

VISUALIZATION OF A SYNTHETIC REPRESENTATION OF ASSOCIATION RULES TO ASSIST EXPERT VALIDATION

Amdouni Hamida¹ and Gammoudi Mohamed Mohsen²

¹PhD Student, FST, University of Tunis ElManar, Tunisia

hamdouni.ecri@gmail.com

²Full Professor, ISAMM, University of Manouba, Tunisia,

mohamed.gammoudi@fst.rnu.tn

ABSTRACT

In order to help the expert to validate association rules, some quality measures are proposed in the literature. We distinguish two categories: objective and subjective measures. The first one depends on a fixed threshold and on data structure from which the rules are extracted. The second one has two subcategories: The first one consists on providing to the expert a tool for rule interactive exploration. In fact, they present these rules in textual form. The second subcategory includes the use of visualization systems to facilitate the task of rules mining. However, this last subcategory assumes that experts have statistical knowledge to interpret and validate association rules. Furthermore, the statistical methods have a lack of semantic representation and could not help the experts during the process of validation. To solve this problem, we propose in this paper a method which visualizes to the experts a synthetic representation of association rules as a formal conceptual graph (FCG). FCG represents his area of interest and allows him to realize the task of rules mining easily due to its semantic richness.

KEYWORDS

Association rules validation, Synthetic representation, Formal conceptual graph, Visualization tool.

1. INTRODUCTION

The Knowledge Discovery in Databases has been widely studied in the literature. In fact, the discovered knowledge can be either a model or a pattern. A model is a representation of all input data, e.g.: the decision tree and neural networks while the pattern is a representation of a subset of data such as association rules which express dependencies or correlations between the attributes (called items). An association rule $R: A \Rightarrow B$ is represented by an antecedent A and a consequent B . A and B are a set of attributes where $A \cap B = \emptyset$

There are two major problems related to the association rules extraction: the extracted rules redundancy, the number and quality of extracted association rules. The last problem is still crucial. In fact, some studies tried to solve it by proposing measures for association rules quality

and some tools of visualization. Two types of measures were proposed: objective measures and subjective measures.

According to the authors in [22], the objective quality measures depend only on the structure of the data from which the rules are extracted. This evaluation technique is based on the following principle: after choosing the measure and the threshold by the expert, only the rules that exceed the threshold are considered.

For subjective measures, the same authors mention that they include two subcategories. The first subcategory consists on providing to the expert a tool for rules interactive exploration. The second includes the use of visualization systems to facilitate the task of rules mining. These systems allow improving the understanding and reduce the time taking by the expert to assimilate the rules [5]. However, these methods require that the expert understands the statistical tools to interpret the results of the visualization [2, 16, 10].

To overcome the last problem, we propose in this paper a method to give a semantically rich synthetic representation of the generated association rules. We show that through visualization rules as a Formal Conceptual Graph which is semantically rich, the expert can validate them while browsing it.

This paper is organized as follows: The section 2 is a state of the art of association rules validity measures. In section 3, we present our contribution in three steps: (i) association rules generation (ii) the definition of a tool for association rules visualization using a Formal Conceptual Graph which represents their synthetic representation in the objective to assist the expert during their validation (iii) Illustration of the use of our system step by step using an example. Finally, we conclude and give some perspectives of our work.

2. RELATED WORKS

In this section, we present the two types of validity measures: objective measures and subjective ones [22, 36, 5, 26].

2.1. Objective quality Measures

Several studies proposed a synthesis of association rules objective measures. To our knowledge, the most referenced one is that of [22]. It has a rating of 20 measurements in 8 criteria. According to [22], the right choice of quality measure based on the nature of the initial data implies a set of extracted rules which are generally satisfactory. In addition, [25] presented 63 objective measures including the 20 measures of [22] with a detailed study according to 12 criteria. We note that [25] have added four criteria related to some kind of sensitivities which are presented as follow:

- Sensitivity to the antecedent that the level of interest' rule depends on the size of the antecedent. Thus, the reduction of the measure increases with the increase of the number of instances associated to the antecedent.
- Sensitivity to the reference situation of equilibrium which means that the number of examples is equal to the number of counter-example. In this case, it is desirable that the measure has a constant value.
- Sensitivity to the reference situation of incompatibility which means that the number of instances associated with the antecedent and the consequent is simultaneously null. In this case, it is desirable that the measure has a constant value.
- Sensitivity to noise is motivated by the fact that in the real world, only few data are not noisy, it is interesting then to consider when choosing measurement.

