A Post-Processing Strategy for Association Rules in Knowledge Discovery

Luiz Fernando da Cunha Cintra[1][a], Rodigo da Silva Dias[2] and Rogerio Salvini[1][b]

[1] *Instituto de Informática, Universidade Federal de Goiás, Goiânia-GO, Brazil*

[2] *Department of Psychiatry, University of Sao Paulo Medical School, São Paulo-SP, Brazil*

Keywords: Association Rule Mining, Post-Processing, Knowledge Discovery.

Abstract: Association Rule Mining (ARM) is a traditional data mining method that describes associations among elements in transactional databases. A well-known problem of ARM is the large number of rules generated, requiring approaches to post-process these rules so that a human expert can analyze the associations found. In certain scenarios, experts focus on exploring a specific element within the data, and a search based on this item can help reduce the problem. Few methods concentrate on post-processing generated rules targeting a specific item of interest. This study aims to highlight relevant associations of a particular element in order to gain knowledge about its role through its interactions and relationships with other factors. The paper introduces a post-processing strategy for association rules, selecting and grouping rules pertinent to a specific item of interest as provided by a domain expert. Additionally, a graphical representation facilitates the visualization and interpretation of associations between rules and their groupings. A case study demonstrates the applicability of the proposed method, effectively reducing the number of relevant rules to a manageable level for expert analysis.

1 INTRODUCTION

Association Rule Mining (ARM) (Agrawal et al., 1993) is a well-known method for extracting patterns from a dataset. Initially proposed to discover associations in supermarket basket data, over the last decades, it has been applied to various other domains, such as construction (Cheng et al., 2016), product development (Karimi-Majd and Mahootchi, 2015), education (Matetic et al., 2015), sports (Weidner et al., 2020), building maintenance (Zhang et al., 2021), medicine (Castro et al., 2018) (Wei and Scott, 2015), and urban planning (Balasubramani et al., 2016). One of the main challenges related to using association rules is the massive number of rules that ARM algorithms can generate. This issue is well-known and has been studied for over 20 years (Baesens et al., 2000).

ARM aims to identify frequent and meaningful associations within a transactional database. In the context of ARM, a transactional database stores data resulting from interactions between two or more parties, with each interaction referred to as a transaction. Typically, each transaction includes an identity number and a list of the items making up the transaction, such as a customer's purchase, a flight booking, or a user's clicks on a web page (Han et al., 2012).

An association rule is an "if-then" type of rule, formalized by Agrawal et al. (Agrawal et al., 1993) that reveals patterns or relationships within a set of transactions. A more general formalization states that an association rule has the form $A \rightarrow B$, where A and B are itemsets, i.e., $A = \{a_1, a_2, ..., a_n\}$ and $B = \{b_1, b_2, ..., b_m\}$, with a_i and b_j being items from a database. A is referred to as the antecedent, and B as the consequent of the rule. Given that I is the set of all items in the database, $A \subset I$, $B \subset I$, and $A \cap B = \emptyset$, meaning that antecedent and consequent are items from the database and do not have repeated items. The size of a rule is the number of items it contains.

Let D be a set of transactions, where each transaction T is an itemset such that $T \subset I$. The *support* of a rule $A \rightarrow B$ is the percentage of transactions in D containing the items of $A \cup B$, i.e., the percentage of transactions where the items from the antecedent and consequent of the rule occur together. The *confidence* of a rule $A \rightarrow B$ is the percentage of transactions in D that, if containing A, also contain B (Agrawal and Srikant, 1994).

Essentially, the ARM problem involves generating, from a set of transactions D, all rules with sup-

[a] https://orcid.org/0009-0002-6709-8789

[b] https://orcid.org/0000-0001-8889-6654

Cintra, L., Dias, R. and Salvini, R.
A Post-Processing Strategy for Association Rules in Knowledge Discovery.
DOI: 10.5220/0012465800003654

In *Proceedings of the 13th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2024)*, pages 119-130
ISBN: 978-989-758-684-2; ISSN: 2184-4313

port and confidence greater than the minimum values specified by the user. Consequently, this problem breaks down into two subproblems: 1) identifying all itemsets with support greater than or equal to the minimum support (referred to as frequent itemsets), and 2) generating rules from the frequent itemsets that have confidence greater than or equal to the specified minimum confidence. Numerous ARM algorithms adopt the support/confidence model to generate association rules. Examples of such algorithms include *Apriori* (Agrawal and Srikant, 1994), *Eclat* (Zaki, 2000), and *FP-Growth* (Han et al., 2004).

ARM, in general, is an exploratory activity. Unlike a classification task, where a target variable of the dataset is used to guide the construction of classification rules, ARM algorithms seek any statistically relevant pattern within the data (Freitas, 2000). This allows the analysis of specific items both in the consequent and the antecedent of a rule. An item of interest positioned in the antecedent enables the extraction of the consequences of its presence, not just what leads to its occurrence, as is the case in classification rules. For example, in the work of Wei and Scott (Wei and Scott, 2015), the item of interest is a vaccine, and the consequent is adverse events. Another aspect that can be explored in a generated set of association rules is the choice of rules that do not include an item of interest but share items with other rules that do, thus highlighting these relationships. Therefore, a method that considers these characteristics is particularly beneficial for experts looking for specific insights into their datasets, contributing to discovering knowledge in particular domains.

The main objective of this work is to propose a post-processing method for association rules that select the rules that, directly or indirectly, are related to a specific object of study for a domain expert based on an item of interest in the database. This method was motivated by a clinical study presented in the work of Castro and colleagues (Castro et al., 2018), in which ARM was used to uncover correlations between events related to the menstrual cycle. Their efforts aimed to construct a network of relationships from groupings of rules that could reveal specific clinical conditions influenced by the menstrual cycle. Essentially, the study addressed the question of whether a particular event forms a continuum related to women's hormonal fluctuations or whether they are isolated epiphenomena. Thus, the present work aims to extract rules that show how the occurrence of an item of interest influences other items and how different items are influenced by the item of interest, highlighting only those most relevant to the expert. We also introduce a graphical representation in the form of a graph to visualize the multiple relationships between the associations around the item of interest.

We extensively reviewed recent literature and did not identify any method specifically designed to handle the specified scenario. While we found four methods that focus on an item of interest, none involve the appearance of this item on both sides of the rule. Moreover, no existing method explicitly seeks to establish connections between rules, as proposed in this work.

The remainder of this paper is structured as follows. Section 2 presents the proposed association rule post-processing method. In Section 3, we present the results obtained when applying the proposed method to real data on deaths due to COVID-19 in Brazil. Section 4 makes a comparison between our method and related methods in the literature. The conclusions of this work are in Section 5.

2 PROPOSED METHOD

The proposed method is tailored to address issues where a specific feature is a focal point for investigation. It strategically selects rule subsets from an association rule set generated by an ARM algorithm, guided by a particular item of the database (referred to as *item of interest*) and a measure M reflecting the importance of the rules. Moreover, the method exclusively considers rules with a maximum size of three, recognizing that more extensive rules pose challenges in interpretation and practical application. The minimum rule size is set at two, aligning with the necessity for a rule antecedent in confidence calculations. Additionally, the method excludes rules containing items with missing values, ensuring a robust and reliable analysis.

The method organizes the selected rules into eight distinct types based on their relational structures. Type 1 addresses bidirectional rules, emphasizing the mutual influence between the item of interest and another element. Types 2, 3, 4, and 5 involve the inclusion of the item of interest in the rule's antecedent, spotlighting its role in strengthening the factors leading to another entity. Conversely, Types 6, 7, and 8 feature the item of interest in the consequent of the rule, revealing how associations with other factors may fortify or amplify the likelihood of the item of interest. Organizing rules into these specific groupings facilitates a deeper understanding of associations related to the research subject. Moreover, the proposed method filters out rules not directly pertinent to the research interest, reducing the number of rules for analysis.

