



# A Web-Based Platform for Mining and Ranking Association Rules

Addi Ait-Mlouk<sup>(✉)</sup> and Lili Jiang

Department of Computing Science, Umeå University, Umeå, Sweden  
{addi.ait-mlouk,lili.jiang}@cs.umu.se

**Abstract.** In this demo, we introduce an interactive system, which effectively applies multiple criteria analysis to rank association rules. We first use association rules techniques to explore the correlations between variables in given data (i.e., database and linked data (LD)), and secondly apply multiple criteria analysis (MCA) to select the most relevant rules according to user preferences. The developed system is flexible and allows intuitive creation and execution of different algorithms for an extensive range of advanced data analysis topics. Furthermore, we demonstrate a case study of association rule mining and ranking on road accident data.

**Keywords:** Data mining · Association rules · Multiple criteria analysis

## 1 Introduction

Association rules mining is a powerful technique in data mining [3] for discovering correlation and relationships between objects. This technique like other data mining techniques has been integrated into a bunch of open-source tools, such as WEKA [11], TANAGRA [10], fpm [4] and DM-MCDA [1]. Most public tools are desktop-based and don't provide intuitive use for the data mining community, hence reduce the chance of re-usability and/or upgradability by developers.

In this demo, we (1) implement a baseline approach to extract association rules from given data, (2) propose an approach based on multiple criteria analysis to rank the extracted rules according to the decision maker's preferences, (3) demonstrate the idea of association rules mining through road accident data as a case study, where the dataset containing information on location, drivers, and the accident characteristics, the vehicles involved and victims [2], and (4) present the preliminary results of the proposed system. Our implementation will be open-source and a live demo can be found at <https://youtu.be/QILaVUghlSM>.

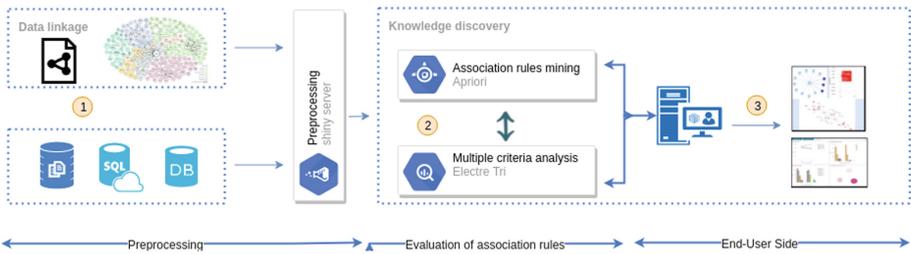
## 2 Mining and Ranking Association Rules

Figure 1 shows our framework architecture which includes three main modules: *data preprocessing*, *association rules mining*, and *multiple criteria analysis*. *Data preprocessing* allows users to process and prepare data to be used by *data mining*

algorithms to extract association rules based on some input thresholds (minimum support and minimum confidence). Afterward, we use *multiple criteria analysis* to rank the relevant rules based on user preferences and assign the result to different relevant categories. The system was implemented on R [7] and r shiny [9], providing interactive and user-friendly interfaces for different modules. The system can be extended with new data sets on demand, rendering information retrieval and query linked data using SPARQL queries (data linkage).

### 2.1 Quality Measurement of Association Rules

In order to select interesting rules from the large set of extracted rules, constraints on various measures of significance and interest are used [5]. The best-known constraints are minimum support, minimum confidence and lift. Due to page limitations, we refer readers to [5] for further references.



**Fig. 1.** Overview of system architecture: (1) Data preprocessing and data linkage, (2) Association rules mining with multiple criteria analysis, and (3) Visualization

### 2.2 Multiple Criteria Analysis (MCA)

MCA is a sub-field of operational research, and management science, dedicated to the development of decision support tools in order to solve complex decision problems involving multiple criteria objectives. When modeling a real decision problem using multiple criteria analysis, several issues [8] can be considered: choice, sorting and ranking. In our context of a large number of extracted rules, we apply the multiple criteria ranking precisely, in a method called ELECTRE (ELimination and Choice Expressing REality) TRI [5,6]. We conduct the following three computations to present the process of ranking:

1. The partial concordance indices  $c_j(a, b_h)$ : (it represents the degree of concordance with the hypothesis of outranking of  $a$  over  $b$  where  $a$  is rule and  $b$  is profile):

$$C_j(a, b_h) = \begin{cases} 0 & \text{if } g_j(b_h) - g_j(a) \geq p_j(b_h) \\ 1 & \text{if } g_j(b_h) - g_j(a) \leq q_j(b_h) \\ & \text{if not } \frac{p_j(b_h) + g_j(a)}{p_j(b_h) - q_j(b_h)} \end{cases} \quad (1)$$