Thus, we observe firstly that [25] have considered the measure is influenced by the nature of the antecedent and the consequent of a rule. While in [22], we observe that only the consequent influence it. Secondly, the contribution from the consideration of the noise sensitivity is very important because many decisions on support application may be based on noisy data.

Furthermore, during the association rules extraction, these measures must be chosen by the expert while proposing a threshold in the objective of extracting only rules covering the area of his interest. However, three problems may arise. The first one is related to the lack of popularization of the formulas defining these measures. The second problem concerns the arbitrary fixing of the threshold which may not cover the desired area. The third problem is related to the high number of extracted rules. In this case, the expert can find difficulties in validating them, because he does not have a synthetic representation of the covering area of his interest.

To remedy these problems, subjective measures have been proposed, as a post-treatment [26]. In the following section, we present the different types of measures.

2.2. Subjective quality Measures

The post-processing of rules (rules mining) is composed of three stages, the first one is to evaluate and filter rules using objective measures [8, 11]. The second stage is to organize the results in an interactive way and the third stage is to present them in graphical form. The last stage is based in two steps which involve the expert. Furthermore, several tools have been introduced in order to facilitate the task of the expert to extract relevant association rules. We find mainly, the introduction of inductive databases and the associated query languages [17, 27, 20, 6], interactive explorations systems [23, 24, 12] and visualization tools [38, 18, 40, 4, 2, 10]. In the following sections, we present in details these different tools.

2.2.1. Database and induction query language

In 1996, the concept of inductive database was introduced that includes data and the knowledge extracted from these data. To treat this, a query language was proposed: DMQL [17], MINE RULE [27], MSQL [20] and Xmine [6]. Although, the expert does not care about the nature of the information (data or knowledge) when he polls the database, several problems can be encountered. The first one is the need to optimize queries. The second one is the inability of the query language to analyze the data [5].

2.2.2. Interactive explorations Systems

In the literature, several interactive explorations systems have been introduced. In 1994, [21] proposed to represent information in textual form. In 1996, this proposal was implemented in software called TASA to analyze problems in the telecommunications network. This system allows the expert to extract only the interesting rules, specifying the thresholds of some objective measures or using some syntactic constraints that determine the items to be used or not in the rules.

In 2000, [24] introduced an explorer treating a subset of representative rules as a generic-abstract to represent the result of the method of [23]. After selecting a rule, the expert accesses specific rules corresponding to the selected one.

In 2002, [35] presented a generic tool for data analysing on gene of microarrays. This tool can do many operations such as filtering, and guiding navigation of the entire association rules that express relationships between genes in post-processing.

In 2004, [12] proposed an explorer of rules called IRSetNav which has many features. It allows the expert to filter and sort the rules by syntactic constraints or by selected measures and saves the selected rules in the exploration.

We observe that exploration systems have two major limitations. The first limitation is related to the rules displayed in textual form. This makes difficulty of interpretation by the expert especially when the number of rules is high. This number may prevent the expert to choose a few relevant rules. The second constraint is the limited number of objective measures implemented in these systems. In fact, it is useful to choose the measure with respect to input data, since each one has specific characteristics and the right choice leads to good results [22]. To facilitate the expert mission to make the right choices, the researchers found that visualization systems can be one of the most suitable solutions.

2.2.3. Visualization Systems

According to [4], the visualization was originally used in the fields of cartography and statistical graphics. Currently, it has become a discipline used in cognitive psychology, in human-machine interfaces and imaging.

The same authors add that visualization can improve cognition through perceptual capabilities of the human visual system. It provides easy detection of similarities, structures and singularities. It improves, also, the level of memorization and reflection of the expert, which facilitates the generation of synthesis.

Moreover, since 1996, a new technique called "Visual Data Mining" has been introduced by [34] to represent easily exploitable and useful knowledge to experts. This technique includes a set of tools that allow experts to intervene during the two phases of pre-treatment and post-treatment. In pre-treatment, this technique allows the expert to better encircle the application domain by guiding the phase of data mining through the choice of data to be analyzed.

In post-processing through result visualizing, the expert can interact with the system to guide knowledge presentation phase according to their expectations. This enables him to interpret and easily deduce the level of relevance of the extracted knowledge.

In 1997, [37] proposed the first method of visualization of association rules. It consists on presenting the rules in the form of a matrix item-item. These rules can be displayed in several filters: by antecedent, by consequent or by metric (support - confidence). However, as mentioned in [2], this method has the handicap of requiring the expert to make repetitive logical inferences. In addition, it is not possible to visualize the rules according to several criteria simultaneously.

