

Metrics for Association Rule Clustering Assessment

Veronica Oliveira de Carvalho¹(✉), Fabiano Fernandes dos Santos²,
and Solange Oliveira Rezende²

¹ Instituto de Geociências e Ciências Exatas,
UNESP - Universidade Estadual Paulista, Rio Claro, Brazil
veronica@rc.unesp.br

² Instituto de Ciências Matemáticas e de Computação,
USP - Universidade de São Paulo, São Carlos, Brazil
{fabianof,solange}@icmc.usp.br

Abstract. Issues related to association mining have received attention, especially the ones aiming to discover and facilitate the search for interesting patterns. A promising approach, in this context, is the application of clustering in the pre-processing step. In this paper, eleven metrics are proposed to provide an assessment procedure in order to support the evaluation of this kind of approach. To propose the metrics, a subjective evaluation was done. The metrics are important since they provide criteria to: (a) analyze the methodologies, (b) identify their positive and negative aspects, (c) carry out comparisons among them and, therefore, (d) help the users to select the most suitable solution for their problems. Besides, the metrics do the users think about aspects related to the problems and provide a flexible way to solve them. Some experiments were done in order to present how the metrics can be used and their usefulness.

Keywords: Association rules · Pre-processing · Clustering · Evaluation metrics

1 Introduction

Association has been highlighted, among the data mining tasks, due to its comprehensibility even by non-experts in the field. To have an idea, the *Apriori* algorithm, broadly used to obtain association patterns, was elected as one of the ten data mining algorithms most employed by the community [1]. For this and other reasons, association has been applied in many domains, as noticed in [2–6] works.

In the last years, researches have adopted some strategies to aid the user to identify the relevant associative patterns of the domain. One of these strategies is to pre-process the data before obtaining the rules. For that, many approaches have been proposed, being clustering a promising one. In this case, as seen in Fig. 1, the data are initially grouped into n groups (GD_1, GD_2, \dots, GD_n). Association rules are extracted within each group and, in the end, n groups of rules are

obtained $(GR_1, GR_2, \dots, GR_n)$. All these rules compose the rule set. According to [7], each group expresses its own associations without the interference of the other groups that contain different association patterns. The aim is to obtain potentially interesting rules that would not be extracted from unpartitioned data sets¹. The user must set the minimum support to a low value to discover these same patterns from unpartitioned data sets, causing a rapidly increase in the number of rules. Recent works have used these ideas in different domains as [8] in maintenance systems, [9] in banking context and [10, 11] in automotive data.

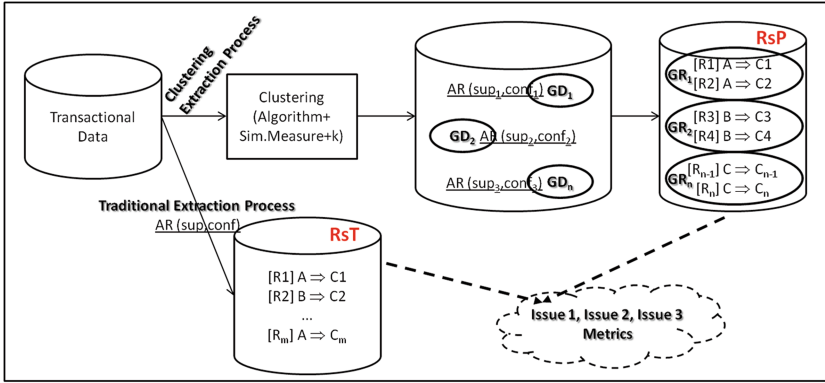


Fig. 1. Overview of the process to extract association rules through clustering in comparison to the traditional one.

Distinct methodologies have been proposed to enable the described process. Each methodology uses a different combination of similarity measures with clustering algorithms to obtain the groups of rules (see Fig. 1). However, little has been done to analyze the performance of the methodologies or even to compare the results. For example, given a specific problem of a domain, how could the user identify the most suitable methodology to use in his problem? Or, how the user could check if the selected methodology was good enough for the problem, considering that different interests may be important for his decision? To aid the user in these tasks, considering the aims of the clustering process above discussed, there are some issues that have to be investigated (see Fig. 1):

Issue 1. Is there overlap between a rule set obtained through partitioned data, i.e., extracted from clustered data, in relation to a rule set obtained through unpartitioned data, i.e., extracted from traditional process? A *r*ule *s*et obtained through a *p*artitioned data is named here as *RsP* and a *r*ule *s*et obtained through a *t*raditional process is named here as *RsT*.

¹ In this work, it is assumed that a pattern is interesting if it is relevant and/or useful to the user – rules having high support and/or high confidence are not necessarily interesting to the user.

Issue 2. Is there overlap between the rules in RsT and RsP regarding the interesting knowledge? In other words, has RsP, in fact, more interesting patterns than RsT?

Issue 3. What is the process behavior regarding the number of rules that are obtained in RsP?

Based on the exposed arguments and on the three presented issues, eleven metrics are proposed in this paper to provide an assessment procedure in order to support the evaluation of the methodologies that use clustering in the pre-processing step. However, to propose the metrics, a subjective evaluation was done with some users to understand each issue. The analysis was done as a way to comprehend the important aspects to be considered by a user when comparing the performance of some methodologies or even their results. Thereby, this paper contributes with current researches since the metrics provide criteria to: (a) analyze the methodologies, (b) identify their positive and negative aspects, (c) carry out comparisons among them and, therefore, (d) help the users to select the most suitable solution for their problems. Besides, the paper does the users think about many aspects to be considered in the presented context and provides to them a flexible way to explore their problems. This paper is an extension of [12]’s work, in which two more metrics are proposed (M_P , M_{RF}) and where the subjective evaluation is used to attest the importance of the previous obtained results.

This paper is organized as follows. Section 2 surveys the related researches. Section 3 presents the subjective evaluation and Sect. 4 the proposed metrics. Section 5 describes some experiments that were carried out to show how the metrics can be used. Section 6 discusses the results obtained in the experiments. Conclusion is given in Sect. 7.

2 Related Works

There are many researches that initially cluster data aiming to discover and facilitate the search for the interesting pattern of the domain. Some of these works are described below.

Plasse et al. [11] propose to split the transactions’ items into groups in order to extract the rules. The aim is to find interesting associations among rare items (less frequent) that would not be discovered if the traditional process were applied, mainly in sparse data. Besides, the authors show that if the data are clustered it is possible to reduce both the amount and the complexity of obtained rules compared to non-clustered data. However, these results depend on the used clustering algorithms and similarity measures. The authors evaluate many hierarchical algorithms (Single, Complete, Average, Ward, Varclus) combined with many similarity measures (Jaccard, Russel Rao, Dice, Ochiai, Pearson). Nevertheless, it is not understandable how the rules are obtained within the groups, since it is necessary to have a set of transactions and not a set of items. This means that it is not clear how the transactions are distributed over the groups. Among the similarity

measures used by them, we emphasize Jaccard (Eq. 1) and Russel Rao (Eq. 2) – the Russel Rao due to its good performance in the experiments presented in [11] and the Jaccard due to its use by the measure described below (Agg). The Jaccard between two items i_1 and i_2 , named here as P-J (*Plasse Jaccard*), is the ratio between the transactions t the items cover simultaneously and the total of transactions the items cover. An item covers a transaction t if the item is in t . The Russel Rao between two items i_1 and i_2 , named here as P-RR (*Plasse Russel Rao*), is computed considering the transactions t the items cover simultaneously over all the transactions. In fact, this measure is equivalent to the support of $i_1 \cap i_2$. In Eq. 2, N_t represents the total number of transactions.

$$\text{P-J}(i_1, i_2) = \frac{|\{t \text{ covered by } i_1\} \cap \{t \text{ covered by } i_2\}|}{|\{t \text{ covered by } i_1\} \cup \{t \text{ covered by } i_2\}|}. \quad (1)$$

$$\text{P-RR}(i_1, i_2) = \frac{|\{t \text{ covered by } i_1\} \cap \{t \text{ covered by } i_2\}|}{N_t}. \quad (2)$$

Aggarwal et al. [13] propose an algorithm, named CLASD, to split the transactions aiming to discover associations on small segments (subsets) of the data. The authors justify that the approach has a considerable impact on the rules that are obtained, since the patterns that cannot be identified in the whole set can be identified in the subsets. To cluster the transactions, [13] use a similarity measure proposed by them (Eq. 3), named here as Agg (*Aggarwal*). The clustering is done by CLASD. In this case, the similarity between two transactions t_1 and t_2 is computed by the affinity (Af) average among the transactions' items. It can be noted that the affinity (Af) is equivalent to the measure P-J. Therefore, after computing P-J among the m items in t_1 and the n items in t_2 , the average among them is obtained. The higher the affinity among the items the more similar the transactions. From this initially clustering the rules are then extracted. Different from [11], the number of rules that are extracted by their process is higher compared to the traditional process, but smaller if the support in the traditional process had to be set with a very low value to obtain the same associations. A limitation of the approach is the number of parameters that has to be set (five plus two constants).

$$\text{Agg}(t_1, t_2) = \frac{\sum_{p=1}^m \sum_{q=1}^n \text{Af}(i_p, j_q)}{m * n}, \text{Af}(i,j) = \frac{\text{sup}(\{i, j\})}{\text{sup}(\{i\}) + \text{sup}(\{j\}) - \text{sup}(\{i, j\})}. \quad (3)$$

Koh and Pears [7] propose a methodology to cluster transactions and then extract a rare association rule set. The algorithm they proposed, named *Apriori Inverse*, applies upon groups of transactions in order to verify if it is possible to extract rare rules that are not discovered considering the whole data. Therefore, the idea is similar to the works above described. The authors demonstrate that the clustering contributes to the discovery of new associations. To cluster the transactions, their algorithm initially finds k seeds (centroids), where k indicates the number of frequent itemsets in transactions that match some conditions. Each seed forms a group. After the seed generation, each transaction t is allocated to one of the groups (seed) considering the similarity obtained through

the measure presented in Eq. 4. The equation is an adaptation of the Jaccard coefficient and it compares the number of common items that occur between the transaction (t) and the group centroid (c_k). In this case, the higher the overlap between t and a centroid the higher the similarity value ($Sim =$ similarity). In this approach it is necessary to set two additional values beyond the minimum support, hindering a little the exploration.

$$Sim(t, c_k) = \frac{|t \cap c_k|}{|t \cup c_k| - |t \cap c_k| + 1}. \quad (4)$$

There are other researches concerned with the clustering of transactions that, although not related to the extraction of association rules, could be used for that purpose [14–17]. In [18] the authors propose an approach to identify, a priori, the potentially interesting items to appear in the antecedents and in the consequents of the association rules without extracting them. The approach is divided in two steps: the clustering of the transactions and the selection of the interesting items. To do the clustering the authors propose the use of incremental K-means with the similarity measure presented in Eq. 5, named here as Denza. Note that this measure is the Jaccard between transactions. Therefore, the similarity between two transactions t_1 and t_2 is computed considering the items the transactions share. After the grouping, statistics are applied upon the groups to identify the items that are relevant to the application.

$$Denza(t_1, t_2) = \frac{|\{items\ in\ t_1\} \cap \{items\ in\ t_2\}|}{|\{items\ in\ t_1\} \cup \{items\ in\ t_2\}|}. \quad (5)$$

Among the papers above described, little has been done to analyze the performance of the methodologies, allowing to identify their positive and negative aspects, or even to compare the results among them. In general, the researchers compare the number of rules and/or itemsets that are obtained from unpartitioned data and clustered data to expose the usefulness of the methodologies. This strategy can be found in [7, 11, 13] and is related to “Issue 3” of Sect. 1. However, [11] also analyze the process considering the complexity of the rules that are obtained – the greater the number of items that compose a rule the higher its complexity. Koh and Pears [7] and Aggarwal et al. [13] discuss about some rules found through clustering to show that the process provides the discovery of interesting patterns, but the analysis of the process is subjective. Aggarwal et al. [13] also consider the execution time. Finally, [7] is the only work that allows a better analysis considering the existing overlap between the rules obtained from unpartitioned data and clustered data. This strategy is related to “Issue 1” of Sect. 1. Based on the presented arguments, as mentioned before, the necessity of an assessment procedure becomes evident.