We introduce the following eight types of groupings, illustrating them with association rules related to dengue fever[1]. These rules are not derived from a database but are based on information provided by a medical professional, serving as illustrative examples of the grouping types. In the subsequent section, we will present results obtained from actual databases. The focal item of interest is analyzing the presence of dengue fever ($dengue = yes$) in the associations. The symbol M indicates a value of a specific measure quantifying the degree of dependence between the antecedent and the consequent of a rule, guiding the grouping process.

Type 1 Group

Type 1 encompasses sets of bidirectional rules, meaning rules where some item implies the item of interest and the item of interest implies that item. This type of information reinforces associations, as it indicates a close connection between the entities. The example below shows that fever for more than seven days and dengue are strongly associated.

$$fever = [>7days] \rightarrow \boldsymbol{dengue = yes}$$
$$\boldsymbol{dengue = yes} \rightarrow fever = [>7days]$$

Type 2 Group

This type of grouping describes sets of rules suggesting a strengthened association between the item of interest and another specific item linked with the same consequent. This strengthening is observed when comparing the values of the particular metric to the individual associations of these items with the same consequent. In the example below, hemophilia and dengue individually cause bleeding. However, when both conditions (hemophilia and dengue) coexist in an individual, the association with bleeding is intensified. This is quantified through the metric M, where M_3 would surpass both M_1 and M_2.

$$hemophilia = yes \rightarrow bleeding = yes,\ (M_1)$$
$$\boldsymbol{dengue = yes} \rightarrow bleeding = yes,\ (M_2)$$
$$hemophilia = yes, \boldsymbol{dengue = yes} \rightarrow bleeding = yes,\ (M_3)$$

Type 3 Group

Type 3 indicates when the item of interest strengthens an existing association. Unlike Type 2, there is no prior rule where the item of interest is already related to the consequent. The example below shows that severe abdominal pain would lead an individual to be admitted to the Intensive Care Unit (ICU). The occurrence of dengue alongside intense abdominal pain reinforces the need for treatment in the ICU ($M_2 > M_1$). It is important to note that there is no rule explicitly stating the association between dengue and ICU admission.

$$abdominalPain = intense \rightarrow ICU = yes,\ (M_1)$$
$$abdominalPain = intense, \boldsymbol{dengue = yes} \rightarrow ICU = yes,\ (M_2)$$

Type 4 Group

Type 4 is similar to Type 3, but this time, it emphasizes when another item reinforces an existing association between the item of interest and the item in the consequent. In this scenario, there is no rule establishing a direct connection between the other element and the consequent. In the example below, we observe that dengue is associated with hospitalization. However, the presence of pregnancy along with dengue intensifies the likelihood of hospitalization ($M_2 > M_1$), even without a rule explicitly stating that pregnancy alone leads to hospitalization.

$$\boldsymbol{dengue = yes} \rightarrow hospitalization = yes,\ (M_1)$$
$$pregnant = yes, \boldsymbol{dengue = yes} \rightarrow hospitalization = yes,\ (M_2)$$

Type 5 Group

Type 5 consists of only one rule of size 3, where the item of interest is associated with an item in the antecedent and another in the consequent, without individual associations between the antecedent items and the consequent item. In the example below, dengue and intense abdominal pain are associated with respiratory distress. However, there are no individual associations between dengue and respiratory distress nor between intense abdominal pain and respiratory distress.

$$\boldsymbol{dengue = yes}, abdominalPain = intense \rightarrow respiratoryDistress = yes$$

Type 6 Group

Type 6 deals with groups of rules where the item of interest appears as the consequent of these rules. In this grouping, three rules aim to demonstrate that the conjunction of other items leading to the item of interest strengthens the association. In the example, we observe that fever for more than seven days and thrombocytopenia can be symptoms of dengue individu-

[1] Dengue fever (CID A90) is a mosquito-borne tropical disease caused by the dengue virus.

ally. The co-occurrence of these two symptoms reinforces the association with the likelihood of dengue (M_3 would be greater than M_1 and M_2).

$$fever = [> 7days] \rightarrow \boldsymbol{dengue = yes},\ (M_1)$$
$$thrombocytopenia = yes \rightarrow \boldsymbol{dengue = yes},\ (M_2)$$
$$fever = [> 7days], thrombocytopenia = yes \rightarrow \boldsymbol{dengue = yes},\ (M_3)$$

Type 7 Group

Similar to Type 6, Type 7 also deals with rules where the item of interest is in the consequent; however, in this case, one of the other items is not individually associated with the item of interest. In the example below, the first rule presents the association between fever for more than seven days and dengue, while the second rule shows that mild bleeding enhances this association (with $M_2 > M_1$). Moreover, there is no rule associating mild bleeding with dengue.

$$fever = [> 7days] \rightarrow \boldsymbol{dengue = yes},\ (M_1)$$
$$mildBleeding = yes, fever = [> 7days] \rightarrow \boldsymbol{dengue = yes},\ (M_2)$$

Type 8 Group

Type 8 highlights items that, when examined individually, are not associated with the item of interest in the consequent but, when combined, demonstrate a significant relationship. In the example, petechiae and intense abdominal pain are associated with dengue. However, there are no other rules associating these symptoms with dengue individually.

$$petechiae = yes, abdominalPain = intense \rightarrow \boldsymbol{dengue = yes}$$

Algorithm

The algorithm for the proposed method selects and groups rules according to the types described above. To this end, the algorithm input is a set R of rules of size 2 or 3 generated by an ARM algorithm based on support and confidence. The user must specify an item of interest, $\bar{a}$, corresponding to a variable with a specific value in the data set. The user must also define an evaluation measure, referred to as M, which is used to evaluate the correlation between the items present in a rule. Measures such as *Lift*, *Conviction*, and *Odds Ratio* are commonly used for this purpose. The value 1 in these measures indicates that there is no correlation between the items in the antecedent and the consequent. This means that the antecedent and consequent of the rule are independent. In this case, the rule is ignored. To ensure that only rules with a strong association between items are selected, the user can specify a threshold around the no-correlation value through the parameter δ. If $\delta = 0$, only rules with the exact no-correlation value will be removed.

A principle of the proposed method is that a larger rule should only be retained if it has some gain (measured by M) over its sub-rules. To prevent this gain from being insignificant, the user can set a parameter α, indicating the minimum relative gain between a size-3 rule and its size-2 sub-rules. The relative gain is the difference in M values between the size-3 rule and a size-2 sub-rule, divided by the M value of the size-2 rule. In other words, the M value for the size-3 rule should be $\alpha\%$ higher than that of the size-2 sub-rule. Equation 1 shows how the relative gain of rule r_2 is calculated concerning rule r_1.

$$RG(r_1, r_2) = \frac{M(r_2) - M(r_1)}{M(r_1)} \qquad (1)$$

In our experiments, we observed that using rules with lower confidence values might be interesting to enable the formation of rule groupings that exhibit relevant relationships. On the other hand, this can cause a side effect, generating many groups around size-3 rules with low confidence. The parameter c' was added to control this issue and specify minimum confidence for size-3 rules. Therefore, the c' parameter of the proposed method is not directly related to the minimum confidence of ARM algorithms. If $c' = 0$, all rules generated by the ARM algorithm will be considered. Algorithm 1 performs the procedure described above.