In [38], authors presented association rules visualization system generated from large databases of texts. This system gives matrices formed from extracted rules and associated item, called matrices item-rule. To represent the supports and the confidence of rules, two lines are added to the height of the matrix and the values of these indications are shown as 3D-bars.

In [19] they focused on the visualization of a subset of association rules generated from the original context. These rules include only items associated with a set of items initially chosen in order to have a better study. The system they developed allow the representation of rules in mosaic way. Each one is represented by a rectangle that its area is the support and its confidence is the height.

Similarly, [40] proposed a visualization tool called MIRAGE. It allows visualizing a non-informative generic base from a lattice of frequent closed Item sets which are labeled by the

minimal generators. The non-informative aspect is confirmed by [14], since the values of support and confidence in derived rules cannot be determined with precision. This tool provides to the experts some features. The first is to choose a root node from which it is possible to display or hide its sons. The second is to choose the minimum rules to display. The expert has the possibility to extract the exact or approximate rules by selecting a node in the lattice. These features can be achieved in two ways. The "Normal" mode consists on generating the rule chosen by the expert without any backup. The Mode "Cache" consists on saving the lattice as well as all rules validated by the expert to give him the opportunity to recover them a posteriori.

In 2003, [39] presented a method to visualize rules based on parallel coordinates tool; it is a graphical display of data in an N-dimensional to a two-dimensional space. The proposed method to organize the items into N groups where each group is represented on an axis. The N-axes are parallel (N equal to the maximum order of association rules). Thus, a rule is in the form of a polyline joining the items on the antecedent followed by another polyline connecting the items in the consequent. However, this visualization method is unreadable and cannot be interpreted as the number of rules becomes important.

In [4], authors proposed a visualization system called ARVis (Association Rule Visualization). It consists on extracting and displaying the rules in the form of a 3D landscape based on expert's quality measures. The expert has the possibility to guide the visual system by choosing to display specific or general rules of a given rule. These authors add that this guidance is provided by the development of a methodology called Rule Focusing. It is based on principles of information visualization proposed in [3] for the development of quality indicators and cognitive principles of data processing. It is made in the context of the decision models [28].

In [2] authors are presented a new visualization system of generic bases, called GenVis. It uses meta-knowledge expressed as fuzzy rule bases to exploit a generic evolutionary way. It uses 2D histograms to representation association rules if they are extracted from a dense contexts and 3D one in the case of scattered contexts. It allows also displaying requested derivable rules and semantic links between them in order to improve the interaction between the system and the expert.

Recently, in 2011, authors of [16] introduced an interactive method called "matrix-based grouped visualization" which consists on visualizing groups of rules in the form of a matrix and providing the ability to representation the contents of each group. The rules are grouped based on the items of their antecedent. Each group is represented by a circle in a plane. The first dimension is formed by antecedent' frequent items, whereas the consequent Itemsets forms the second dimension. In this work, the authors considered that the clustering is a solution to the high number of rules. However, the clustering causes a heaviness treatment to their creation. In addition, the need for expert intervention to determine the size of these groups is subjective, since sizing differs from one expert to another in the same domain. Then, the problem of redundant rules is not treated by this method.

Finally, in 2012, authors of [10] presented a method for pre-processing using the dual scaling which is a technique for analyzing discrete multivariate data [30]. This method helps the expert to identify the elements (attribute / item) that can be removed from the database. This will reduce database size without affecting the quality of association rules extracted by the A-priori algorithm. Extracted rules are displayed in 2D form groups using the k-means algorithm. However, k is defined in an empirical way and as they mention in their work, the processing time increases as a function of the number of iterations required for the generation of groups. In addition, the redundancy of rules has not been processed in this method.

After reviewing the literature related to visualization system, we can deduce that the purpose of visualization systems is to facilitate the task of the experts during the mining rules by offering them the opportunity to guide their discovery. However, it is very difficult to assimilate all of them especially if the number of associated attributes is important. In addition, most of these methods require that the expert must understand the statistical tools to interpret the results of the visualization.