It is important to mention that cluster validation indices (CVI), independently of being relative, internal, or external [19, 20], are not reviewed here. It is understood that these CVI can be used to evaluate the groups of transactions before obtaining the association rules: in this case, the aim is to check if the partition fits the structure underlying the data. However, after obtaining the groups of transactions, the rules are obtained and, therefore, the CVI can no longer be

used, since the data used as input (for example, transactions) must be the same of the output (for example, groups of transactions); however, in this case, the input are transactions and the output are rules. Thus, the metrics presented here are important as a final step to evaluate the usefulness of a clustering process in obtaining groups of rules. Therefore, the works described in this section are the ones that aim to apply clustering as a previous step to obtain association rules and their limitations regarding an assessment procedure.

3 Understanding the Issues Through a Subjective Evaluation

A subjective evaluation was done with some users to understand each issue before proposing the metrics. The analysis was done as a way to comprehend the important aspects to be considered by a user when comparing the performance of some methodologies or even their results.

The evaluation was done through a questionnaire, which was elaborated to present to the users scenarios that can occur in the considered context. For each scenario a question was formulated in order to comprehend the importance of the aspects related to the exposed problem. The full questionnaire can be found in *Appendix*. Before presenting the Questions, a brief introduction was given to the users about the questionnaire. The questionnaire comprehends 10 Questions: Questions 1 and 2 are related to “Issue 1”, Questions 3, 4, 5 and 6 to “Issue 2” and Question 7 to “Issue 3”. Question 8 is related to a more general advantage the clustering process can provide. Questions 9 and 10 are general questions about the evaluation. The questionnaire was answered by 5 users, all of them researches on data mining and/or text mining. All of them have knowledge on clustering and association rules, including their particularities, problems, means of validation, etc., and, so, they could understand the exposed problem and contribute with the current research. A brief overview of their experiences is here presented: (user (1)) worked with objective measures to post-process association rules obtained to extract related words in a collection of documents to construct a enriched bag-of-words representation; worked with clustering to construct a topic hierarchy based on features extracted through association rules; (user (2)) worked with clustering to post-process association rules and developed a labeling method to label groups of rules; worked with complex networks and semi-supervised learning to post-process association rules; (user (3)) worked with clustering and semi-supervised learning to build a dynamic topic hierarchy; worked with association rules in some applications; (user (4)) worked with objective and subjective measures to post-process association rules; worked with clustering to organize document collections; (user (5)) worked with taxonomies to generalize association rules to help users during the post-processing phase; worked with association rules to develop recommendation systems; worked with clustering in some applications. The users’ answers are presented in Table 1, which are discussed below along with each Question.

Table 1. Users’ evaluation about the exposed scenarios.

Question	User				
	1	2	3	4	5
1	desirable	desirable	desirable	desirable	desirable
2	desirable	desirable	desirable	desirable	indifferent
3	desirable	desirable	indifferent : no	desirable	desirable
4	indifferent : no	indifferent	indifferent : no	desirable	desirable
5	indifferent : no	desirable	desirable	desirable	indifferent
6	desirable	no desirable	desirable	no desirable	desirable
7	low	high	low	average	average
8	yes	yes	yes	yes	yes

Question 1 (Issue 1). The idea behind this question, presented in Fig. 2, was to analyze if it is important not to lose knowledge during the clustering process. The other idea was to analyze, regarding the repetition over the clusters, if the knowledge must be in subsets that express its own associations. All of them considered that is *desirable* that rules extracted in RsT be found in RsP, directing the analysis to the aspects embedded in question. Besides, all of them stated that this aspect is important for an assessment procedure (1.a.) – the “yes” answers, in all the questions, regarding the importance of the considered aspect, are omitted in Table 1. In relation to the comments made by the users (1.b.), we can mention: (a) the necessity to differentiate the rules that don’t repeat over the clusters from the ones that repeat; (b) the understanding that if the patterns are present in both sets it is because they are relevant (even if they, probably, represent a frequent relation of the domain), being desirable that this scenario occurs. To formalize these aspects, two metrics are proposed in the next section: M_{O-RsP} and $M_{R-O-RsP}$.

1. In your opinion, observing "Scenario-A" below, how do you consider the occurrence of rules obtained in RsT in RsP (cases in green and orange)? Both the cases, green and orange, represent rules obtained in both of the sets, but the rules in orange are extracted more than once in RsP along the groups. If needed to distinguish the green cases of the orange cases, please let it indicated.
 desirable indifferent no desirable

a. Do you think important to consider this scenario in an assessment procedure to be used in the presented context?
 yes no

b. Would you like to make any comment about the scenario (advantage, disadvantage, etc.)?

Fig. 2. Question 1 (see *Appendix* for details).

Question 2 (Issue 1). The idea behind this question, presented in Fig. 3, was to analyze if it is important to obtain new knowledge during the clustering process. The other idea was to analyze, regarding the repetition over the clusters, if the new knowledge must be in subsets that express its own associations. Almost all of them (four out five) considered that is *desirable* that rules extracted in RsP be not found in RsT, directing the analysis to the aspects embedded in question. However, all of them stated that this aspect is important for an assessment procedure (2.a.). In relation to the comments made by the users (2.b.), we can mention: (a) the necessity to differentiate the rules that don't repeat over the clusters from the ones that repeat; (b) the understanding that the patterns appearing only in RsP represent new knowledge, probably interesting to the users, that would be difficult to be found through the traditional process. To formalize these aspects, two metrics are proposed in the next section: M_{N-RsP} and $M_{R-N-RsP}$.

<p>2. In your opinion, observing "Scenario-A" below, how do you consider the non-occurrence of rules obtained in RsP in RsT (cases in purple and red)? Both the cases, purple and red, represent rules obtained only in RsP, but the rules in red are extracted more than once in RsP along the groups. If needed to distinguish the purple cases of the red cases, please let it indicated.</p> <p><input type="checkbox"/> desirable <input type="checkbox"/> indifferent <input type="checkbox"/> no desirable</p> <p>a. Do you think important to consider this scenario in an assessment procedure to be used in the presented context?</p> <p><input type="checkbox"/> yes <input type="checkbox"/> no</p> <p>b. Would you like to make any comment about the scenario (advantage, disadvantage, etc.)?</p>

Fig. 3. Question 2 (see *Appendix* for details).

Question 3 (Issue 2). The idea behind this question, presented in Fig. 4, was to analyze if it is important to discovery new *interesting* knowledge; if so, the cost of the clustering process would be minimized since new interesting knowledge would be found. Almost all of them (four out five) considered that is *desirable* that some (or none) of the n most interesting rules in RsP be not found in RsT, directing the analysis to the aspect embedded in question. Besides, almost all of them (four out five) stated that this aspect is important for an assessment procedure (3.a.). In relation to the comments made by the users (3.b.), we can mention: (a) the importance of identifying new interesting rules not extracted through the traditional process; (b) the understanding that these new interesting patterns are the ones that stand in a certain group of data. To formalize this aspect, one metric is proposed in the next section: $M_{N-I-RsP}$.

Question 4 (Issue 2). The idea behind this question, presented in Fig. 5, was to analyze if it is important not to lose interesting knowledge during the clustering process. Most of them (three out five) considered that is *indifferent* that some (or none) of the n most interesting rules in RsT be not found in RsP; thus, in this case, any conclusion was done in relation to the aspect embedded in question. However, most of them (three out five) stated that

3. In your opinion, observing "Scenario-B" below, how do you consider the non-occurrence of some (or none) of the n more interesting rules in RsP in RsT (cases in blue)? Notice that the blue rules belong only to the RsP set.
 desirable indifferent no desirable

a. Do you think important to consider this scenario in an assessment procedure to be used in the presented context?
 yes no

b. Would you like to make any comment about the scenario (advantage, disadvantage, etc.)?

Fig. 4. Question 3 (see *Appendix* for details).

this aspect is important for an assessment procedure (4.a.). In relation to the comments made by the users (4.b.), there were few relevant observations. Only one user mentioned that it is important to analyze the considered scenario because it is understood that the clustering process could be losing knowledge, making the user to think about the truth quality of the clusters (directing to the aspect embedded in question). Although no conclusion could be made in relation to the main question, considering the importance for an assessment procedure (4.a.), one metric is proposed in the next section to formalize the aspect: $M_{O-I-N-RsP}$.

4. In your opinion, observing "Scenario-B" below, how do you consider the reverse scenario? This is, the non-occurrence of some (or none) of the n more interesting rules in RsT in RsP (cases in orange)? Notice that the orange rules belong only to the RsT set.
 desirable indifferent no desirable

a. Do you think important to consider this scenario in an assessment procedure to be used in the presented context?
 yes no

b. Would you like to make any comment about the scenario (advantage, disadvantage, etc.)?

Fig. 5. Question 4 (see *Appendix* for details).

Question 5 (Issue 2). The idea behind this question, presented in Fig. 6, was to analyze if it is important that an intersection exists between the new interesting knowledge, discovery through the clustering process, in relation to the interesting knowledge already found through the traditional process. Most of them (three out five) considered as *desirable* the existing intersection between the n most interesting rules in RsP and the n most interesting rules in RsT, directing the analysis to the aspect embedded in question. Besides, almost all of them (four out five) stated that this aspect is important for an assessment procedure (5.a.). In relation to the comments made by the users (5.b.), we can mention: (a) the understanding that interesting rules appearing in both sets have a high probability of being really interesting to the users; (b) however, as an opposed idea, the understanding that different knowledge must be extracted from the clustering process compared to the traditional one (*indifferent* cases). Thus, although distinct views exist (5.b.),

considering the importance for an assessment procedure (5.a.), one metric is proposed in the next section to formalize the aspect: M_{C-I} (which can be interpreted according to the user's view).

5. In your opinion, observing "Scenario-B" below, how do you consider the existing intersection between the n more interesting rules in RsP and the n more interesting rules in RsT (cases in red)?
 desirable indifferent no desirable
- a. Do you think important to consider this scenario in an assessment procedure to be used in the presented context?
 yes no
- b. Would you like to make any comment about the scenario (advantage, disadvantage, etc.)?