Visualizing the Interconnections Among Rule Groups

We observe that specific rules may appear in multiple groupings generated by the proposed method. Therefore, we have developed a graphical representation capable of condensing the information derived from these groupings. This is accomplished through a graph encompassing all generated groups connected by some common rule. Figure 1 provides a general example of this graph. In the graph, the red nodes represent rules shared by the groupings, referred to as pivot rules. The blue nodes depict groupings of Types 2, 4, 6, and 7, while the yellow nodes represent groupings of Type 1. The connections between groupings are established exclusively through pivot rules, a strategy employed to reduce the number of edges and enhance visualization. Establishing connections between groups of Types 3, 5, and 8 is not feasible. Types 5 and 8 consist of only one rule and lack a pivot rule. Similarly, Type 3 also lacks a pivot rule, given

```
Data: $R$: rule set generated by ARM; $\bar{a}$: item of interest; $\delta$: correlation threshold; $\alpha$: minimum relative gain; $c'$: minimum confidence for size three rules
Result: rule groupings categorized into 8 different types
Select from R the rules whose M value is greater than $1.0+\delta$:
  $R' = \{r_j | r_j \in R \wedge M(r_j) > 1.0+\delta\}$;
Select from $R'$ only the rules that have the item of interest $\bar{a}$: $R'' = \{r_j | r_j \in R \wedge \bar{a} \subset r_j\}$;
Remove 3 size rules with confidence less than $c'$: $R''' = \{r_j | r_j \in R'' \wedge (len(r_j) = 2 \vee Conf(r_j) \geq c')\}$;
Organize the rules into the following types:
Type 1: rule pairs of size 2, such as:
  $\{r_1 : a \rightarrow \bar{a}; r_2 : \bar{a} \rightarrow a\}$, where $r_1, r_2 \in R'''$.
Type 2: trios of rules, such as:
  $\{r_1 : a_1 \rightarrow a_2; r_2 : \bar{a} \rightarrow a_2; r_3 : a_1, \bar{a} \rightarrow a_2\}$,
  where $r_1 \in R'$, $r_2, r_3 \in R'''$, $RG(r_1, r_3) \geq \alpha$ and $RG(r_2, r_3) \geq \alpha$.
Type 3: pairs of rules, such as:
  $\{r_1 : a_1 \rightarrow a_2; r_2 : a_1, \bar{a} \rightarrow a_2\}$,
  where $r_1 \in R'$, $r_2 \in R'''$, $\bar{a} \rightarrow a_2 \notin R'''$ and $RG(r_1, r_2) \geq \alpha$.
Type 4: pairs of rules, such as:
  $\{r_1 : \bar{a} \rightarrow a_2; r_2 : a_1, \bar{a} \rightarrow a_2\}$,
  where $r_1, r_2 \in R'''$, $a_1 \rightarrow a_2 \notin R'$ and $RG(r_1, r_2) \geq \alpha$.
Type 5: unitary sets of rules, such as:
  $\{r : a_1, \bar{a} \rightarrow a_2\}$,
  where $r \in R'''$, $a_1 \rightarrow a_2 \notin R'$ and $\bar{a} \rightarrow a_2 \notin R'''$.
Type 6: trios of rules, such as:
  $\{r_1 : a_1 \rightarrow \bar{a}; r_2 : a_2 \rightarrow \bar{a}; r_3 : a_1, a_2 \rightarrow \bar{a}\}$,
  where $r_1, r_2, r_3 \in R'''$ , $RG(r_1, r_3) \geq \alpha$ and $RG(r_2, r_3) \geq \alpha$.
Type 7: pairs of rules, such as:
  $\{r_1 : a_1 \rightarrow \bar{a}; r_2 : a_1, a_2 \rightarrow \bar{a}\}$,
  where $r_1, r_2 \in R'''$, $a_2 \rightarrow \bar{a} \notin R'''$ and $RG(r_1, r_2) \geq \alpha$.
Type 8: unitary sets of rules, such as:
  $\{r : a_1, a_2 \rightarrow \bar{a}\}$,
  where $r \in R'''$, $a_1 \rightarrow \bar{a} \notin R'''$ and $a_2 \rightarrow \bar{a} \notin R'''$
```

Algorithm 1: Algorithm for rule selection and grouping.

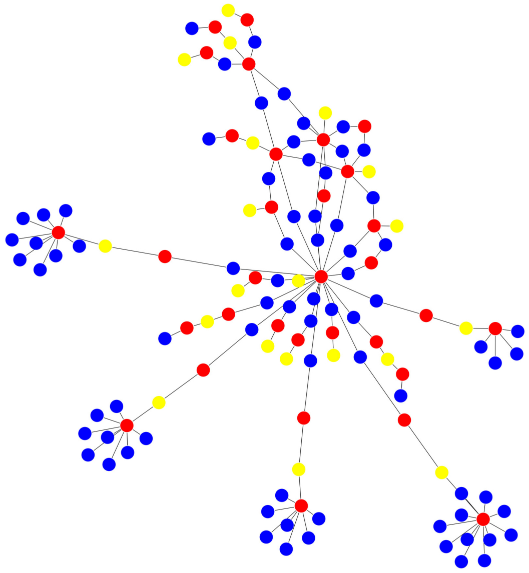

Figure 1: Graph illustrating the connections between rule groups. Red nodes denote pivot rules, yellow nodes correspond to Type 1 groups (bidirectional rules), and blue nodes represent other groups connected through pivot rules.

that the size-2 rule can only appear in a single grouping as it does not include the item of interest.

Figure 2 illustrates a subgraph example where the central red node represents a pivot rule establishing an association where dengue leads to bleeding. This pivot rule is also present in other groups, thereby connecting various associations with bleeding. For instance, medications such as acetylsalicylic acid (ASA), ibuprofen, and escitalopram enhance the association between dengue and bleeding despite lacking direct individual associations with bleeding (no specific association rules exist between these medications and bleeding). Additionally, the pivot rule connects associations related to severe liver disease, hemophilia, and bleeding. In this scenario, there are rules indicating that these diseases individually lead to bleeding, and other rules suggest that the presence of dengue with these diseases amplifies the occurrence of the association with bleeding. Lastly, the yellow node encompasses a bidirectional rule indicating that dengue and bleeding share a mutual relationship and can serve as a linkage point with other nodes featuring associations leading to dengue in different group rules.

3 CASE STUDIES

We conducted experiments on four real case studies to evaluate the effectiveness of the proposed method.

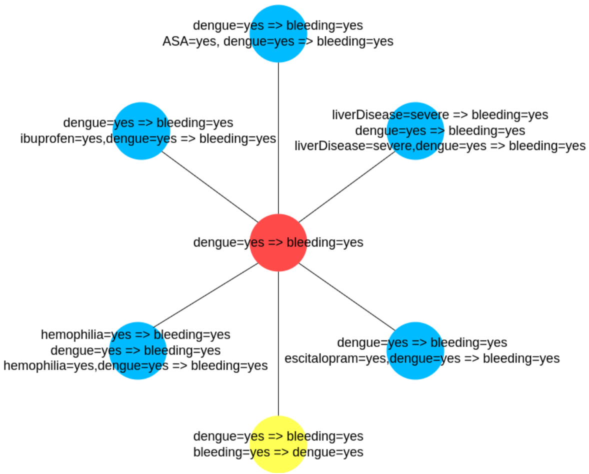
dengue=yes => bleeding=yes
ASA=yes, dengue=yes => bleeding=yes
dengue=yes => bleeding=yes
ibuprofen=yes,dengue=yes => bleeding=yes
liverDisease=severe => bleeding=yes
dengue=yes => bleeding=yes
liverDisease=severe,dengue=yes => bleeding=yes
dengue=yes => bleeding=yes
hemophilia=yes => bleeding=yes
dengue=yes => bleeding=yes
hemophilia=yes,dengue=yes => bleeding=yes
dengue=yes => bleeding=yes
escitalopram=yes,dengue=yes => bleeding=yes
dengue=yes => bleeding=yes
bleeding=yes => dengue=yes

Figure 2: Illustration of a subgraph exemplifying multiple associations interconnected through the relationship between dengue and bleeding.