Furthermore, few systems offer to the experts minimum rules (generic) [40, 2] and give them the opportunity to see a derivable rules for selected generic rules. To our knowledge, the integration of semantics by exploiting the wealth of items and links between association rules still very important. Thus, we achieve this section by the presentation (see table 1) of a classification summary including visualization systems presented according to three criteria: "Display type of rules," "level of interaction" (with the expert) and "Bases":

- "Type display rules": matrix, graph 2D or 3D, Galois lattices, etc.
- "Level of interaction" denoted NI: defines the level of interaction with the expert, high or low.
- "Bases": includes two subcategories. The first one is entitled "G" and Indicates whether the system displays the generic bases. While the second one is "Infor" Which indicates if visualized generic bases are informative or not.

Table 1. Comparison between visualization systems

Systems	Display Type	I	Bases	
			G.	Info
[37]	Matrix item-to-item	L	No	-
[38]	Matrix item-to-rule	L	No	-
[19]	Representation on mosaic rules	L	No	-
MIRAGE [40]	Galois lattice	L	Yes	No
[39]	2D with parallel coordinates	L	No	-
ARVis [4]	3D landscape	L	No	-
GenVis [2]	Histogram 2D ; Histogram3D	H	Yes	Yes
[16]	2D Scatter	H	No	-
[10]	2D Scatter	H	No	-

3. MOTIVATION AND CONTRIBUTION

In this section, we present our visualization method to help the expert evaluating the association rules. We will show that the visualized rules are generic satisfying the thresholds fixed by the expert. They are represented as a rich semantically graph which is based on Formal Concept Analysis approach (FCA) [14]. The FCA has been used in several areas such as: natural language processing [7, 9], object-oriented databases, information retrieval, data mining, web service, etc. Indeed, in the field of language processing, the extraction of semantic relations by the FCA in linguistic applications is presented according to [29] in [31, 32]. Furthermore, the formalization of the Worldnet ontology is done by FCA in [33]. The Galois lattice structure was also used to represent the class hierarchy or inheritance graph generation in object-oriented approach [15, 13]. The inheritance graph hierarchy is represented by a sub-graph of a Galois lattice called Formal Conceptual Graph which covers all formal concepts of the lattice.

In order to be clearer, we recall some basic notions of this approach such as: formal context, Galois connexion, Formal concept, Partial order relation, Formal Conceptual Graph (FCG).

- **Formal context:** a triplet $k = (O, I, R)$, O set of objects, I set of items (attributes), $R \subseteq O \times I$ binary relation between objects and items. If $(o, i) \in R$ it means that an object o is related to an item i .
- **Galois connection:** let a formal context (O, I, R) , $A \in O$, $B \in I$ and R binary relation between O and I . $A \in E$, $B \in F$ et R a binary relation ; the antitone Galois connection are :

$$f(A) = \{m / \forall g \in A ; R(g,m) = 1\}$$

$$h(B) = \{g / \forall m \in B ; R(g,m) = 1\}$$
- **Formal concept:** let a formal context (O, I, R) , $A \in O$, $B \in I$: if $A = h(B)$ and $B = f(A)$, (A, B) is said a formal concept. A and B are called respectively the extent and the intent of (A, B) .
- **Partial order relation between concepts:** Let (A_1, B_1) and (A_2, B_2) two formal concepts. $(A_1, B_1) \ll (A_2, B_2)$ if and if $A_1 \subseteq A_2 \Leftrightarrow B_2 \subseteq B_1$.
The concepts in FCA of a given formal context are ordered by sub-concept and super-concept relation.
- **Galois lattice:** The ordered set of formal concepts extracted from a formal context is Galois Lattice. Its representation is done by a Hass Diagram.
- **Formal Conceptual Graph (FCG) [13]:** is a subset of formal concepts of a Galois lattice which covers the formal context from which it is extracted.

Formal Conceptual Graph (FCG) allows us the reduction of Galois lattice nodes preserving its properties. Thus, we exploit this structure to extract association rule semantic by visualizing easily a synthetic representation of these rules according to the expert interest. We will show in the following section how our method is interactive since it offers to the expert (i) the possibility of displaying derivable rules from the generic one, (ii) the possibility of validating or pruning extracted rules and (iii) the possibility to save the activities made by the expert.

The principle of our method consists on two steps: the first one is to extract generic rules entitled GenCondClose. The second one is to give a synthetic representation of the items related to the association rules which will be visualized.

In the following, we detail these two steps.

3.1. Step of generating rules: GenCondClose

This step consists on applying the algorithm of CondClose [1] to a binary context in order to build a Formal Conceptual Graph (FCG) with each formal concept is labeled by its generators. The generated FCG is used to extract the generic exact and approximate association rules. The general architecture of this step is shown in (Figure 1).