Fig. 6. Question 5 (see *Appendix* for details).

Question 6 (Issue 2). The idea behind this question, presented in Fig. 7, was to analyze if it is important that the new interesting knowledge be found in a small number of the groups. Most of them (three out five) considered as *desirable* the spread of the n most interesting rules in RsP in a small number of clusters, directing the analysis to the aspect embedded in question. Besides, all of them stated that this aspect is important for an assessment procedure (6.a.). In relation to the comments made by the users (6.b.), we can mention: (a) the understanding that the user could explore a group of interesting knowledge, directing the user to a reduced exploration space; (b) however, as an opposed idea, the understanding that it could be more relevant to explore the interesting rules within each cluster, since each group express different concepts of the domain (*no desirable* cases). Thus, although distinct views exist (6.b.), considering the importance for an assessment procedure (6.a.), one metric is proposed in the next section to formalize the aspect: $M_{NC-I-RsP}$ (which can be interpreted according to the user's view).

6. In your opinion, how do you would consider the spread of the n more interesting rules in RsP in a small number of clusters?
 desirable indifferent no desirable
- a. Do you think important to consider this scenario in an assessment procedure to be used in the presented context?
 yes no
- b. Would you like to make any comment about the scenario (advantage, disadvantage, etc.)?

Fig. 7. Question 6 (see *Appendix* for details).

Question 7 (Issue 3). The idea behind this question, presented in Fig. 8, was to analyze the clustering process regarding the number of extracted rules in relation to the traditional one. *Two* of them considered that the amount of rules to be extracted through clustering, compared to the traditional process, should be *low*, *two* of them *average* and *one high*. However, all of them

stated that this aspect is important for an assessment procedure (7.a.). In relation to the comments made by the users (7.b.), we can mention: (a) the understanding that, since each group contains a specific bunch of items, the amount of rules should be low (*low* cases); (b) the understanding that each group will contain a small number of rules (due to the same arguments of (a)) and, as a consequence, an average amount considering all the groups (*average* cases); (c) the understanding that since each group will contain similar items, the amount of rules should be high (due to the low diversity of items compared to the traditional process considering the same value of support/confidence) (*high* cases). Thus, although distinct views exist (7.b.), considering the importance for an assessment procedure (7.a.), one metric is proposed in the next section to formalize the aspect: M_{NR-RsP} (which can be interpreted according to the user's view).

<p>7. In your opinion, do you consider that the amount of rules to be extracted through clustering, compared to the traditional process, should be: <input type="checkbox"/> low <input type="checkbox"/> average <input type="checkbox"/> high</p> <p>a. Do you think important to consider this scenario in an assessment procedure to be used in the presented context? <input type="checkbox"/> yes <input type="checkbox"/> no</p> <p>b. Would you like to make any comment about the scenario (advantage, disadvantage, etc.)?</p>
--

Fig. 8. Question 7.

Question 8 (Others). The idea behind this question, presented in Fig. 9, was to analyze the clustering process regarding the spread of the concepts of the domain. All of them considered that is *desirable* that the clustering process should, as a consequence, enable each cluster to express a distinct topic of the domain. Besides, all of them stated that this aspect is important for an assessment procedure (8.a.). In relation to the comments made by the users (8.b.), we can mention, unanimously, the understanding that the aim of the clustering process is to split the distinct topics of the domain in order to aid rule exploration. To formalize this aspect, two metrics are proposed in the next section: M_P and M_{RF} .

<p>8. In your opinion, only in relation to RsP, do you consider that the clustering process should, as a consequence, enable each cluster to express a distinct topic of the domain? <input type="checkbox"/> yes <input type="checkbox"/> indifferent <input type="checkbox"/> no</p> <p>a. Do you think important to consider this scenario in an assessment procedure to be used in the presented context? <input type="checkbox"/> yes <input type="checkbox"/> no</p> <p>b. Would you like to make any comment about the scenario (advantage, disadvantage, etc.)?</p>

Fig. 9. Question 8.

Question 9 and 10 (Others). Questions 9 and 10 are general questions about the evaluation. In relation to “Question 9”, presented in Fig. 10, none of them gave any suggestion. However, in relation to “Question 10”, also presented in Fig. 10, two comments were made: (a) the understanding that an assessment procedure should consider two or more scenarios but maintaining a good trade-off among them; (b) the understanding that the presentation of the rules in different scenarios, as the ones exposed, can aid the user during an evaluation process, since the user can give more attention to some aspects according to his aims. Therefore, as observed, this paper contributes with current researches making the users think about some aspects to be considered in the presented context and providing to them a flexible way to explore their problems (since the users can interpret a metric according to their view).

<p>9. Can you identify other scenario(s), not previously explored, that can be relevant to the presented context? Give an example of the scenario(s) that you identified.</p> <p>a. Do you think important to consider this(these) scenario(s) in an assessment procedure to be used in the presented context? <input type="checkbox"/> yes <input type="checkbox"/> no</p> <p>10.If you want to leave any comment/observation, please do it below.</p>
--

Fig. 10. Questions 9 and 10.

4 Evaluation Metrics: Providing an Assessment Procedure

Eleven metrics are proposed to provide an assessment procedure in order to support the evaluation of the methodologies that use clustering in the pre-processing step (as the ones described in Sect. 2). These metrics formalize the aspects related to each issue, which were analyzed by some users through a subjective evaluation. Each metric, which was implicit explored through a subjective question, is related to an issue. For each issue there are one or more metrics. All metrics, with exception to M_{NR-RsP} , range from 0 to 1. Since RsP contains all the rules extracted within each group, repeated rules may exist in the set; therefore, RsP can be, in some cases, a multiset. In RsT the same doesn't occur since the rules are unique.

Issue 1. Regarding the existing overlap among the rules in RsP and RsT, four metrics are proposed, which are described as follows:

M_{O-RsP} Related to “Question 1”, measures the ratio of “old” rules in RsP, i.e., the ratio of rules in RsT found in RsP (Eq. 6). A rule is considered “old” if it is in RsT, i.e., in the rule set obtained through the traditional process. Therefore, considering users’ analysis, the higher the value the better the

metric, since the value indicates that there was no loss of knowledge during the process.

$$M_{O-RsP} = \frac{|RsT \cap RsP|}{|RsT|}. \quad (6)$$

$M_{R-O-RsP}$ Related to “Question 1”, measures the ratio of “old” rules that repeat in RsP (Eq. 7). In this work, it is assumed that a rule should exist in only one of the clustering groups, since it has to be in a subdomain that expresses its own associations. Therefore, considering users’ analysis and authors’ understanding on the problem, the lower the value the better the metric, since the value indicates that the knowledge, already known, is in subsets that express its own associations.

$$M_{R-O-RsP} = \frac{FindRepetitionRsP(RsT \cap RsP)}{|RsT \cap RsP|}, \quad (7)$$

FindRepetitionRsP: function that receives by parameter a set of non repeated rules and returns the number of rules in the set that repeat in RsP.

M_{N-RsP} Related to “Question 2”, measures the ratio of “new” rules in RsP, i.e., the ratio of rules in RsP not found in RsT (Eq. 8). A rule is “new” if it isn’t in RsT, i.e., in the rule set obtained through the traditional process. Although it is important that any knowledge be lost (metric M_{O-RsP}), it is expected that the ratio of “new” rules in RsP be greater than the ratio of “old” rules. Therefore, considering users’ analysis, the higher the value the better the metric, since the value indicates the amount of knowledge, previously unknown, obtained during the process.

$$M_{N-RsP} = \frac{|RsP^* - RsT|}{|RsP|}, \quad (8)$$

RsP^* : all rules in RsP found in RsT are considered “old” rules and, therefore, are computed as having only one occurrence in RsP.

$M_{R-N-RsP}$ Related to “Question 2”, measures the ratio of “new” rules that repeat in RsP (Eq. 9). Idem to $M_{R-O-RsP}$. Therefore, considering users’ analysis and authors’ understanding on the problem, as in $M_{R-O-RsP}$, the lower the value the better the metric, since the value indicates that the knowledge, previously unknown, is in subsets that express its own associations.

$$M_{R-N-RsP} = \frac{FindRepetition(RsP^* - RsT)}{|RsP^* - RsT|}, \quad (9)$$

FindRepetition: function that receives by parameter a set that may contain repeated rules and returns the number of rules in the set that repeat; RsP^* : idem Eq. 8.

Issue 2. Regarding the existing overlap among the rules in RsP and RsT considering the interesting aspect of the knowledge, four metrics are proposed, which are described as follows:

$M_{N-I-RsP}$ Related to “Question 3”, measures the ratio of “new” rules among the h -top interesting rules in RsP (Eq. 10). Given a subset of h -top interesting rules, selected from RsP, it is expected that the ratio of “new” rules in this subset be as large as possible. The h -top rules are the h rules that contain the highest values regarding an objective measure, where h is a number to be chosen². Therefore, considering users’ analysis and authors’ understanding on the problem, the higher the value the better the metric, since the value indicates that the cost of the process is minimized by the discovery of interesting knowledge, previously unknown, in RsP.

$$M_{N-I-RsP} = \frac{\text{CountTopRules}(h_{top} \text{ of } RsP, RsP^* - RsT)}{|h_{top} \text{ of } RsP|}, \quad (10)$$

CountTopRules: function that receives by parameter a set of h -top interesting rules and a set of rules and returns the number of rules that appears among the h -top; RsP^* : idem Eq. 8.

$M_{O-I-N-RsP}$ Related to “Question 4”, measures the ratio of “old” rules not in RsP among the h -top interesting rules in RsT (Eq. 11). Given a subset of h -top interesting rules, selected from RsT, it is expected that all these rules are present in RsP. Since any conclusion could be done considering users’ analysis, it is understood, in this work, that is not desirable that patterns in RsT disappear in RsP, which would imply in the loss of relevant knowledge. Thus, this metric measures the ratio of “old” interesting rules not in RsP. The h -top rules are as described in $M_{N-I-RsP}$. Therefore, considering the importance of this aspect for an assessment procedure, according to users’ view, and authors’ understanding on the problem, the lower the value the better the metric, since the value indicates that the interesting knowledge in RsT was not lost during the process.

$$M_{O-I-N-RsP} = \frac{\text{CountTopRules}(h_{top} \text{ of } RsT, RsT - RsP)}{|h_{top} \text{ of } RsT|}, \quad (11)$$

CountTopRules: idem Eq. 10.

M_{C-I} Related to “Question 5”, measures the ratio of common rules among the h -top interesting rules in RsP and the h -top interesting rules in RsT (Eq. 12). Consider two subsets, S_1 and S_2 , containing, respectively, the h -top interesting rules in RsP and the h -top interesting rules in RsT. This metric measures the existing intersection between these two subsets, which is expected to be as large as possible according to the users (different from what was assumed in [12], that presented a metric interpretation as in the *indifferent* cases (see the justification below)). Therefore, the higher the value the better the metric, since the users understand that rules present in both sets have a high probability of being really interesting to them. However, the value of the metric can be interpreted according to the user’s view, being the lower the value the better the metric (*indifferent* cases – in this condition,

² Any other criteria could be adopted to select the h -top interesting rules.

the process would not provide any additional relevant information, since all the knowledge already known as interesting in RsT is also identified as interesting in RsP, as understood in [12]). Therefore, as noticed, the metric provides a flexible way to analyze the problem.

$$M_{C-I} = \frac{|h_{top} \text{ of } RsP \cap h_{top} \text{ of } RsT|}{h}, \quad (12)$$

h is the number to be chosen to realize the selection of the rules in both sets, i.e., RsP and RsT .