The initial case study is derived from research by (Slyepchenko et al., 2017), later utilized by (Castro et al., 2018) as an Association Rule Mining (ARM) task. The primary objective is to investigate Premenstrual Dysphoric Disorder (PMDD) in women with Bipolar Disorder (BD). The other three studies utilize information from open sources provided by the Brazilian government on Severe Acute Respiratory Syndrome (SARS)[2]. Due to space constraints, we will focus on one of the SARS case studies related to COVID-19 and mortality. However, the results of rule groups from all case studies are available at *https://github.com/Luiz-Cintra-Experiments/masters-degree-experiments/tree/main/results*.

The ARM algorithm used to generate the association rules was the *Apriori* implemented in the R language, version 4.3.1, available in the library *arules* version 1.7-6 (Hahsler, 2023). The parameters used were: minimum support of 1% , minimum confidence of 30%, and maximum rule size equal to 3. This support value was chosen because some attributes have a low frequency at specific values. The confidence value was chosen to demonstrate that rules of size 2 with lower confidence may have their associations enhanced if the item in its antecedent occurs concomitantly with another item, forming a rule of size 3 that is more reliable.

The proposed method was implemented in Python language version 3.8, available at *https://github.com/Association-Rules-Post-Processing/ARPPL*. The parameters of the proposed method were: interest measure $M = Odds\ Ratio$ (OR), dependence margin $\delta = 0.1$, minimum improvement $\alpha = 10\%$, and minimum confidence for size 3 $c' = 50\%$. The graphs were generated using the library *Networkx* (Hagberg et al., 2008) in version 2.8.8 (NetworkX Developers, 2022), with the library *ForceAtlas2* version 1.0 (Shinn, 2016) being used to improve the layout of nodes in the graph.

3.1 Case Study: COVID-19-Related Mortality

The data used in this case study was compiled from four distinct SARS databases spanning the years 2019 to 2022. The final dataset comprises $2,166,443$ entries from patients with a confirmed diagnosis of COVID-19 and includes 34 variables [3]. In the final dataset, 95.75% of patients were hospitalized. Therefore, the results presented concern hospitalized patients and not a general perspective of the effects of COVID-19.

The *Apriori* algorithm generated $187,407$ association rules. These rules were then processed through the proposed method, with $evolution = covid\ death$ as the item of interest for rule grouping. This resulted in the formation of 116 groups, encompassing 215 distinct association rules. Only groupings of types 7 and 8 were not formed. Table 1 presents a subset of rules from these groups.

The two initial groupings reveal a bidirectional association (Type 1 group) between medical procedures such as the use of invasive respiratory support ($vent\_sup = inv$) and admission to the Intensive Care Unit (ICU) ($ICU = yes$) with death due to COVID-19 ($evolution = covid\ death$). In the first grouping (G_1), it is observed that patients using invasive respiratory support (rule 1) had a high occurrence of mortality (Conf. = 74.5%). On the other hand, among patients who died from COVID-19 (rule 2), there is an occurrence of invasive respiratory support, but to a lesser extent (Conf. = 41.3%). In the second grouping (G_2), the occurrence of items in the antecedent and consequent of the rules appears balanced. Of the patients admitted to the ICU (rule 3), around half died from COVID-19 (Conf. = 53.1). Likewise, of the patients who died from COVID-19 (rule 4), just over half were admitted to the ICU (Conf. = 55.3%). The following three groupings show bidirectional associations of respiratory symptoms such as respiratory discomfort ($resp\_discomfort = yes$), low blood oxygen saturation ($blood\_oxygen = [< 95]$), and dyspnea ($dypnea = yes$) with death. We can observe that given these respiratory symptoms (rules 5, 7, and 9),

[2] https://opendatasus.saude.gov.br/dataset?tags=SRAG

[3] Available at: https://github.com/Luiz-Cintra-Databases/SRAG-OpenDataSUS-2019-2022/tree/main/srag/database

Table 1: Subset of rule groups formed for the analysis of COVID death (item of interest $evolution = covid\ death$). Gr.: group number; Tp.: group type; Id: rule identification; Sup.: rule support (in %); Conf.: rule confidence (in %); OR: rule odds ratio.

Gr. (Tp.)	Id: Rule	Sup.	Conf.	OR
G_1 (1)	1: $vent\_sup = inv \rightarrow evolution = covid\ death$	12.8	74.5	10.4
	2: $evolution = covid\ death \rightarrow vent\_sup = inv$	12.8	41.3	10.4
G_2 (1)	3: $ICU = yes \rightarrow evolution = covid\ death$	17.1	53.1	4.4
	4: $evolution = covid\ death \rightarrow ICU = yes$	17.1	55.3	4.4
$G_3(1)$	5: $resp\_discomfort = yes \rightarrow evolution = covid\ death$	19.3	35.3	1.6
	6: $evolution = covid\ death \rightarrow resp\_discomfort = yes$	19.3	62.3	1.6
$G_4(1)$	7: $blood\_oxygen = [< 95] \rightarrow evolution = covid\ death$	21.8	35.3	1.7
	8: $evolution = covid\ death \rightarrow blood\_oxygen = [< 95]$	21.8	70.2	1.7
$G_5(1)$	9: $dyspnea = yes \rightarrow evolution = covid\ death$	23.4	33.9	1.6
	10: $evolution = covid\ death \rightarrow dyspnea = yes$	23.4	75.6	1.6
$G_6(1)$	11: $risk\_fac = yes \rightarrow evolution = covid\ death$	22.5	37.3	2.2
	12: $evolution = covid\ death \rightarrow risk\_fac = yes$	22.5	72.7	2.2
$G_7(2)$	12: $evolution = covid\ death \rightarrow risk\_fac = yes$	22.5	72.7	2.2
	14: $vaccinated = yes \rightarrow risk\_fac = yes$	14.1	70.5	1.7
	15: $vaccinated = yes, evolution = covid\ death \rightarrow risk\_fac = yes$	4.8	81.1	3.0
$G_8(2)$	16: $ICU = yes \rightarrow vent\_sup = inv$	14.4	44.8	19.2
	2: $evolution = covid\ death \rightarrow vent\_sup = inv$	12.8	41.3	10.4
	18: $ICU = yes, evolution = covid\ death \rightarrow vent\_sup = inv$	11.1	64.7	23.1
$G_9(6)$	1: $vent\_sup = inv \rightarrow evolution = covid\ death$	12.8	74.5	10.4
	20: $age = (75y+) \rightarrow evolution = covid\ death$	10.0	51.5	3.0
	21: $age = (75y+), vent\_sup = inv \rightarrow evolution = covid\ death$	3.3	86.9	16.4
$G_{10}(6)$	3: $ICU = yes \rightarrow evolution = covid\ death$	17.1	53.1	4.4
	20: $age = (75y+) \rightarrow evolution = covid\ death$	10.0	51.5	3.0
	24: $ICU = yes, age = (75y+) \rightarrow evolution = covid\ death$	4.8	68.8	5.6
$G_{11}(6)$	5: $resp\_discomfort = yes \rightarrow evolution = covid\ death$	19.3	35.3	1.6
	20: $age = (75y+) \rightarrow evolution = covid\ death$	10.0	51.5	3.0
	27: $resp\_discomfort = yes, age = (75y+) \rightarrow evolution = covid\ death$	6.1	56.8	3.4
G_{12} (6)	28: $sx\_date = [5/2020\text{-}8/2020] \rightarrow evolution = covid\ death$	5.5	33.0	1.1
	20: $age = (75y+) \rightarrow evolution = covid\ death$	10.0	51.5	3.0
	30: $sx\_date = [5/2020\text{-}8/2020], age = (75y+) \rightarrow evolution = covid\ death$	2.1	58.7	3.3

the occurrence of death is close to one-third (confidence of the rules is 35.5% and 33.9%). However, patients who died (rules 6, 8, and 10) had a higher occurrence of these symptoms (confidence of the rules is 62.3%, 70.2%, and 75.6%, respectively). The last Type 1 grouping (G_6) presented in the Table 1 shows a bidirectional association between a patient's risk factor ($risk\_fac = yes$) and death from COVID-19. Like the three previous groupings, the risk factor (rule 11) implies death from COVID-19 in close to a third of cases (Conf. = 37.3%). However, given a patient who died from COVID-19 (rule 12), the occurrence of them having some risk factor is high (Conf. = 72.7%).