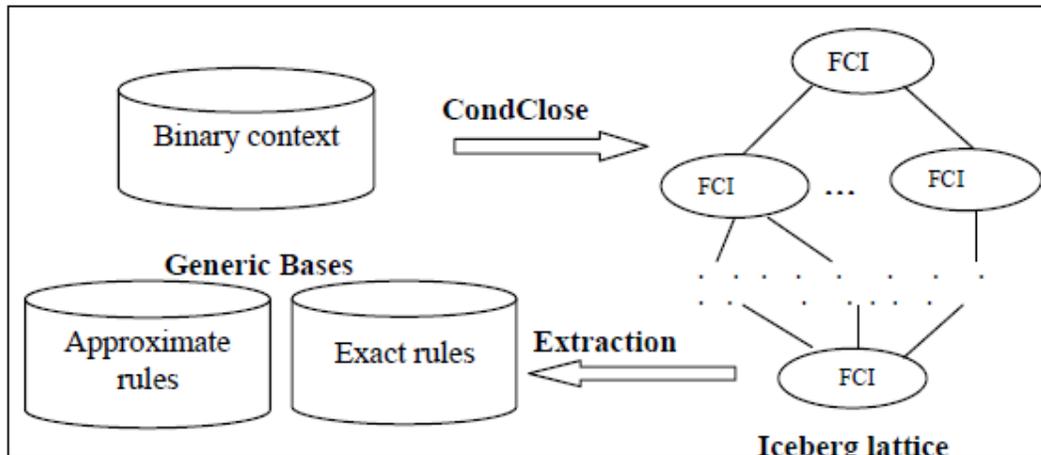


Figure 1. The general architecture of GenCondClose

After obtaining generic bases of association rules, the expert can validate them. To achieve this task, we propose to the expert a tool called SVEA_ARV for visualizing a synthetic representation of the minimum rules.

3.2. System of Visualization and Expert Assistance for Association rules Validation (SVEA_ARV)

The second step of our method is the design and the development of a SVEA_ARV tool. The general architecture of this tool is shown in (Figure 2).

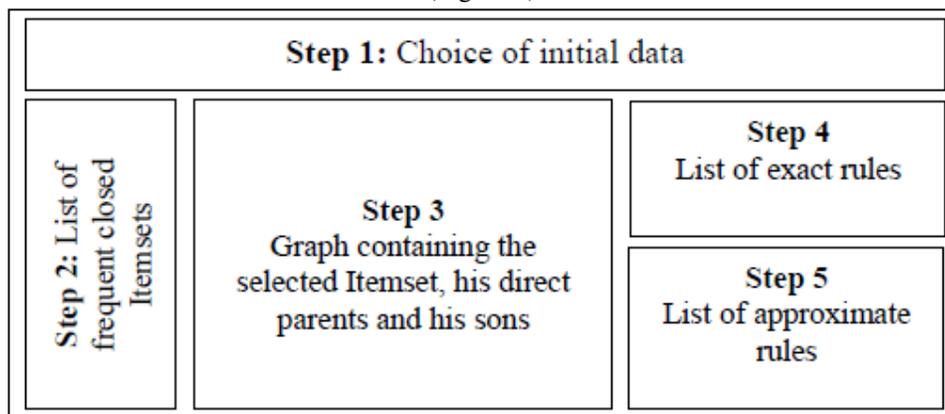


Figure 2. The general architecture of SVEA_ARV

In the following, we explain these five steps mentioned in Figure 2:

- **Step 1:** In the beginning, the expert specifies the initial context, the minimum value of the support, the choice of measures validation (confidence, lift, conviction, etc.) and the associated threshold. If the expert has already selected these values in a previous session, the system retrieves the associated FCG, the generated rules and those that were validated by the expert. Otherwise, the system generates the FCG and labels each formal concept with the minimal generators of corresponding closed frequent Itemsets.
- **Step 2:** The system displays a list of closed Itemsets associated to the chosen context.
- **Step 3:** Having the list of frequent closed Itemsets, the expert can choose an Itemset to display by generalization or specialization its direct fathers and children's as a FCG. The

nodes of this graph are a frequent closed itemset. It shows a synthetic representation of Itemsets that are connected by semantic links expressing specialization/ generalization between formal concepts. In addition, each semantic link is labeled by the number of approximate rules based on the threshold chosen by the expert. Thus, the expert has the possibility to navigate into this graph.

- **Step 4:** Following the selection of a frequent closed Itemset in step 2, the set of the generic exact rules associated are displayed.
- **Step 5:** for approximate rules presentation, the expert selects the link between two chosen frequent closed itemset.