$M_{NC-I-RsP}$ Related to “Question 6”, measures the ratio of groups in the clustering related to RsP that contains the h -top interesting rules in RsP (Eq. 13). Therefore, considering users’ analysis, the lower the value the better the metric. This means that just some of the groups would have to be explored by the user, which will contain the “new” relevant knowledge extracted during the process. However, the value of the metric can be interpreted according to the user’s view, being the higher the value the better the metric (*indifferent* cases – in this condition, each cluster would express its own interesting knowledge). As noticed, the metric provides a flexible way to analyze the problem.

$$M_{NC-I-RsP} = \frac{FindGroups(h_{top} \text{ of } RsP)}{N}, \quad (13)$$

N : number of groups in the clustering; $FindGroups$: function that receives by parameter a set of h -top interesting rules, finds their groups and returns the number of distinct selected groups.

Issue 3. Regarding the process behavior related to the number of rules that are obtained in RsP, a unique metric is proposed, which is described as follows:

M_{NR-RsP} Related to “Question 7”, measures the ratio of rules in RsP in relation to RsT (Eq. 14). It is important to analyze the process in relation to the number of rules in RsP. It is not desirable, according to the authors’ understanding, to have a large increase in the volume of rules, because even if new patterns are discovered, it will be harder for the user to identify them. Therefore, the lower the value the better the metric, since the value indicates that although new patterns have been extracted, the number of extracted rules is not big enough to overload the user. However, the value of the metric can be interpreted according to the user’s view, since each user has a different opinion (although all of them agree with the importance of this aspect for an assessment procedure). As noticed, the metric provides a flexible way to analyze the problem.

$$M_{NR-RsP} = \frac{|RsP|}{|RsT|}. \quad (14)$$

Others. “Question 8” tried to capture a more general advantage the clustering process can provide: to enable each cluster expresses a distinct topic of the domain.

According to the users, this is a *desirable* aspect. Carvalho et al. [21] evaluated some labeling methods for association rule clustering through two measures. Precision (P) measures how much a labeling method finds labels that represent as accurately as possible the rules contained in their own groups – if the labels don't represent the knowledge inside each cluster the user will have difficult to explore the existing concepts related to the topics the labels express. Repetition Frequency (RF) measures how the labels are distributed over the clusters – if the labels appear repeatedly over the clusters the user will have difficult to identify the existing topics. As known, a labeling method is applied over the clusters obtained through a clustering process. Thus, considering that a good labeling method exists³, it is assumed that a methodology, in the presented context, provides a good distribution of the topics if it presents high values for P and RF , since which will impact the results is the clustering itself. Therefore, these measures are used as metrics in the considered context, which are described as follows:

M_P Considering that a good labeling method is available, measures how much the labels, built over the obtained clustering, represent the rules contained in the clusters (Eq. 15). The more a methodology provides groups that express their own associations, more specific domain knowledge the groups will contain and, probably, the more the labels will represent the rules in the clusters. Therefore, the higher the value the better the metric, since the value indicates that the methodology succeed to enable a suitable distribution of the domain topics.

$$M_P = \frac{\sum_{i=1}^N P(C_i)}{N}, P(C_i) = \frac{\#\{\text{rules covered in } C_i \text{ by } C_i \text{ labels}\}}{\#\{\text{rules in } C_i\}}, \quad (15)$$

N refers to the number of clusters. A rule is covered (represented) by a set of labels if the rule contains at least one of the labels.

M_{RF} Considering that a good labeling method is available, measures how much the distinct labels, built over the obtained clustering, don't repeat (Eq. 16). The more a methodology provides groups that express their own associations, more specific domain knowledge the groups will contain and, probably, the more distinct the labels will be over the clusters. Therefore, the higher the value the better the metric, since the value indicates that the methodology succeed to enable a suitable distribution of the domain topics.

$$M_{RF} = 1 - \frac{\#\{\text{distinct labels that repeat in the clusters}\}}{\#\{\text{distinct labels in the clusters}\}}. \quad (16)$$

Table 2 summarizes the metrics above described, indicating the suitable use of each one. As noticed, the metrics provide a flexible way to analyze the problem, since the user can analyze the results according to his interests (users can disagree about some aspects), both in relation to the metrics values' interpretation as in relation to the importance (weight) of each one to a given domain – as in other contexts, such as objective measures for association rules [23], where the

³ In this work, it is considered that this labeling method is the one presented by [22].

Table 2. Summary and recommended use of the proposed evaluation metrics.

Metric	Description
$M_{O-RsP}[\uparrow]$	Ratio of “old” rules in RsP Recommended use: to measure the loss of knowledge obtained by the clustering process
$M_{R-O-RsP}[\downarrow]$	Ratio of “old” rules in RsP that repeat Recommended use: to measure if the knowledge, already known, is in subsets that express their own associations
$M_{N-RsP}[\uparrow]$	Ratio of “new” rules in RsP Recommended use: to measure the amount of knowledge, previously unknown, obtained by the clustering process
$M_{R-N-RsP}[\downarrow]$	Ratio of “new” rules in RsP that repeat Recommended use: to measure if the knowledge, previously unknown, is in subsets that express their own associations
$M_{N-I-RsP}[\uparrow]$	Ratio of “new” rules among the h -top interesting rules in RsP Recommended use: to measure the amount of previously unknown patters among the knowledge, obtained by the clustering process, identified as interesting
$M_{O-I-N-RsP}[\downarrow]$	Ratio of “old” rules not in RsP among the h -top interesting rules in RsT Recommended use: to measure the amount of knowledge already known as interesting among the knowledge obtained by the clustering process
$M_{C-I}[\uparrow]$	Ratio of common rules among the h -top interesting rules in RsP and the h -top interesting rules in RsT Recommended use: to measure the amount of common interesting knowledge between the patterns already known as interesting and the patterns previously unknown that were identified as interesting
$M_{NC-I-RsP}[\downarrow]$	Ratio of groups in the clustering related to RsP that contains the h -top interesting rules in RsP Recommended use: to measure if the patterns, previously unknown, identified as interesting, are spread over a small number of groups
$M_{NR-RsP}[\downarrow]$	Ratio of rules in RsP in relation to RsT Recommended use: to measure if the number of patterns obtained by the clustering process don't overload the user
$M_P[\uparrow]$	Behavior of topics' distribution regarding precision Recommended use: to measure if an obtained clustering enables a suitable distribution of the domain topics regarding precision
$M_{RF}[\uparrow]$	Behavior of topics' distribution regarding repetition frequency Recommended use: to measure if an obtained clustering enables a suitable distribution of the domain topics regarding repetition frequency

user chooses the measures (and sometimes their thresholds cut) to evaluate the obtained patterns according to his interests. Besides, the metrics also contribute in the sense of enabling the users to think about important aspects related to the presented context. Finally, relating the proposed metrics by the researchers found in the literature (Sect. 2), it can be observed that: (a) [7] is the only work that provides a similar analysis related to the metrics M_{O-RsP} and M_{N-RsP} in *Issue 1*; (b) none of them provide an analysis related to the aspects covered by *Issue 2* and *Others*; [7, 11, 13] provide a similar analysis related to the metric M_{NR-RsP} in *Issue 3*. Thus, as mentioned in Sect. 2, the necessity of an assessment procedure becomes evident.

5 Experiments

Some experiments were carried out in order to present how the metrics can be used. Therefore, two contexts were defined. Suppose a user decides to apply clustering in the pre-processing step. First of all, he has to find the most suitable methodology to use in his problem. After that, he has to check if the selected methodology was good enough for the problem, considering that different interests may be important for his decision. Thus, two different situations were regarded: (i) identify among some clustering setups the most suitable; (ii) analyze the process itself. A clustering setup is obtained by the application of a clustering algorithm combined with a similarity measure. Therefore, the metrics provide the support to evaluate each situation under the discussed issues: while in (i) the data is initially clustered through some clustering setups in order to identify the one that obtains a good association set, in (ii) the usefulness of the process itself is analyzed. Four data sets and four clustering setups were selected to be used in the experiments. It is important to mention that the choices below could be changed without affecting the paper relevance, since the aim here is to present the metrics and to demonstrate how they can be used, independently of the clustering setup.

The four data sets were 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 in the R Project for Statistical Computing through the package “arules”⁴. The last one was donated by a supermarket located in São Carlos city, Brazil. Adult and Income are relational sets and Groceries and Sup transactional. Therefore, before extracting the rules on Adult and Income, the sets were converted to a transactional format, where each transaction was composed by pairs of the form “attribute=value”.

The four clustering setups were obtained by the combination of the algorithms and similarity measures presented in Table 3. Each combination gives a clustering setup, i.e., a different way to analyze the process. PAM is a partitional medoid-based algorithm that splits n objects in k groups, in which each object is closer to the medoid that defines its own cluster than to the medoid of any

⁴ <http://cran.r-project.org/web/packages/arules/index.html>.

other cluster – a medoid is an object that has the minimal average dissimilarity to all the other objects in the same cluster [24]. Therefore, the algorithm works searching for the k representative objects in n to build the k clusters by assigning each object to its nearest medoid. On the other hand, Ward is an agglomerative hierarchical algorithm that starts by considering each object as a singleton cluster and then, at each iteration, merges the two closest clusters until a single cluster remains. The two closest clusters are the ones that minimize the increase of the within-cluster sum of the squared errors (SSE) [24]. Despite the existence of algorithms designed for transactions, such as ROCK, the choices of the algorithms were made based on works that cluster the rules in the post-processing phase, as [25, 26], aiming a *posteriori* comparison. The similarity measures were chosen considering the works described in Sect. 2 – only the similarities among transactions were selected based on previous experiments. Finally, although it is necessary to set k , the number of groups, in order to obtain a clustering setup, this value was only used to analyze the clustering setups on different views. We understand that even though k is an important parameter, its values were ranged and then averaged (see Sect. 6) without affecting the experiments' results, since the aim of the paper, as mentioned before, is to present some metrics and to demonstrate how they can be used, independently of the clustering setup. Besides, as it will be seen in Sect. 6, almost all the metrics present a low standard deviation, showing a good homogeneity among the results obtained over the values of k ⁵.