Two Type 2 groupings are illustrated in Table 1. The grouping G_7 shows the previous association rule 12 in which patients who died from Covid were more likely to have some risk factor. Additionally, there is an association between vaccinated COVID patients ($vaccinated = yes$) who also had a higher occurrence of having some risk factor (rule 14). Rule 15 then establishes that the association of vaccinated patients who died from COVID increases the possibility of them having had a risk factor. This is concluded from the increase in confidence and odds ratio of rule 15 (Conf.=81.1% and OR=3.0) concerning rules 12 (Conf.=72.7% and OR=2.2) and 14 (Conf.= 70.5% and OR=1.7). The grouping, G_8 shows an association in which patients admitted to the ICU lead to the occurrence of invasive respiratory support (rule 16). Additionally, the previous association rule 2 indicates that patients who died from COVID-19 also involved in the use of invasive respiratory support. Rule 18 reinforces these two previous associations, showing that patients admitted to the ICU who died increase the occurrence of using invasive respiratory support. This is also concluded by the increase in confidence and odds ratio of rule 18 (Conf.=64.7% and OR=23.1) concerning rules 16 (Conf.=44.8% and OR=19.2) and 2 (Conf.=41.3% and OR=10.4).

Type 6 groupings deal with the item of interest, death from COVID-19, in the consequent of the rules. For these groupings, we highlight associations that reflect how the need for specific procedures in individuals over 75 years old ($age = 75y+$) presents a higher risk of death. Grouping G_9 shows a previously seen association where the use of invasive respiratory support leads to death (rule 1). Besides, it shows that age over 75 also leads to death (rule 20). However, age acts as an important reinforcement when it occurs with ventilatory support (rule 21). We can verify that both confidence and odds ratio have a significant increase in rule 21 (Conf. = 86.9% and OR = 16.4) compared to rules 1 (Conf. = 74.5% and OR = 10.4) and 20 (Conf. = 51.5% and OR = 3.0). Similarly, rule 24 in grouping G_{10} indicates that elderly patients admitted to the ICU have a higher occurrence of death when looking at their confidence (68.8%) and odds ratio (5.6), which are higher than those of rules 3 and 20 seen earlier. In turn, rule 27 in grouping G_{11} indicates that the symptom of respiratory discomfort in the elderly also intensifies the evolution of covid to death, according to its higher confidence value (56.8%) and odds ratio (3.4) compared to the previous rules 25 and 20. G_{12} is the last grouping presented in Table 1. It shows that patients with COVID-19 symptoms in the second third of 2020 ($sx\_date = [5/2020 - 8/2020]$) are associated with death from COVID-19 (rule 28). Although this association is not as strong (Conf. = 33.0% and OR = 1.1), this type of association was not generated for other dates. However, the association is enhanced in patients over 75 years old (rule 30) (Conf. = 58.7% and OR = 3.3).

Visualization of Rule Groupings

Figure 3 shows the graph of generated groups that present relationship between groups. The image has three subgraphs corresponding to groups of Type 6 and Type 7 (the subgraphs that look like a fireworks explosion), they are linked to the center of the graph through bidirectional rules (yellow nodes). The center of the graph shows the relationship between groups of Type 2 and Type 4. A subgraph with only three groups was generated and is not related to the rest of the graph. The two subgraphs highlighted in the Figure 3 were adjusted manually, due to limitation on the number of the pages, and will be presented below.

Figure 4 highlights all the forms of relationships between the groups that the method is able to capture. A bidirectional rule will show that the item of interest also acts as a factor in the occurrence of the associated item. Therefore, the yellow nodes provide a bridge to verify both the factors that lead to the occurrence of the item of interest, what the item of interest acts as a factor and whether any item can reinforce such an association. For example, $age = (75y+) \rightarrow evolution = covid\ death$ and $evolution = covid\ death \rightarrow age = (75y+)$, furthermore, the latter has a connection with another group that shows that in patients with neurological diseases and who have had Covid there is a greater chance of being over 75 years old.

The graph also shows that $resp\_discomfort = yes$, $ven\_sup = inv$ e $icu = yes$ are associated with death due to COVID-19 through the blue nodes linked to rule $age = (75y+) \rightarrow evolution = covid\ death$. The remaining bidirectional rules convey that items are inherently associated to death due to COVID-19 ($resp\_discomfort = yes$ is linked to a bidirectional, which in turn is linked to center of graph, see Figure 3). Thereby, we can see that not only $ven\_sup = inv \rightarrow evolution = covid\ death$ but also $evolution = covid\ death \rightarrow ven\_sup = inv$, additionally the association is reinforced when $evolution = covid\ death$ and $icu = yes$ occurs together. A similar relationship can also be seen in the rule $icu = yes \rightarrow evolution = covid\ death$, where $evolution = covid\ death \rightarrow icu = yes$ also occurs, in addition, the association is reinforced when $evolution = covid\ death$ occurs concomitantly with $tomo\_res = typical\ covid$ or $obesity = yes$.

Relationship between groups of Types 2 and 4 are showed in Figure 5. This type of relationship shows when the item of interest acts as a factor for another item to occurs, moreover to showing the other items that reinforced this association. For example, in Figure 5, in addition to showing that $evolution = covid\ death \rightarrow risk\_fac = yes$ occurs, the graph also shows that several items reinforced this association, such as, $icu = yes$, $blood\_oxygen = [< 95]$, $fever = not$ e $sore\_throat = not$. The subgraph also shows a bidirectional that conveys that $evolution = covid\ death$ and $risk\_fac = yes$ are intrinsically associated.

4 DISCUSSION

Few studies have been found on post-processing association rules based on an item of interest in the data. In a literature review since 2015 on this subject, only three works were found that focus on an item of interest in the consequent (Berka, 2018) (Cheng et al., 2016) (Hahsler and Karpienko, 2017), and a single work that focuses on an item of interest in the antecedent (Wei and Scott, 2015). However, no works address an item of interest on either side of the rule, as in our proposed method.