- The discordance indices  $d_j(a, b_h)$ : (it represents the degree of discordance with the hypothesis that  $a$  outranks  $b$  on elementary criterion  $g_j$ ).

$$d_j(a, b_h) = \begin{cases} 0 & \text{if } g_j(a_h) \preceq g_j(b_h) + p_j(b_h) \\ 1 & \text{if } g_j(a_h) \succ g_j(b_h) + v_j(b_h) \\ & \text{if not } \in [0, 1] \end{cases} \quad (2)$$

- The credibility indices  $\sigma(a, b_h)$ : (it represents the credibility of the outranking of  $a$  over  $b$  on criterion  $j$ ):

$$\sigma(a, b_h) = C(a, b_h) \prod_{j \in \bar{F}} \frac{1 - d_j(a, b_h)}{1 - C(a, b_h)} \quad (3)$$

where,  $K_j$  is the weight of criteria  $j$ ,  $C_j(a, b_h)$  is the partial concordance index of criteria  $j$ , and  $F = \{j \in F : d_j(a, b_h) > C(a, b_h)\}$

The pseudo-code of the proposed algorithm takes as input a list of extracted rules  $(a_1, a_2, \dots, a_n)$ , a list of quality measurement (support, confidence, lift), a list of profiles  $(b_1, b_2)$  and the preference ( $p$ ), indifference ( $q$ ), veto ( $v$ ), cutting level  $\lambda$  thresholds. It computes the concordance, discordance and credibility indices for each rule under a given quality measurement. After iterating all extracted rules, the algorithm compares the credibility indices  $\sigma(a, b_h)$  and the default value of  $\lambda$  (0.7) then generates a list of relevant rules [5].

### 3 Demonstration

Firstly, users are allowed to upload a new CSV dataset or use SPARQL query to retrieve data from linked data (Dbpedia, wikidata, etc.). As shown in Fig. 2, the left panel displays data file input for a dataset (support CSV file) and parameter settings for some thresholds (attributes, minsup, minconf, etc). Secondly, users extract association rules and the main panel displays seven different tabs to characterize association rules across data sources and their summaries information. *Summary* tab shows the summary of rules mining algorithm, *ScatterPlot* tab shows the scatter-plot of extracted rules, *FrequentItemsets* tab shows frequent itemsets from data, and the rest of tabs shows interactive techniques to visualize association rules.

Figure 3 shows the association rules evaluation interface. The parameters configuration panel is displayed on the left side including method choice and the list of association rules with quality measurement. The main panel presents different steps for ELECTRE TRI (Eqs (1)–(3)), afterward, user can move to the bottom panels and clicks ‘Decision matrix’ tab to see the performance table of association rules associated with predefined thresholds (Table 2) and quality

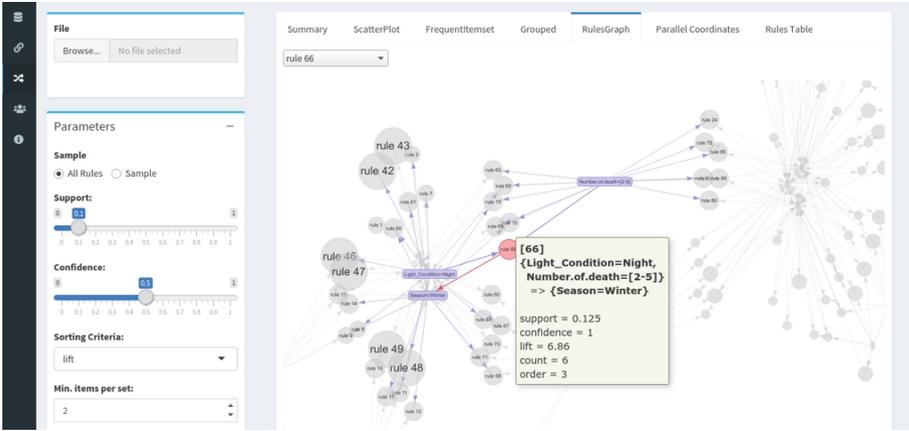


Fig. 2. Association rules mining and visualization interface

measurement. Furthermore, user can move to other tabs for different computations and assign association rules to different categories (The set of categories to which the rules must be assigned to is completely ordered from the best to the worst), the rules in the first category will be the most relevant ones to user preferences (Table 1).