As described before, the rules are extracted within each group after clustering the data. The values of the minimum support (min-sup) and minimum confidence (min-conf) have to be set in order to extract a set of association rules. To automate the specification of the min-sup in each group, the following procedure was adopted: (i) find the 1-itemsets of the group with their supports, (ii) compute the average of these supports, (iii) use this average support as the min-sup of the group. This strategy was adopted since some groups can have few transactions and others many transactions, which implies in choosing suitable parameter values to avoid an explosion of rules within a group or the obtainment of an empty rule set. Regarding min-conf, the following values were used for each data set: Adult 50 %; Income 50 %; Groceries 10 %; Sup 100 %. Thus, the same min-conf value was applied in all the groups of a given data set. These values were chosen experimentally. Although it is known that min-sup and min-conf impact on the set of rules that are obtained, it was assumed that the focus was on the use of the metrics and, so, that the values were adequate for the proposed problem (the same argument is also applied to algorithms, similarity measures and k). The rules were extracted with an Apriori implementation developed by Christian Borgelt⁶ with a minimum of two items and a maximum of five items per rule.

⁵ In this work, each dendrogram obtained by Ward were cut considering each one of the values of k .

⁶ <http://www.borgelt.net/apriori.html>.

Considering the four clustering setups, the RsP sets were obtained. However, once almost all the metrics are based on the rules obtained through the traditional process, the four data sets were also processed to obtain the RsT sets. For that, the min-sup was set automatically, as described before. Regarding min-conf, the same values used in RsP were considered, i.e., Adult 50 %, Income 50 %, Groceries 10 % and Sup 100 %. Furthermore, as some of the metrics are based on the h -top interesting rules of a given rule set, an objective measure should be selected. Instead of choosing a specific measure, the average rating obtained through 18 objective measures (see Table 3) was considered as follows: (i) the value of 18 measures was computed for each rule; (ii) each rule received 18 ID's, each ID corresponding to the rule position in one of the ranks related to a measure; (iii) the average was then calculated based on the rank positions (ID's). Thus, the h -top rules were selected considering the best average ratings. h , also a number to be set, was defined, in all the sets (RsT and RsP), to assume 10 % of the total of rules in RsT (always the smallest set). Therefore, each rule set contains its own values that are proportional in all of them.

To finish, as mentioned before, the labeling method applied over the clusters of each obtained clustering was the one presented by [22], named GLM (*Genetic Labeling Method*). In GLM the labels of each cluster are chosen by optimizing Precision (P) and Repetition Frequency (RF), the two measures previous described in Sect. 4. In fact, 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 $(P + RF)$ shows 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 . Table 3 summarizes the configurations used to apply the proposed metrics.

Table 3. Configurations used to apply the proposed metrics.

Data sets	Adult; Income; Groceries; Sup
Algorithms	PAM; Ward [algorithms details in [24]]
Similarity measures	Agg; Denza
k	5 to 25, steps of 5
h	10 % of the total of rules in RsT
Objective measures [measures details in [23]]	Added Value, Certainty Factor, Collective Strength, Confidence, Conviction, IS, ϕ -coefficient, Gini Index, J-Measure, Kappa, Kloggen, λ , Laplace, Lift, Mutual Information (asymmetric), Novelty, Support, Odds Ratio
Labeling Method	GLM [details in [22]]

6 Results and Discussion

Considering the configurations presented in Table 3 and the RsT sets above described, the experiments were carried out and the values of each metric obtained. Regarding the first proposed situation, i.e., identify among some clustering setups the most suitable (Sect. 5), an analysis based on the average of each metric was carried out. Table 4 presents the results for each data set. In this case, the metrics will help the users to find a suitable methodology for their problems. In order to aid the comparison of the results, all the metrics that present better results when their values are the smallest ($M_{R-O-RsP}$, for example) were processed to store the complement of the information. Therefore, all the metric, with exception to M_{NR-RsP} , have the same interpretation: the higher the value the better the performance. Furthermore, all the metrics can be seen in terms of percentage if multiplied by 100 (ex.: $0.807 \cdot 100 = 80.7\%$).

Each average in Table 4 was obtained from the results of the experiments related to the presented configuration. The value 0.807 in M_{O-RsP} at Adult:PAM:Agg, for example, was obtained by the average of the values in M_{O-RsP} at Adult:PAM:Agg over the values of k . The table also presents, between “[]”, the standard deviations; regarding the last example, the standard deviation is ± 0.02 ($0.807 [\pm 0.02]$). It can be observed that almost all the metrics present a low standard deviation, showing a good homogeneity among the results obtained over the values of k . The major exception is M_{NR-RsP} , which presents high standard deviation values, since the higher the number k of groups the higher the number of rules. For each data set, the highest averages are marked with \blacktriangle in each metric. The only exception is M_{NR-RsP} , where the lowest averages are highlighted. For Adult, for example, the best average for $M_{R-O-RsP}$ is the one related to PAM:Agg (0.807). Thereby, it is possible to visualize, for each data set, the most suitable clustering setup. It is important to mention that the results are deterministic and, therefore, no statistical test was done to check if there is a significant difference among the averages. It can be noticed that:

Adult. The most suitable configuration for this data set is **PAM:Agg**, since it presents better results in 8 of the 11 metrics. Furthermore, it can be noticed that in some cases the values in PAM:Agg are more representative than the others – observe, for example, that while in PAM:Agg M_{O-RsP} presents a performance above 80%, the others presents a performance below 60% (see also $M_{O-I-N-RsP}$ and M_{NR-RsP}).

Income. The most suitable configuration for this data set is **PAM:Denza**, since it presents better results in 4 of the 11 metrics. However, PAM:Agg can also be useful, since it presents better results in 3 of the 11 metrics and the two setups present close values in almost all the metrics. Thus, in this case, the user can choose one of them based on the importance each metric represents to him in the considered problem.

Groceries. The most suitable configuration for this data set is **Ward:Agg**, since it presents better results in 7 of the 11 metrics (in 5 excluding the ties). Furthermore, it can be noticed that in some cases the values in Ward:Agg are more

Table 4. Average of the proposed metrics, for each data set, in the considered clustering setups.

Adult							
Algorithm	Measure	M_{O-RsP}	$M_{R-O-RsP}$	M_{N-RsP}	$M_{R-N-RsP}$	$M_{N-I-RsP}$	$M_{O-I-N-RsP}$
PAM	Agg	0.807 [±0.02]▲	0.343 [±0.07]▲	0.834 [±0.02]	0.890 [±0.01]▲	0.588 [±0.02]	0.840 [±0.01]▲
	Denza	0.585 [±0.03]	0.314 [±0.06]	0.888 [±0.03]▲	0.863 [±0.03]	0.824 [±0.03]	0.333 [±0.05]
		M_{C-I}	$M_{NC-I-RsP}$	M_{NR-RsP}	M_P	M_{RF}	
	Agg	0.778 [±0.02]▲	0.551 [±0.08]▲	12.221 [±2.53]▲	0.699 [±0.03]▲	0.830 [±0.06]▲	
	Denza	0.886 [±0.03]	0.483 [±0.15]	20.460 [±9.63]	0.657 [±0.03]	0.789 [±0.12]	
		M_{O-RsP}	$M_{R-O-RsP}$	M_{N-RsP}	$M_{R-N-RsP}$	$M_{N-I-RsP}$	$M_{O-I-N-RsP}$
Ward	Agg	0.565 [±0.01]	0.253 [±0.07]	0.867 [±0.03]	0.854 [±0.02]	0.932 [±0.02]▲	0.150 [±0.01]
	Denza	0.565 [±0.04]	0.298 [±0.03]	0.878 [±0.03]	0.854 [±0.03]	0.870 [±0.04]	0.237 [±0.10]
		M_{C-I}	$M_{NC-I-RsP}$	M_{NR-RsP}	M_P	M_{RF}	
	Agg	0.993 [±0.01]▲	0.394 [±0.11]	17.587 [±7.99]	0.619 [±0.04]	0.805 [±0.11]	
	Denza	0.921 [±0.05]	0.385 [±0.22]	18.057 [±7.61]	0.618 [±0.04]	0.804 [±0.12]	
		M_{O-RsP}	$M_{R-O-RsP}$	M_{N-RsP}	$M_{R-N-RsP}$	$M_{N-I-RsP}$	$M_{O-I-N-RsP}$
Income							
Algorithm	Measure	M_{O-RsP}	$M_{R-O-RsP}$	M_{N-RsP}	$M_{R-N-RsP}$	$M_{N-I-RsP}$	$M_{O-I-N-RsP}$
PAM	Agg	0.909 [±0.03]▲	0.134 [±0.05]	0.979 [±0.01]	0.871 [±0.04]	0.844 [±0.10]	0.942 [±0.04]▲
	Denza	0.876 [±0.03]	0.156 [±0.10]	0.983 [±0.00]▲	0.894 [±0.03]	0.924 [±0.05]▲	0.833 [±0.08]
		M_{C-I}	$M_{NC-I-RsP}$	M_{NR-RsP}	M_P	M_{RF}	
	Agg	0.898 [±0.07]	0.640 [±0.15]	238.519 [±120.36]▲	0.654 [±0.06]	0.759 [±0.16]	
	Denza	0.938 [±0.03]	0.643 [±0.23]	300.608 [±169.34]	0.666 [±0.06]▲	0.786 [±0.13]▲	
		M_{O-RsP}	$M_{R-O-RsP}$	M_{N-RsP}	$M_{R-N-RsP}$	$M_{N-I-RsP}$	$M_{O-I-N-RsP}$
Ward	Agg	0.871 [±0.07]	0.173 [±0.09]▲	0.980 [±0.00]	0.888 [±0.04]	0.909 [±0.05]	0.782 [±0.10]
	Denza	0.879 [±0.06]	0.120 [±0.10]	0.978 [±0.01]	0.901 [±0.04]▲	0.869 [±0.08]	0.782 [±0.08]
		M_{C-I}	$M_{NC-I-RsP}$	M_{NR-RsP}	M_P	M_{RF}	
	Agg	0.942 [±0.05]▲	0.612 [±0.12]	241.775 [±134.24]	0.632 [±0.07]	0.773 [±0.13]	
	Denza	0.916 [±0.04]	0.681 [±0.15]▲	249.930 [±139.44]	0.644 [±0.08]	0.773 [±0.11]	
		M_{O-RsP}	$M_{R-O-RsP}$	M_{N-RsP}	$M_{R-N-RsP}$	$M_{N-I-RsP}$	$M_{O-I-N-RsP}$
Groceries							
Algorithm	Measure	M_{O-RsP}	$M_{R-O-RsP}$	M_{N-RsP}	$M_{R-N-RsP}$	$M_{N-I-RsP}$	$M_{O-I-N-RsP}$
PAM	Agg	0.981 [±0.03]	0.429 [±0.17]	0.807 [±0.05]▲	0.948 [±0.03]	0.500 [±0.11]	1.000 [±0.00]▲
	Denza	1.000 [±0.00]▲	0.100 [±0.07]	0.798 [±0.07]	0.880 [±0.02]	0.567 [±0.17]▲	1.000 [±0.00]▲
		M_{C-I}	$M_{NC-I-RsP}$	M_{NR-RsP}	M_P	M_{RF}	
	Agg	1.000 [±0.00]▲	0.909 [±0.06]▲	11.238 [±4.41]	0.613 [±0.10]	0.833 [±0.10]	
	Denza	0.933 [±0.08]	0.714 [±0.19]	18.178 [±8.58]	0.558 [±0.10]	0.800 [±0.09]	
		M_{O-RsP}	$M_{R-O-RsP}$	M_{N-RsP}	$M_{R-N-RsP}$	$M_{N-I-RsP}$	$M_{O-I-N-RsP}$
Ward	Agg	1.000 [±0.00]▲	0.894 [±0.13]▲	0.245 [±0.27]	0.997 [±0.01]▲	0.367 [±0.45]	1.000 [±0.00]▲
	Denza	1.000 [±0.00]▲	0.509 [±0.37]	0.538 [±0.27]	0.949 [±0.06]	0.567 [±0.36]▲	1.000 [±0.00]▲
		M_{C-I}	$M_{NC-I-RsP}$	M_{NR-RsP}	M_P	M_{RF}	
	Agg	0.367 [±0.45]	0.792 [±0.09]	1.866 [±1.05]▲	0.893 [±0.09]▲	0.966 [±0.04]▲	
	Denza	0.700 [±0.40]	0.850 [±0.03]	5.978 [±4.29]	0.758 [±0.15]	0.950 [±0.08]	
		M_{O-RsP}	$M_{R-O-RsP}$	M_{N-RsP}	$M_{R-N-RsP}$	$M_{N-I-RsP}$	$M_{O-I-N-RsP}$
Sup							
Algorithm	Measure	M_{O-RsP}	$M_{R-O-RsP}$	M_{N-RsP}	$M_{R-N-RsP}$	$M_{N-I-RsP}$	$M_{O-I-N-RsP}$
PAM	Agg	0.778 [±0.03]	0.990 [±0.01]	0.996 [±0.00]▲	0.999 [±0.00]	1.000 [±0.00]▲	1.000 [±0.00]▲
	Denza	0.849 [±0.09]	0.920 [±0.09]	0.971 [±0.05]	0.999 [±0.00]	0.924 [±0.15]	0.972 [±0.03]
		M_{C-I}	$M_{NC-I-RsP}$	M_{NR-RsP}	M_P	M_{RF}	
	Agg	1.000 [±0.00]▲	0.909 [±0.06]▲	528.422 [±735.64]	0.496 [±0.06]	0.971 [±0.03]	
	Denza	0.959 [±0.08]	0.855 [±0.13]	845.059 [±942.35]	0.538 [±0.09]	0.960 [±0.03]	
		M_{O-RsP}	$M_{R-O-RsP}$	M_{N-RsP}	$M_{R-N-RsP}$	$M_{N-I-RsP}$	$M_{O-I-N-RsP}$
Ward	Agg	0.952 [±0.05]▲	1.000 [±0.00]▲	0.245 [±0.25]	1.000 [±0.00]▲	0.055 [±0.07]	0.993 [±0.01]
	Denza	0.946 [±0.05]▲	1.000 [±0.00]▲	0.848 [±0.11]	0.999 [±0.00]	0.641 [±0.24]	0.993 [±0.01]
		M_{C-I}	$M_{NC-I-RsP}$	M_{NR-RsP}	M_P	M_{RF}	
	Agg	0.228 [±0.10]	0.864 [±0.04]	1.439 [±0.54]▲	0.961 [±0.02]▲	1.000 [±0.00]▲	
	Denza	0.669 [±0.20]	0.885 [±0.05]	437.856 [±863.97]	0.862 [±0.05]	0.978 [±0.03]	
		M_{O-RsP}	$M_{R-O-RsP}$	M_{N-RsP}	$M_{R-N-RsP}$	$M_{N-I-RsP}$	$M_{O-I-N-RsP}$