Berka's work (Berka, 2018) focuses on describing concepts by fixing the chosen concept in the con-

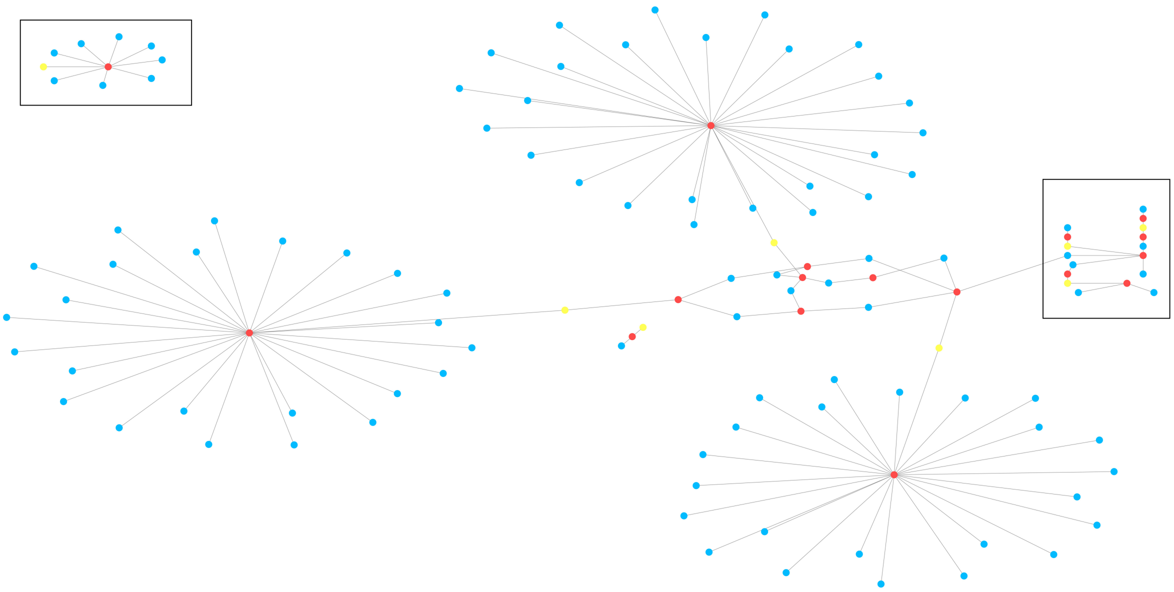

Figure 3: Graph showing the relationships among all rule groupings. The detailed view of the highlighted square on the right side is presented in Figure 4, and the detailed view of the highlighted square in the upper left corner is provided in Figure 5.

sequent of the rule and generating meta-rules. His method involves, from rules obtained by the ARM algorithm, filtering those with a specific consequent (the concept to be explored) and generating a new database where each rule is mapped as a row, maintaining the columns of the original database except for the column that was fixed in the consequent. The ARM algorithm is then applied to this new database, and the result of this application provides information associated with the sought concept. Thus, a second level of information is obtained compared to traditional rules. As a consequence of this approach, he can get the association between items that lead to the occurrence of a specific concept. However, associations in which the concept acts as a factor for the presence of another item are lost. Our method is more generic and retains these associations in which the item of interest acts as a factor for the occurrence of other items (item of interest in the antecedent of the rule).

Cheng and colleagues (Cheng et al., 2016) propose a visualization based on a set of association rules with the same consequent, where an expert can “assemble” a rule by adding items. This strategy allows the expert to try to build rules they already have a prior assumption. As this construction progresses, the expert validates the relevance of the constructed rules based on the support and confidence provided by the visualization. Hahsler and Karpienko (Hahsler and Karpienko, 2017) also propose a visualization method where clusters of antecedents are generated for rules with the same consequent. First, a matrix is created where the rows represent the consequent, and the columns represent the antecedents (both itemsets). Then, the columns are grouped using the *k-means* algorithm, checking the lift measure. The authors argue that the lift measure allows matching conditions of synonyms or similar items, such as butter and margarine. The visualization consists of a balloon plot where rows are the consequent, columns are the clusters of antecedents, and a point on the plot represents a rule. The points correspond to balloons with properties such as color, size, and balloon position being used to highlight the clusters. Although the method does not focus on a specific item of interest, as all consequents are arranged in the rows of the visualization, it facilitates an analysis of a particular item in the consequent. Moreover, as their goal is to group similar items, the clustering method does not highlight the individual associations contained in less general rules. In both mentioned works that use visualization, only associations with an item of interest in the consequent are highlighted, but not those where an item of interest appears in the antecedent. In our approach, we consider that rules with an item of interest in the antecedent are essential because they allow for the analysis of the consequences of this occurrence, and it is possible to visualize these relationships in the graph in a grouped manner.

Wei and Scott (Wei and Scott, 2015) combine pruning, summarization, and visualization to find patterns in adverse reactions to vaccines in the United

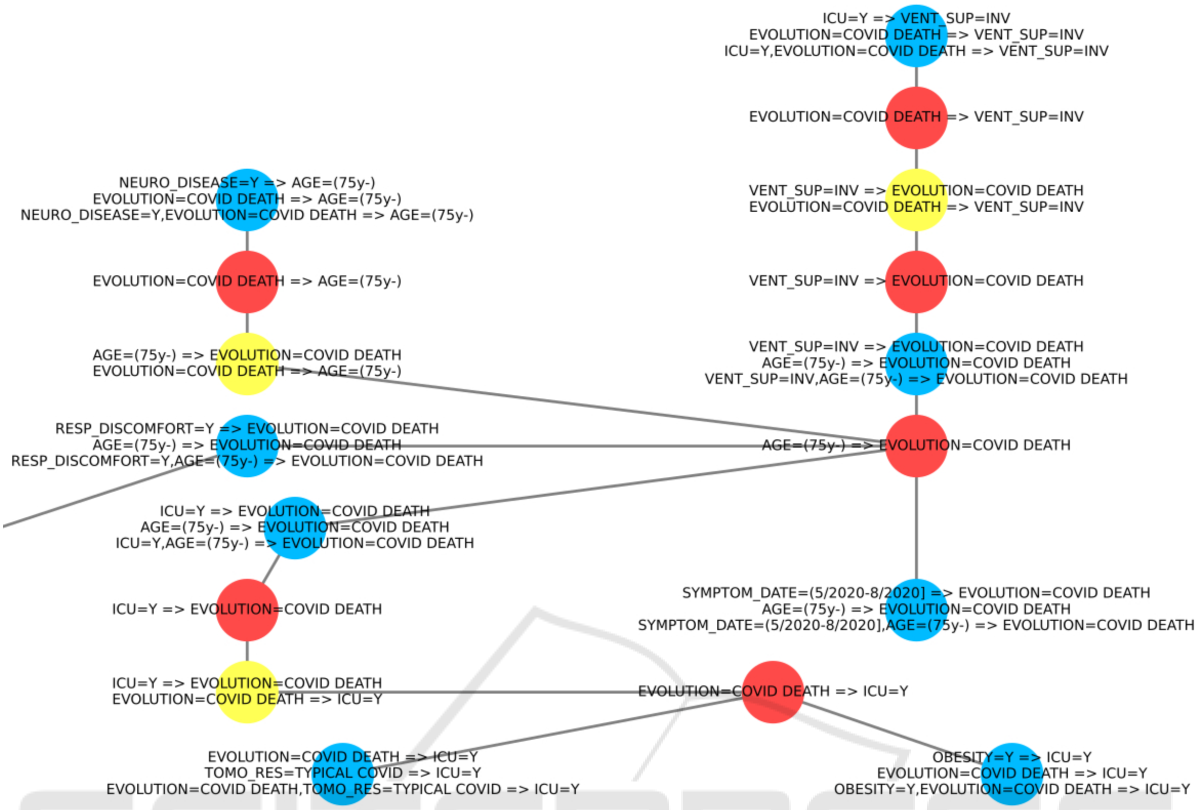
ICU=Y => VENT_SUP=INV
EVOLUTION=COVID DEATH => VENT_SUP=INV
ICU=Y,EVOLUTION=COVID DEATH => VENT_SUP=INV
EVOLUTION=COVID DEATH => VENT_SUP=INV
NEURO_DISEASE=Y => AGE=(75y-)
EVOLUTION=COVID DEATH => AGE=(75y-)
NEURO_DISEASE=Y,EVOLUTION=COVID DEATH => AGE=(75y-)
VENT_SUP=INV => EVOLUTION=COVID DEATH
EVOLUTION=COVID DEATH => VENT_SUP=INV
EVOLUTION=COVID DEATH => AGE=(75y-)
VENT_SUP=INV => EVOLUTION=COVID DEATH
AGE=(75y-) => EVOLUTION=COVID DEATH
EVOLUTION=COVID DEATH => AGE=(75y-)
VENT_SUP=INV => EVOLUTION=COVID DEATH
AGE=(75y-) => EVOLUTION=COVID DEATH
VENT_SUP=INV,AGE=(75y-) => EVOLUTION=COVID DEATH
RESP_DISCOMFORT=Y => EVOLUTION=COVID DEATH
AGE=(75y-) => EVOLUTION=COVID DEATH
RESP_DISCOMFORT=Y,AGE=(75y-) => EVOLUTION=COVID DEATH
AGE=(75y-) => EVOLUTION=COVID DEATH
ICU=Y => EVOLUTION=COVID DEATH
AGE=(75y-) => EVOLUTION=COVID DEATH
ICU=Y,AGE=(75y-) => EVOLUTION=COVID DEATH
SYMPTOM_DATE=(5/2020-8/2020] => EVOLUTION=COVID DEATH
AGE=(75y-) => EVOLUTION=COVID DEATH
SYMPTOM_DATE=(5/2020-8/2020],AGE=(75y-) => EVOLUTION=COVID DEATH
ICU=Y => EVOLUTION=COVID DEATH
ICU=Y => EVOLUTION=COVID DEATH
EVOLUTION=COVID DEATH => ICU=Y
EVOLUTION=COVID DEATH => ICU=Y
EVOLUTION=COVID DEATH => ICU=Y
TOMO_RES=TYPICAL COVID => ICU=Y
EVOLUTION=COVID DEATH,TOMO_RES=TYPICAL COVID => ICU=Y
OBESITY=Y => ICU=Y
EVOLUTION=COVID DEATH => ICU=Y
OBESITY=Y,EVOLUTION=COVID DEATH => ICU=Y