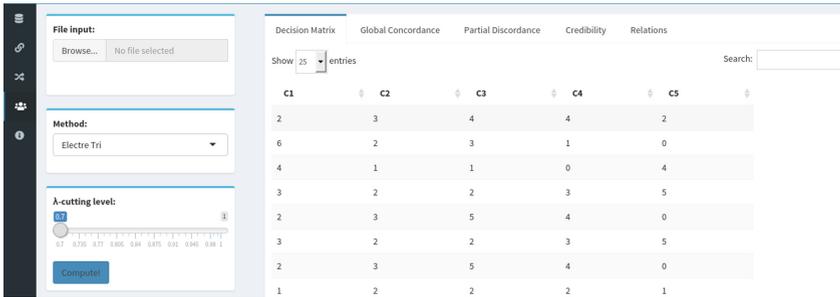


Fig. 3. Select and rank extracted association rules through MCA

Table 1. Profiles defining the category limits

Profile	Support	Confidence	Lift
$b_1$	0.3	0.6	1.0
$b_1$	0.4	0.7	0.9

**Table 2.** Thresholds for ELECTRE TRI method

Thresholds	Support	Confidence	Lift
$weight(k_j)$	0.4	0.7	0.9
$p_j(b_1)$	0.3	0.6	1.0
$q_j(b_1)$	0.4	0.5	0.9
$v_j(b_1)$	0.4	0.8	0.8
$p_j(b_2)$	0.2	0.7	1.0
$q_j(b_2)$	0.5	0.6	0.9
$v_j(b_2)$	0.4	0.5	1.0

Association rules mining algorithms produce a large number of rules. It is, therefore, necessary to choose relevant ones according to the decision-makers' preferences and predefined thresholds  $p_j$ ,  $q_j$ ,  $v_j$ . Based on this study, the integration of multiple criteria analysis within the association rules process performs well and produces relevant rules. After eliminating the rules users are not interested in, eleven significant rules were obtained in this case study (R1, R2, R5, R10, R15, R20, R30, R31, R15, R22, R13).

## 4 Conclusion

This demo paper presents an open-source software for information retrieval by combining data mining especially association rules analysis and multiple criteria analysis. The system starts with data processing by uploading a new CSV dataset or uses a SPARQL query to retrieve data from linked data. Afterward, users obtain association rules extracted by the Apriori algorithm and the relevant rules are chosen using multiple criteria analysis approach. In the future, we aim to extend the system with more data mining algorithms in the context of big data.

**Acknowledgement.** This research is funded by Umeå University in Sweden on federated database research.

## References

1. Ait-Mlouk, A., Agouti, T.: DM-MCDA: a web-based platform for data mining and multiple criteria decision analysis: a case study on road accident. *SoftwareX* **10**, 100323 (2019). <https://doi.org/10.1016/j.softx.2019.100323>
2. Ait-Mlouk, A., Gharnati, F., Agouti, T.: An improved approach for association rule mining using a multi-criteria decision support system: a case study in road safety. *Eur. Transp. Res. Rev.* **9**(3), 40 (2017)
3. Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P.: *Advances in Knowledge Discovery and Data Mining*, pp. 1–34 (1996)
4. FPM (2019). <http://www.borgelt.net/software.html>

5. Le Bras, Y., Meyer, P., Lenca, P., Lallich, S.: A robustness measure of association rules. In: Balcázar, J.L., Bonchi, F., Gionis, A., Sebag, M. (eds.) ECML PKDD 2010. LNCS (LNAI), vol. 6322, pp. 227–242. Springer, Heidelberg (2010). [https://doi.org/10.1007/978-3-642-15883-4\\_15](https://doi.org/10.1007/978-3-642-15883-4_15)
6. Mousseau, V., Figueira, J., Naux, J.P.: Using assignment examples to infer weights for ELECTRE TRI method: some experimental results. *Eur. J. Oper. Res.* **130**(2), 263–275 (2001)
7. R (2019). <https://www.r-project.org/>
8. Roy, B., Vincke, P.: Multicriteria analysis: survey and new directions. *Eur. J. Oper. Res.* **8**(3), 207–218 (1981)
9. Rstudio: R shiny (2019). <https://shiny.rstudio.com/>
10. TANAGRA (2019). <http://eric.univ-lyon2.fr/~ricco/tanagra/fr/tanagra.html>
11. WEKA (2019). <https://www.cs.waikato.ac.nz/~ml/weka/index.html>