representative than the others – observe, for example, that while in Ward:Agg $M_{R-O-RsP}$ presents a performance above 89%, the others presents a performance below 50% (see also M_{NR-RsP} and M_P).

Sup. The most suitable configuration for this data set is **Ward:Agg**, since it presents better results in 6 of the 11 metrics (in 5 excluding the ties). Furthermore, it can be noticed that in some cases the values in Ward:Agg are more

representative than the others (see M_{NR-RsP} and M_P). However, PAM:Agg can also be useful, since it presents better results in 5 of the 11 metrics (also 5 excluding the ties). Thus, in this case, the user can choose one of them based on the importance each metric represents to him in the considered problem.

Therefore, considering the user selected and used the configurations presented in Table 3, it can be observed that the most suitable clustering setup according to the metrics is PAM:Agg for Adult, PAM:Denza for Income and Ward:Agg for Groceries and Sup. In other words, the user will obtain better results, i.e., reasonable rule sets, if he initially clusters Adult through PAM:Agg, Income through PAM:Denza and Groceries and Sup through Ward:Agg. However, in other situations, different aspects can be of interesting, providing to the user a flexible way to explore his problems: the user can choose the metrics to apply and their better interpretations (lower/higher values (as mentioned in some of the metrics (Sect. 4))). Thus, in this first situation, the metrics provide criteria to carry out comparisons, helping the user to select the most suitable methodology for his problem.

From that point, supposing that PAM:Agg is a suitable solution for the user's problem regarding Adult, it is possible to analyze the process itself, i.e., to check if good results are really obtained (the same is valid for the other data sets). Observe that different interests may be important for his decision. Thus, the metrics provide criteria not only to analyze the process, but also to identify its positive and negative aspects, helping the user to reach a conclusion. To discuss this second situation, Table 5 presents the values of the metrics, in the selected clustering setup, regarding Adult (although this second situation is only discussed on Adult, the same analysis can be done to the other data sets). These values are the ones presented in Table 4, but in their original scales, since the smaller scales (\downarrow) were previous converted – the larger scales (\uparrow) remain the same. The scale, in each metric, is found between “[]”. It can be noticed that:

M_{O-RsP} : little knowledge is lost during the process, around 20 %, since more than 80 % of the rules in RsT are found in RsP – being a positive aspect.

$M_{R-O-RsP}$: the repetition of “old” rules in RsP is high, around 66 %, indicating that the knowledge, already known, is not in subdomains that express its own associations – being a negative aspect.

M_{N-RsP} : most of the rules in RsP are “new”, around 83 %, indicating the discovery of a great amount of knowledge previously unknown – being a positive aspect.

$M_{R-N-RsP}$: the repetition of “new” rules in RsP is low, around 11 %, indicating that the knowledge, previously unknown, is in subdomains that express its own associations – being a positive aspect.

$M_{N-I-RsP}$: nearly half of the h -top interesting rules in RsP are “new”, around 59 %, indicating that the cost of the process is minimized by the discovery of interesting knowledge, previously unknown, in RsP – being a positive aspect.

$M_{O-I-N-RsP}$: the loss of “old” and interesting knowledge is low, around 16 %, since a great amount of the h -top interesting rules in RsT are found in RsP, around 84 % ($100\% - 16\%$) – being a positive aspect.

M_{C-I} : the intersection between the h -top interesting rules in RsP and the h -top interesting rules in RsT is high, around 78%, indicating that a great amount of the knowledge already known as interesting in RsT is found in RsP – being a positive aspect.

$M_{NC-I-RsP}$: the number of groups that contain the h -top interesting rules in RsP is high, around 45% – being a negative aspect, since many groups would have to be explored once the “new” relevant knowledge of the domain would be spread over the clusters.

M_{NR-RsP} : the number of rules in RsP is only around 12 times greater in relation to RsT – being a positive aspect (it can be seen in Table 4 that this ratio can be very high).

M_P : the selected clustering setup provides a suitable distribution of the domain topics, since the labels represent the rules in the clusters at a ratio around 70% – being a positive aspect.

M_{RE} : the selected clustering setup provides a suitable distribution of the domain topics, since the distinct labels don't repeat over the clusters at a ratio around 83% – being a positive aspect.

Table 5. Average of the proposed metrics in Adult, PAM:Agg clustering setup.

Data set	M_{O-RsP} [↑]	$M_{R-O-RsP}$ [↓]	M_{N-RsP} [↑]	$M_{R-N-RsP}$ [↓]	$M_{N-I-RsP}$ [↑]	$M_{O-I-N-RsP}$ [↓]
Adult	0.807	0.657	0.834	0.110	0.588	0.160
PAM:Agg	M_{C-I} [↑]	$M_{NC-I-RsP}$ [↓]	M_{NR-RsP} [↓]	M_P [↑]	M_{RF} [↑]	
	0.778	0.449	12.221	0.699	0.830	

As seen, considering the positive and negative aspects of the process, the user can analyze the results, according to his interests, and conclude if good results were reached. It is relevant to mention that the importance of each percentage depends on the user's needs, on the data sets, etc., and, therefore, the metrics provide a flexible way for him to explore the problems. Regarding the presented context, it can be said that the process obtains reasonable results, since 9 of the 11 aspects were considered positives. However, if the weight of the negative aspects is more important to the user, he can discard the results. Moreover, he can try to improve the process to obtain better results in these metrics, since he has an overview of all the aspects. Thus, in this second situation, the metrics provide criteria to analyze the process based on different interests, identifying its positive and negative aspects, helping the user to reach a conclusion.

Finally, to complement and finalize the discussion, Fig. 11 presents the variation of h parameter in the four metrics that depend on that value: $M_{N-I-RsP}$, $M_{O-I-N-RsP}$, M_{C-I} , $M_{NC-I-RsP}$. The clustering setup related to each data set is the one above identified as the most suitable. Axis x is related to h and axis y to the metrics' values. The metrics are represented by the different lines in the graphics. Note that the metrics' values related to $h=10\%$ are the same as the ones presented in Table 4 (as before, each metric's value was obtained by the average of the values in the metric over the values of k). It can be seen that:

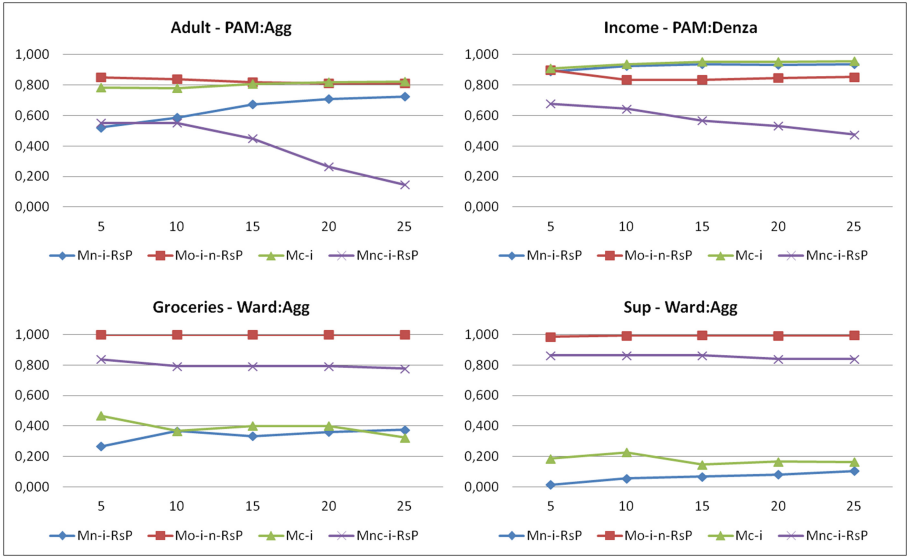


Fig. 11. Behavior of the h parameter in $M_{N-I-RsP}$, $M_{O-I-N-RsP}$, M_{C-I} , $M_{NC-I-RsP}$ in each data set (Color figure online).