Figure 4: Subgraph displaying interconnections among various types of groups of rules.

States. In this work, the item of interest acts as the sole antecedent of the rule, and the consequent is a set of items representing the adverse effects of a vaccine. This work aligns with ours in studying the item of interest in the antecedent, but it does not address when the item is in the consequent, making our proposal more general as we can analyze the causes and consequences of an item of interest in other items in the database.

We can categorize the post-processing of association rules into four main types of tasks: pruning, grouping, summarization, and visualization (Baesens et al., 2000). Our method performs three of these tasks: pruning (selecting rules based on an item of interest and an evaluation measure), grouping (grouping the selected rules into types), and visualization (generating a graph to visualize the interconnections of the formed rule groups).

The pruning performed by our method is based, in parts, on the concept presented by (Bayardo et al., 2000), but using a minimum percentage improvement in the pruning process. The grouping carried out by our method differs significantly from the groupings found in recent literature, as it employs a predefined format of the relationship between rules (subsumed rules). Works such as (de Padua et al., 2018) and (Karimi-Majd and Mahootchi, 2015) perform groupings based on the similarity of items between rules, enabling the display of other related rules. However, if a specialist is interested in verifying, from a rule of size 3, whether the antecedents are individually associated, these additional rules will increase the effort required for analysis. The same issue occurs in the study by (Zhang et al., 2019), which groups rules using rule semantics, making it unclear to the specialist the individual associations in relation to size 3 associations.

Some works use graphs to visualize rules, but the approach of fixing a pivot rule was not found. The idea of using the pivot rule came from adapting the approaches of (Wei and Scott, 2015) and (Kwon and Kim, 2019), where the rules are nodes, and the items are represented with different visual marks. Since the work of (Wei and Scott, 2015) fixes an antecedent, it does not need to worry about showing the antecedent, unlike (Kwon and Kim, 2019), which needs to show the antecedent by connecting to the node and the node connecting to the consequent.

Microsoft provides an alternative visualization method through its proprietary software, MS Analysis

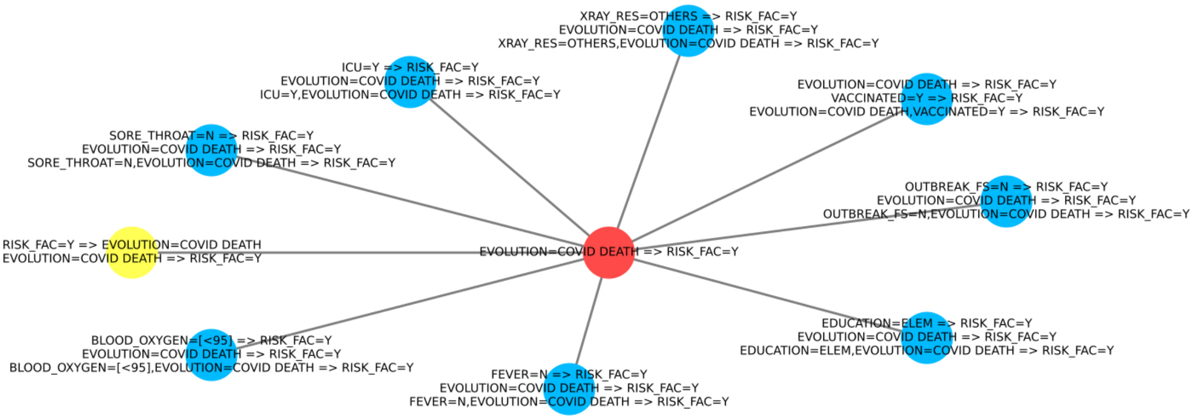
XRAY_RES=OTHERS => RISK_FAC=Y
EVOLUTION=COVID DEATH => RISK_FAC=Y
XRAY_RES=OTHERS,EVOLUTION=COVID DEATH => RISK_FAC=Y
ICU=Y => RISK_FAC=Y
EVOLUTION=COVID DEATH => RISK_FAC=Y
ICU=Y,EVOLUTION=COVID DEATH => RISK_FAC=Y
EVOLUTION=COVID DEATH => RISK_FAC=Y
VACCINATED=Y => RISK_FAC=Y
EVOLUTION=COVID DEATH,VACCINATED=Y => RISK_FAC=Y
SORE_THROAT=N => RISK_FAC=Y
EVOLUTION=COVID DEATH => RISK_FAC=Y
SORE_THROAT=N,EVOLUTION=COVID DEATH => RISK_FAC=Y
OUTBREAK_FS=N => RISK_FAC=Y
EVOLUTION=COVID DEATH => RISK_FAC=Y
OUTBREAK_FS=N,EVOLUTION=COVID DEATH => RISK_FAC=Y
RISK_FAC=Y => EVOLUTION=COVID DEATH
EVOLUTION=COVID DEATH => RISK_FAC=Y
EVOLUTION=COVID DEATH => RISK_FAC=Y
EDUCATION=ELEM => RISK_FAC=Y
EVOLUTION=COVID DEATH => RISK_FAC=Y
EDUCATION=ELEM,EVOLUTION=COVID DEATH => RISK_FAC=Y
BLOOD_OXYGEN=[<95] => RISK_FAC=Y
EVOLUTION=COVID DEATH => RISK_FAC=Y
BLOOD_OXYGEN=[<95],EVOLUTION=COVID DEATH => RISK_FAC=Y
FEVER=N => RISK_FAC=Y
EVOLUTION=COVID DEATH => RISK_FAC=Y
FEVER=N,EVOLUTION=COVID DEATH => RISK_FAC=Y

Figure 5: Subgraph showing multiple associations interconnected through the relationship between COVID-19-related deaths and a risk factor.

