agrawal, R., Imielinski, T., Swami, A.N.: Mining association rules between sets of items in large databases. In: Proceedings of the ACM SIGMOD International Conference on Management of Data, pp. 207–216 (1993)
2. Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., Verkamo, A.I.: Fast discovery of association rules. In: Advances in Knowledge Discovery and Data Mining, pp. 307–328. AAAI Press (1996)
3. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules. In: Proceedings of the 20th International Conference on Very Large Data Bases, pp. 487–499 (1994)
4. Carvalho, V.O., Biondi, D.S., Santos, F.F., Rezende, S.O.: Labeling methods for association rule clustering. In: Proceedings of the 14th International Conference on Enterprise Information Systems, pp. 105–111 (2012)
5. Carvalho, V.O., Santos, F.F., Rezende, S.O.: Post-processing association rules with clustering and objective measures. In: Proceedings of the 13th International Conference on Enterprise Information Systems, pp. 54–63 (2011)

6. Carvalho, V.O., Santos, F.F., Rezende, S.O.: Agrupamento de regras de associação no pré-processamento e no pós-processamento: O que vale mais a pena? Technical Report 381, Instituto de Ciências Matemáticas e de Computação - ICMC - USP (2012)
7. Chang, J., Boyd-Graber, J.L., Gerrish, S., Wang, C., Blei, D.M.: Reading tea leaves: How humans interpret topic models. In: Neural Information Processing Systems, pp. 288–296 (2009)
8. Han, J., Cheng, H., Xin, D., Yan, X.: Frequent pattern mining: Current status and future directions. Data Mining and Knowledge Discovery 15(1), 55–86 (2007)
9. Hipp, J., Mangold, C., Güntzer, U., Nakhaeizadeh, G.: Efficient rule retrieval and postponed restrict operations for association rule mining. In: Chen, M.-S., Yu, P.S., Liu, B. (eds.) PAKDD 2002. LNCS (LNAI), vol. 2336, pp. 52–65. Springer, Heidelberg (2002)
10. Jorge, A.: Hierarchical clustering for thematic browsing and summarization of large sets of association rules. In: Proceedings of the 4th SIAM International Conference on Data Mining, 10 p. (2004)
11. Liu, B.: Association rules and sequential patterns. In: Web Data Mining, pp. 17–62 (2011)
12. Manning, C.D., Raghavan, P., Schütze, H.: An Introduction to Information Retrieval. Cambridge University Press, 544 p.(2009)
13. Moura, M.F., Rezende, S.O.: A simple method for labeling hierarchical document clusters. In: Proceedings of the 10th IASTED International Conference on Artificial Intelligence and Applications, pp. 336–371 (2010)
14. Natarajan, R., Shekar, B.: Interestingness of association rules in data mining: Issues relevant to e-commerce. SĀDHANĀ – Academy Proceedings in Engineering Sciences (The Indian Academy of Sciences) 30(Parts 2&3), 291–310 (2005)
15. de Padua, R., de Carvalho, V.O., de Souza Serapião, A.B.: Labeling association rule clustering through a genetic algorithm approach. In: Catania, B., Cerquitelli, T., Chiusano, S., Guerrini, G., Kämpf, M., Kemper, A., Novikov, B., Palpanas, T., Pokorny, J., Vakali, A. (eds.) New Trends in Databases and Information Systems. AISC, vol. 241, pp. 45–52. Springer, Heidelberg (2014)
16. Popescul, A., Ungar, L.: Automatic labeling of document clusters. Unpublished manuscript (2000), `http://www.cis.upenn.edu/popescul/~Publications/popescul00labeling.pdf`
17. Rathinasabapathi, R., Ramesh, G.: Comparison of association rules and decision trees for disease prediction and data mining for improved cardiac care. International Journal of Computer Science and Management Research 2, 1716–1721 (2013)
18. Reynolds, A.P., Richards, G., de la Iglesia, B., Rayward-Smith, V.J.: Clustering rules: A comparison of partitioning and hierarchical clustering algorithms. Journal of Mathematical Modelling and Algorithms 5(4), 475–504 (2006)
19. Sahar, S.: Exploring interestingness through clustering: A framework. In: Proceedings of the IEEE International Conference on Data Mining, pp. 677–680 (2002)
20. Toivonen, H., Klemettinen, M., Ronkainen, P., Hätönen, K., Mannila, H.: Pruning and grouping discovered association rules. In: Workshop Notes of the ECML Workshop on Statistics, Machine Learning, and Knowledge Discovery in Databases, pp. 47–52 (1995)
21. Xu, R., Wunsch, D.: Clustering. IEEE Press Series on Computational Intelligence. Wiley (2008)
22. Zhao, Y., Zhang, C., Cao, L.: Post-mining of Association Rules: Techniques for Effective Knowledge Extraction, 372 p. Information Science Reference (2009)