(a) $M_{N-I-RsP}$ [\uparrow] (blue line) tends to become higher as h increases or tends to assume a value close to the other h values; therefore, this metric presents better results with high values of h ; (b) $M_{O-I-N-RsP}$ [\downarrow] (red line) tends to decrease as h increases or tends to assume a value close to the other h values; therefore, this metric presents better results with high values of h ; (c) M_{C-I} [\uparrow] (green line) does not present a pattern; however, while in the relational data sets the values are close and high regardless the h value in the transactional ones the values are close and low regardless the h value; (d) $M_{NC-I-RsP}$ [\downarrow] (purple line) tends to decrease as h increases or tends to assume a value close to the other h values; therefore, this metric presents better results with high values of h .

7 Conclusion

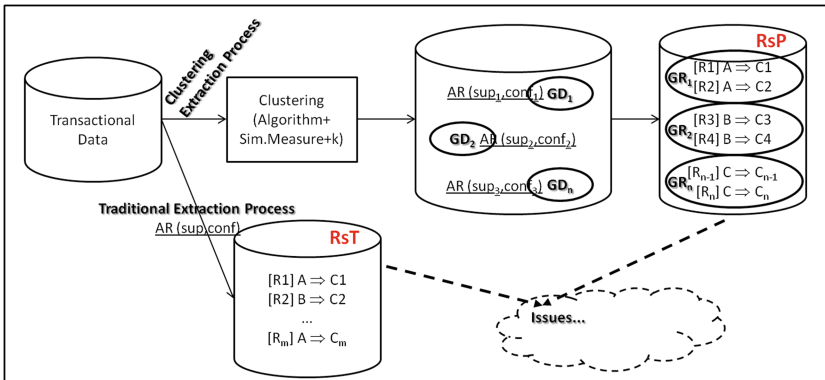
In this paper, eleven metrics were proposed to provide an assessment procedure in order to support the evaluation of methodologies that use clustering in the pre-processing step. The metrics were developed to answer three main issues. However, to propose the metrics, a subjective evaluation was done with some users to understand each issue. Some experiments were carried out in order to present how the metrics can be used. For that, two different situations were regarded: (i) identify among some clustering setups the most suitable; (ii) analyze the process itself. Through the experiments, it could be noticed that the metrics provide criteria to: (a) analyze the methodologies, (b) identify their positive and negative aspects, (c) carry out comparisons among them and, therefore,

(d) help the users to select the most suitable solution for their problems. Besides, the metrics do the users think about aspects to be considered in the presented context and provide to them a flexible way to explore the problems. Finally, this paper complements [12]’s work, since the subjective evaluation is used to attest the importance of the previous obtained results.

Acknowledgments. We wish to thank Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (processes numbers: 2010/07879-0 and 2011/19850-9) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (process number DS-6345378/D) for the financial support. Besides, we also want to thank the reviewers for the great contributions.

Appendix: Questionnaire

Introduction. Many issues related to association rule mining have received attention in the last years, especially the ones aiming to discover and facilitate the search for the interesting patterns of the domain. One approach related to this issue is the application of clustering in the pre-process step. In this case, as noticed in the figure below, data are initially grouped in n groups (GD_1, GD_2, \dots, GD_n). From this initial clustering, the rules are then extracted within each group (cluster), obtaining n groups of rules (GR_1, GR_2, \dots, GR_n). *The aim is to obtain potentially interesting rules that would not be extracted from unpartitioned data sets, for not having enough support, without overloading the user with a great amount of patterns.* The user must set the minimum support to a low value to discover these same patterns from unpartitioned data sets, causing a rapidly increase in the number of rules. Thereby, data are initially split and the rules are extracted within each group, in a manner that *each group expresses its own associations without the interference of the other groups that contain different association patterns.* Distinct methodologies have been proposed to enable this process. Each methodology uses a different combination of clustering algorithms and similarity measures in order to obtain the groups of rules.



It is in this context that this evaluation should be done. Some scenarios that can occur in this scope are shown below, waiting for your contribution for a better understanding of the problem. In all the cases, it is assumed that two rule sets are available, in order to evaluate the presented scenarios: one extracted through traditional process, RsT⁷, and one extracted through clustering (process above described), RsP⁸ – the examples presented below are merely illustrations of the scenarios and, therefore, should not be evaluated considering the knowledge they express. Based on this evaluation, the aim is to propose an assessment procedure to support the analysis of the existing methodologies.

Scenarios

1. In your opinion, observing “Scenario-A” (Table 6), how do you consider the occurrence of rules obtained in RsT in RsP (cases in green and orange)? Both the cases, green and orange, represent rules obtained in both of the sets, but the rules in orange are extracted more than once in RsP over the groups. If needed to distinguish the green cases of the orange cases, please let it indicated.
 desirable indifferent no desirable
 - a. Do you think important to consider this scenario in an assessment procedure to be used in the presented context?
 yes no
 - b. Would you like to make any comment about the scenario (advantage, disadvantage, etc.)?
2. In your opinion, observing “Scenario-A” (Table 6), how do you consider the non-occurrence of rules obtained in RsP in RsT (cases in purple and red)? Both the cases, purple and red, represent rules obtained only in RsP, but the rules in red are extracted more than once in RsP over the groups. If needed to distinguish the purple cases of the red cases, please let it indicated.
 desirable indifferent no desirable
 - a. Do you think important to consider this scenario in an assessment procedure to be used in the presented context?
 yes no
 - b. Would you like to make any comment about the scenario (advantage, disadvantage, etc.)?

For questions “3” to “6”, consider that for each rule set, RsP and RsT, it is shown only the subset related to the n most interesting rules of the domain. These subsets can be identified, for example, automatically, based on a set of objective measures – assuming that objective measures are suitable to find the most interesting knowledge of a given domain.

3. In your opinion, observing “Scenario-B” (Table 7), how do you consider the non-occurrence of some (or none) of the n most interesting rules in RsP in RsT (cases in blue)? Notice that the blue rules belong only to the RsP set.
 desirable indifferent no desirable

⁷ Rule set obtained through a *t*raditional process.

⁸ Rule set obtained through a *p*artitioned data.

Table 6. Scenario-A. This scenario was formulated based on the Sup data set described in Sect. 5. In this scenario the rules in RsP are presented in their own clusters since the aim here is to detach to the user the situations that can occur among the groups (repetitions of rules) and between RsT and RsP (occurrence/non-occurrence of rules between the sets).

RsT	RsP
CALDO_MAGGI & FERMROYAL ⇒ CREME_DE_LEITE_NESTLE PAPEL_ALUMROLITTO & GELATINA_ROYAL ⇒ FARTRIGO_RENATA	Cluster _n CALDO_MAGGI & FERMROYAL ⇒ CREME_DE_LEITE_NESTLE PAPEL_ALUMROLITTO & GELATINA_ROYAL ⇒ FARTRIGO_RENATA MILHO_VERDE_QUERO & FEJAO_TORRESAN ⇒ ACHOCNESCAU FERMROYAL & DETERGLIMPOL ⇒ LEITE_MOCA Cluster _m PAPEL_ALUMROLITTO & GELATINA_ROYAL ⇒ FARTRIGO_RENATA FERMROYAL & DETERGLIMPOL ⇒ LEITE_MOCA

- a. Do you think important to consider this scenario in an assessment procedure to be used in the presented context?
 yes no
- b. Would you like to make any comment about the scenario (advantage, disadvantage, etc.)?
4. In your opinion, observing “Scenario-B” (Table 7), how do you consider the reverse scenario? This is, the non-occurrence of some (or none) of the n most interesting rules in RsT in RsP (cases in orange)? Notice that the orange rules belong only to the RsT set.
 desirable indifferent no desirable
- a. Do you think important to consider this scenario in an assessment procedure to be used in the presented context?
 yes no
- b. Would you like to make any comment about the scenario (advantage, disadvantage, etc.)?
5. In your opinion, observing “Scenario-B” (Table 7), how do you consider the existing intersection between the n most interesting rules in RsP and the n most interesting rules in RsT (cases in red)?
 desirable indifferent no desirable
- a. Do you think important to consider this scenario in an assessment procedure to be used in the presented context?
 yes no
- b. Would you like to make any comment about the scenario (advantage, disadvantage, etc.)?
6. In your opinion, how do you would consider the spread of the n most interesting rules in RsP in a small number of clusters?
 desirable indifferent no desirable
- a. Do you think important to consider this scenario in an assessment procedure to be used in the presented context?
 yes no
- b. Would you like to make any comment about the scenario (advantage, disadvantage, etc.)?

Table 7. Scenario-B. This scenario was formulated based on the Sup data set described in Sect. 5. In this scenario the rules in RsP are presented all together since, in this case, only the n most interesting rules in the set are exhibited to the user, independently of the group they were extracted – the aim here is to detach to the user the situations that can occur between the subsets containing the n most interesting rules.