Services[4]. The visualization in the MS software represents a dependency network, i.e., a directed graph where items are nodes, linking antecedents to consequents. A key distinction from our approach is that our graph nodes can represent individual rules (if it is a pivot rule) or groups of rules rather than individual items. This distinction simplifies the visualization by avoiding complex cycles between items and reducing the number of edges. Additionally, each group connects to only one pivot rule and vice versa, further minimizing edges. Our method is tailored to extract and emphasize more intricate relationships within associations for the usefulness of specialists.

Thus, although the proposed method incorporates adaptations of existing methods and concepts, our work presents some notable differences compared to the most similar approaches found in the literature on association rule post-processing. Primarily, it is focused on the user's interest in a specific item. Due to this emphasis, we address the item of interest more comprehensively, acknowledging its relevance both in the antecedent and consequent of the rule. Therefore, pruning, grouping, and visualization operations are optimized to meet the user's needs.

5 CONCLUSIONS

Developing an effective method in Association Rule Mining (ARM) is crucial to harness the exploratory nature inherent in this process, as it can result in many rules, making it challenging for a domain-specific expert to manipulate and analyze them later. Unlike the inductive bias of supervised approaches, ARM seeks statistically relevant patterns without relying on a specific target variable. Instead of simply filtering by the item of interest and potentially concealing relevant relationships, the proposed method allows the selection of rules that, even without directly containing the item of interest, share other items with those that do. This approach promotes knowledge discovery by highlighting complex and valuable connections in the dataset and is beneficial for domain experts seeking specific insights into their datasets, contributing significantly to understanding and interpreting the relationships in the data.

A limitation of the work is that the proposed method only considers rules with a maximum size of three, which may lead to the exclusion of potentially relevant rules and limit the scope of the analysis. Furthermore, the method depends on the rules generated by the ARM algorithm and the parameters defined to create them. For example, rare items may only appear if the minimum support used to generate the rules is lower than the frequency of the item in the database. For future work, we plan to assess differences in the formed rule groupings by considering various evaluation measures. In the graphical representation, there is room for improvement in visualization, particularly in areas with significant edge overlap. Additionally, we aim to enhance the level of detail in the graphs by incorporating additional information about the rules, such as support, confidence, and the evaluation measure used for forming the groupings. These enhancements are expected to contribute to a more comprehensive and complete analysis of the associations within rule groupings.

[4] https://learn.microsoft.com/en-us/analysis-services/data-mining/browse-a-model-using-the-microsoft-association-rules-viewer?view=asallproducts-allversions#BKMK_Dependency

REFERENCES

Agrawal, R., Imieliński, T., and Swami, A. (1993). Mining association rules between sets of items in large databases. ACM SIGMOD Record, 22:207–216.

Agrawal, R. and Srikant, R. (1994). Fast algorithms for mining association rules in large databases. In Proceedings of the 20th International Conference on Very Large Data Bases, page 487–499. Morgan Kaufmann Publishers Inc.

Baesens, B., Viaene, S., and Vanthienen, J. (2000). Post-processing of association rules. In KDD '00: Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Special workshop on post-processing in machine learning and data mining: interpretation, visualization, integration, and related topics. Association for Computing Machinery.

Balasubramani, B. S., Shivaprabhu, V. R., Krishnamurthy, S., Cruz, I. F., and Malik, T. (2016). Ontology-based urban data exploration. In Proceedings of the 2nd ACM SIGSPATIAL Workshop on Smart Cities and Urban Analytics - UrbanGIS '16, pages 1–8. ACM Press.

Bayardo, R. J., Agrawal, R., and Gunopulos, D. (2000). Constraint-based rule mining in large, dense databases. Data Mining and Knowledge Discovery, 4:217–240.

Berka, P. (2018). Comprehensive concept description based on association rules: A meta-learning approach. Intelligent Data Analysis, 22:325–344.

Castro, G., Salvini, R., Soares, F. A., Nierenberg, A. A., Sachs, G. S., Lafer, B., and Dias, R. S. (2018). Applying Association Rules to Study Bipolar Disorder and Premenstrual Dysphoric Disorder Comorbidity. In 2018 IEEE Canadian Conference on Electrical & Computer Engineering (CCECE), pages 1–4. IEEE.

Cheng, C.-W., Sha, Y., and Wang, M. D. (2016). Intervisar: An interactive visualization for association rule search. In Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, pages 175–184. ACM.

de Padua, R., do Carmo, L. P., Rezende, S. O., and de Carvalho, V. O. (2018). An analysis on community detection and clustering algorithms on the post-processing of association rules. In 2018 International Joint Conference on Neural Networks (IJCNN), volume 2018-July, pages 1–7. IEEE.

Freitas, A. A. (2000). Understanding the crucial differences between classification and discovery of association rules. ACM SIGKDD Explorations Newsletter, 2:65–69.

Hagberg, A. A., Schult, D. A., and Swart, P. J. (2008). Exploring network structure, dynamics, and function using networkx. In Varoquaux, G., Vaught, T., and Millman, J., editors, Proceedings of the 7th Python in Science Conference, pages 11 – 15, Pasadena, CA USA.

Hahsler, M. (2023). R package arules - mining association rules and frequent itemsets. https://www.rdocumentation.org/packages/arules/versions/1.7-6. Accessed: 2023-08-05.

Hahsler, M. and Karpienko, R. (2017). Visualizing association rules in hierarchical groups. Journal of Business Economics, 87:317–335.

Han, J., Kamber, M., and Pei, J. (2012). Data Mining: Concepts and Techniques. Elsevier Inc.

Han, J., Pei, J., Yin, Y., and Mao, R. (2004). Mining frequent patterns without candidate generation: A frequent-pattern tree approach. Data Mining and Knowledge Discovery, 8:53–87.

Karimi-Majd, A.-M. and Mahootchi, M. (2015). A new data mining methodology for generating new service ideas. Information Systems and e-Business Management, 13:421–443.

Kwon, J.-H. and Kim, E.-J. (2019). Accident prediction model using environmental sensors for industrial internet of things. Sensors and Materials, 31:579.

Matetic, M., Bakaric, M. B., and Sisovic, S. (2015). Association rule mining and visualization of introductory programming course activities. In Proceedings of the 16th International Conference on Computer Systems and Technologies - CompSysTech '15, volume 1008, pages 374–381. ACM Press.

NetworkX Developers (2022). Networkx - network analysis in python. https://networkx.org/documentation/networkx-2.8.8/. Accessed: 2023-08-05.

Shinn, M. (2016). Forceatlas2 for python. https://github.com/mwshinn/forceatlas2-python/. Accessed: 2023-08-05.

Slyepchenko, A., Frey, B. N., Lafer, B., Nierenberg, A. A., Sachs, G. S., and Dias, R. S. (2017). Increased illness burden in women with comorbid bipolar and premenstrual dysphoric disorder: data from 1 099 women from STEP-BD study. Acta Psychiatrica Scandinavica, 136(5):473–482.

Wei, L. and Scott, J. (2015). Association rule mining in the us vaccine adverse event reporting system (vaers). Pharmacoepidemiology and Drug Safety, 24:922–933.

Weidner, D., Atzmueller, M., and Seipel, D. (2020). Finding Maximal Non-redundant Association Rules in Tennis Data, volume 12057 LNAI, pages 59–78. Springer.

Zaki, M. (2000). Scalable algorithms for association mining. IEEE Transactions on Knowledge and Data Engineering, 12:372–390.

Zhang, C., Xue, X., Zhao, Y., Zhang, X., and Li, T. (2019). An improved association rule mining-based method for revealing operational problems of building heating, ventilation and air conditioning (hvac) systems. Applied Energy, 253:113492.

Zhang, C., Zhao, Y., Lu, J., Li, T., and Zhang, X. (2021). Analytic hierarchy process-based fuzzy post mining method for operation anomaly detection of building energy systems. Energy and Buildings, 252:111426.