RsT <i>n</i> most interesting rules in <i>RsT</i>	RsP <i>n</i> most interesting rules in <i>RsP</i>
SPO,MINERVA & BISCNESTLE & LEITE,MOCA ⇒ ACHOCSNECAU SHELSEVE & COCA,COLA ⇒ LEITE,MOCA LUSTRA,MOVOPOLIFLOR & CREME,DE,LEITE,NESTLE ⇒ LEITE,MOCA CALDO,MAGGI & FERROYAL & LEITE,MOCA ⇒ CREME,DE,LEITE,NESTLE NISSIM,L,AMIEM & PAPEL,HIGPERSONAL & BISCNESTLE ⇒ ACHOCSNECAU SCLIXO,PLASLIXO & GELATINA,ROYAL ⇒ ARROZ,PRATO,FINO LUSTRA,MOVOPOLIFLOR & ACUCAR,UNIAO ⇒ LEITE,MOCA OLEO,SOJA,SOYA & BISCNESTLE & COCA,COLA ⇒ LEITE,MOCA FELIAO,TORRESAN & PAPEL,HIGPERSONAL & BISCNESTLE ⇒ LEITE,MOCA LUSTRA,MOVOPOLIFLOR & VEJA,MUSO ⇒ LEITE,MOCA VEJA,MUSO & FERROYAL ⇒ LEITE,MOCA OLEO,SOJA,SOYA & FERROYAL & PAPEL,HIGPERSONAL ⇒ ACUCAR,DA,BARRA DESINFINHO & CREME,DE,LEITE,NESTLE ⇒ LEITE,MOCA PAPEL,ALUMROLITTO & GELATINA,ROYAL ⇒ FARTRIGO,RENATA CALDO,KNORR & PAPEL,HIGPERSONAL ⇒ LEITE,MOCA FELIAO,TORRESAN & OLEO,SOJA,SOYA & ACHOCSNECAU ⇒ LEITE,MOCA CARGA,GILLETTE & BOMBRIIL ⇒ LEITE,MOCA SABPROTEX & AGUA,SANTICANDURA ⇒ PAPEL,HIGPERSONAL OLEO,SOJA,SOYA & PAPEL,HIGPERSONAL & BISCNESTLE ⇒ LEITE,MOCA PAPEL,ALUMROLITTO & GELATINA,ROYAL & BISCNESTLE ⇒ FARTRIGO,RENATA OLEO,SOJA,SOYA & FELIAO,BROTO,LEGAL ⇒ LEITE,MOCA NESCAFÉ,TRADICAO & COCA,COLA ⇒ LEITE,MOCA AGUA,SANTICANDURA & GELATINA,ROYAL ⇒ FARTRIGO,RENATA AMACOMFORT & FARTRIGO,RENATA & CREME,DE,LEITE,NESTLE ⇒ LEITE,MOCA REQCATUPIRY & CREME,DE,LEITE,NESTLE ⇒ LEITE,MOCA ESPONJA,BOMBRIIL & ACHOCSNECAU ⇒ LEITE,MOCA LUSTRA,MOVOPOLIFLOR & BISCNESTLE ⇒ LEITE,MOCA ESPONJA,BOMBRIIL & LETERGILMPO & LEITE,MOCA ⇒ ACHOCSNECAU CALDO,MAGGI & FERROYAL ⇒ CREME,DE,LEITE,NESTLE	ACUCAR,CRISTAL,DA,BARRA & OLEO,SOJA,SOYA & SAL,CISNE ⇒ FELIAO,TORRESAN FANTA & ACUCAR,DA,BARRA & LEITE,MOCA ⇒ OLEO,SOJA,SOYA MILHO,VERDE,QUERO & FELIAO,TORRESAN ⇒ ACHOCSNECAU SHELSEVE & COCA,COLA ⇒ LEITE,MOCA LUSTRA,MOVOPOLIFLOR & CREME,DE,LEITE,NESTLE ⇒ LEITE,MOCA CALDO,MAGGI & FERROYAL & LEITE,MOCA ⇒ CREME,DE,LEITE,NESTLE NISSIM,L,AMIEM & PAPEL,HIGPERSONAL & BISCNESTLE ⇒ ACHOCSNECAU PAPEL,ALUMROLITTO & GELATINA,ROYAL & FARTRIGO,RENATA ⇒ BISCNESTLE SCLIXO,PLASLIXO & GELATINA,ROYAL ⇒ ARROZ,PRATO,FINO PAPEL,ALUMROLITTO & GELATINA,ROYAL ⇒ BISCNESTLE PAPEL,ALUMROLITTO & GELATINA,ROYAL & CREME,DE,LEITE,NESTLE ⇒ FARTRIGO,RENATA OLEO,SOJA,SOYA & BISCNESTLE & COCA,COLA ⇒ LEITE,MOCA SCLIXO,PLASLIXO & GELATINA,ROYAL & LEITE,MOCA ⇒ ARROZ,PRATO,FINO FELIAO,TORRESAN & PAPEL,HIGPERSONAL & BISCNESTLE ⇒ LEITE,MOCA OLEO,SOJA,SOYA & FERROYAL & PAPEL,HIGPERSONAL ⇒ ACUCAR,DA,BARRA DESINFINHO & CREME,DE,LEITE,NESTLE ⇒ LEITE,MOCA ESPONJA,BOMBRIIL & ACHOCSNECAU & BISCNESTLE ⇒ LEITE,MOCA PAPEL,ALUMROLITTO & GELATINA,ROYAL ⇒ FARTRIGO,RENATA FELIAO,TORRESAN & OLEO,SOJA,SOYA & ACHOCSNECAU ⇒ LEITE,MOCA CALDO,MAGGI & FERROYAL & BISCNESTLE ⇒ CREME,DE,LEITE,NESTLE SABPROTEX & AGUA,SANTICANDURA ⇒ PAPEL,HIGPERSONAL OLEO,SOJA,SOYA & PAPEL,HIGPERSONAL & BISCNESTLE ⇒ LEITE,MOCA PAPEL,ALUMROLITTO & GELATINA,ROYAL & BISCNESTLE ⇒ FARTRIGO,RENATA OLEO,SOJA,SOYA & FELIAO,BROTO,LEGAL ⇒ LEITE,MOCA REQCATUPIRY & CREME,DE,LEITE,NESTLE ⇒ LEITE,MOCA ESPONJA,BOMBRIIL & FERROYAL & LEITE,MOCA ⇒ CREME,DE,LEITE,NESTLE AGUA,SANTICANDURA & GELATINA,ROYAL ⇒ FARTRIGO,RENATA LUSTRA,MOVOPOLIFLOR & BISCNESTLE ⇒ LEITE,MOCA REQCATUPIRY & CREME,DE,LEITE,NESTLE ⇒ LEITE,MOCA ESPONJA,BOMBRIIL & ACHOCSNECAU ⇒ LEITE,MOCA CD,CLOSEUP & ACHOCSNECAU & LEITE,MOCA ⇒ BISCNESTLE PAPEL,ALUMROLITTO & GELATINA,ROYAL & LEITE,MOCA ⇒ FARTRIGO,RENATA CALDO,MAGGI & FERROYAL ⇒ CREME,DE,LEITE,NESTLE

7. In your opinion, do you consider that the amount of rules to be extracted through clustering, compared to the traditional process, should be:
 - () low () average () high
 - a. Do you think important to consider this scenario in an assessment procedure to be used in the presented context?
 - () yes () no
 - b. Would you like to make any comment about the scenario (advantage, disadvantage, etc.)?
8. In your opinion, only in relation to RsP, do you consider that the clustering process should, as a consequence, enable each cluster to express a distinct topic of the domain?
 - () yes () indifferent () no
 - a. Do you think important to consider this scenario in an assessment procedure to be used in the presented context?
 - () yes () no
 - b. Would you like to make any comment about the scenario (advantage, disadvantage, etc.)?
9. Can you identify other scenario(s), not previously explored, that can be relevant to the presented context? Give an example of the scenario(s) that you identified.
 - a. Do you think important to consider this(these) scenario(s) in an assessment procedure to be used in the presented context?
 - () yes () no
10. If you want to leave any comment/observation, please do it below.

References

1. Wu, X., Kumar, V.: *The Top Ten Algorithms in Data Mining*. Chapman & Hall/CRC, Boca Raton (2009)
2. Dadaser-Celik, F., Celik, M., Dokuz, A.S.: Associations between stream flow and climatic variables at Kizilirmak river basin in Turkey. *Glob. NEST J.* **14**(3), 354–361 (2012)
3. Xiao, G.: Association rules algorithm in bank risk assessment. In: Lee, J. (ed.) *Advanced Electrical and Electronics Engineering*. LNEE, vol. 87, pp. 675–681. Springer, Heidelberg (2011)
4. Nuwangi, S.M., Oruthotaarachchi, C.R., Tilakaratna, J.M.P.P., Caldera, H.A.: Usage of association rules and classification techniques in knowledge extraction of diabetes. In: *Proceedings of the 6th International Conference on Advanced Information Management and Service*, pp. 372–377 (2010)
5. Rajasekar, U., Weng, Q.: Application of association rule mining for exploring the relationship between urban land surface temperature and biophysical/social parameters. *Photogram. Eng. Remote Sens.* **75**(3), 385–396 (2009)
6. Changguo, Y., Nianzhong, W., Tailei, W., Qin, Z., Xiaorong, Z.: The research on the application of association rules mining algorithm in network intrusion detection. In: *Proceedings of the 1st International Workshop on Education Technology and Computer Science*, vol. 2, pp. 849–852 (2009)
7. Koh, Y.S., Pears, R.: Rare association rule mining via transaction clustering. In: *7th Australasian Data Mining Conference*. CRPIT, vol. 87, pp. 87–94 (2008)
8. Maquee, A., Shojaie, A.A., Mosaddar, D.: Clustering and association rules in analyzing the efficiency of maintenance system of an urban bus network. *Int. J. Syst. Assur. Eng. Manage.* **3**(3), 175–183 (2012)
9. Faraajian, M.A., Mohammadi, S.: Mining the banking customer behavior using clustering and association rules methods. *Int. J. Ind. Eng. Prod. Res.* **21**(4), 239–245 (2010)
10. Fan, L.: Research on classification mining method of frequent itemset. *J. Convergence Inf. Technol.* **5**(8), 71–77 (2010)
11. Plasse, M., Niang, N., Saporta, G., Villeminot, A., Leblond, L.: Combined use of association rules mining and clustering methods to find relevant links between binary rare attributes in a large data set. *Comput. Stat. Data Anal.* **52**(1), 596–613 (2007)
12. de Carvalho, V.O., dos Santos, F.F., Rezende, S.O.: Metrics to support the evaluation of association rule clustering. In: Bellatreche, L., Mohania, M.K. (eds.) *DaWaK 2013*. LNCS, vol. 8057, pp. 248–259. Springer, Heidelberg (2013)
13. Aggarwal, C.C., Procopiuc, C., Yu, P.S.: Finding localized associations in market basket data. *IEEE Trans. Knowl. Data Eng.* **14**(1), 51–62 (2002)
14. Wang, K., Xu, C., Liu, B.: Clustering transactions using large items. In: *8th International Conference on Information and Knowledge Management*, pp. 483–490 (1999)
15. Yun, C.-H., Chuang, K.-T., Chen, M.-S.: An efficient clustering algorithm for market basket data based on small large ratios. In: *25th International Computer Software and Applications Conference on Invigorating Software Development*, pp. 505–510 (2001)
16. Wang, J., Karypis, G.: Summary: efficiently summarizing transactions for clustering. In: *4th IEEE International Conference on Data Mining*, pp. 241–248 (2004)

17. Yang, L.: Pruning and visualizing generalized association rules in parallel coordinates. *IEEE Trans. Knowl. Data Eng.* **17**(1), 60–70 (2005)
18. D’Enza, A.I., Palumbo, F., Greenacre, M.: Exploratory data analysis leading towards the most interesting binary association rules. In: 11th Symposium on Applied Stochastic Models and Data Analysis, pp. 256–265 (2005)
19. Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J.M., Perona, I.: An extensive comparative study of cluster validity indices. *Pattern Recogn.* **46**(1), 243–256 (2013)
20. Halkidi, M., Batistakis, Y., Vazirgiannis, M.: On clustering validation techniques. *J. Intell. Inf. Syst.* **17**(2/3), 107–145 (2001)
21. 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. 105–109 (2012)
22. Padua, R., Carvalho, V.O., Serapião, A.B.S.: Labeling association rule clustering through a genetic algorithm approach. In: Proceedings of the 17th East European Conference on Advances in Databases and Information Systems, pp. 45–52 (2013)
23. Tan, P.-N., Kumar, V., Srivastava, J.: Selecting the right objective measure for association analysis. *Inf. Syst.* **29**(4), 293–313 (2004)
24. Xu, R., Wunsch, D.: Clustering. Computational Intelligence. IEEE Press/Wiley, New York (2008)
25. Carvalho, V.O., Santos, F.F., Rezende, S.O., Padua, R.: PAR-COM: a new methodology for post-processing association rules. *Lect. Notes Bus. Inf. Process.* **102**, 66–80 (2012)
26. Carvalho, V.O., Santos, F.F., Rezende, S.O.: Post-processing association rules with clustering and objective measures. In: Proceedings of 13th International Conference on Enterprise Information Systems, vol. 1, pp. 54–63 (2011)