After visualizing association rules (approximates and exacts) by our system, He could validate them one by one. Furthermore, he has the possibility to display the list of generated rules by the system and those that were already validated by him.

4. THE USE OF OUR SYSTEM STEP BY STEP

To illustrate our method, we use a Car Evaluation Data Set¹ (see Table 2). The values of their attributes have a nominal type. To apply the first step of our method, we use a binarization method implemented in Weka² Software to obtain a formal context $K = (O, I, R)$. O is the set of instances (or objects), I is the set of attributes and R is a binary relation between O and I.

Table 2. Description of the Car Evaluation Data Set

Number of instances	Number of attributes	Number of binary attributes
1728	7	25

Using our system SVEA_ARV, the expert chooses the formal context K, the support's threshold, the confidence as a measure of validation and the threshold associated which are in this example, respectively 0.1 and 0.2 (see Figure 3). After this step (Phase 1 shown in Figure 2), the system generates the corresponding FCG and extracts the association rules.

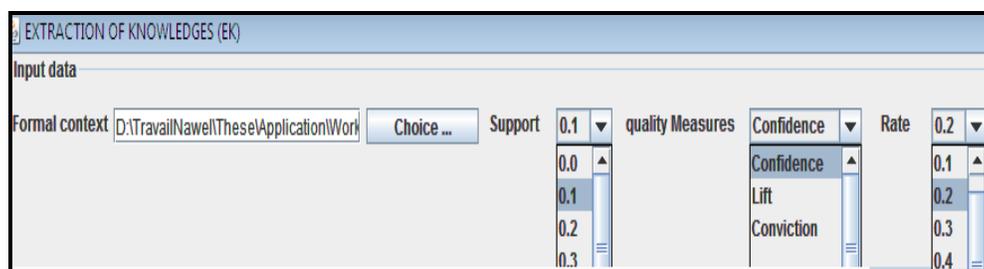


Figure 3. Choice of Input Data

The system displays a list of Frequent Closed Itemset (FCI). The expert could select one of them according into his interest. A synthetic representation represented by a FCG associated to this FCI is displayed as well as the exact association rules. In the graph, the selected FCI is represented by a red rectangle. Its intent is $\{\text{lug_boot}=\text{big}, \text{class}=\text{unacc}\}$. The children and the parent of the select FCI are presented with blue rectangles (see Figure 4).

¹ Available at <http://archive.ics.uci.edu/ml/datasets/Car+Evaluation>

² Available at <http://www.cs.waikato.ac.nz/ml/Weka>

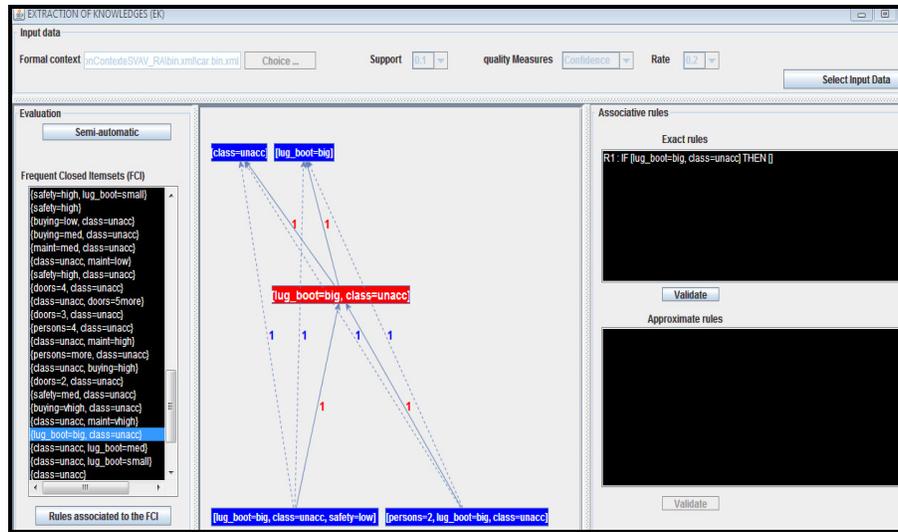


Figure 4. Interface of our system SVEA_ARV

The expert can validate them. Knowing that each edge of the graph is labeled by the number of approximate association rules and thanks to the richness of the graph, the system offers to the expert the possibility to consult the list of approximate association rules by clicking on an edge between two FCI in the graph (see Figure 5).

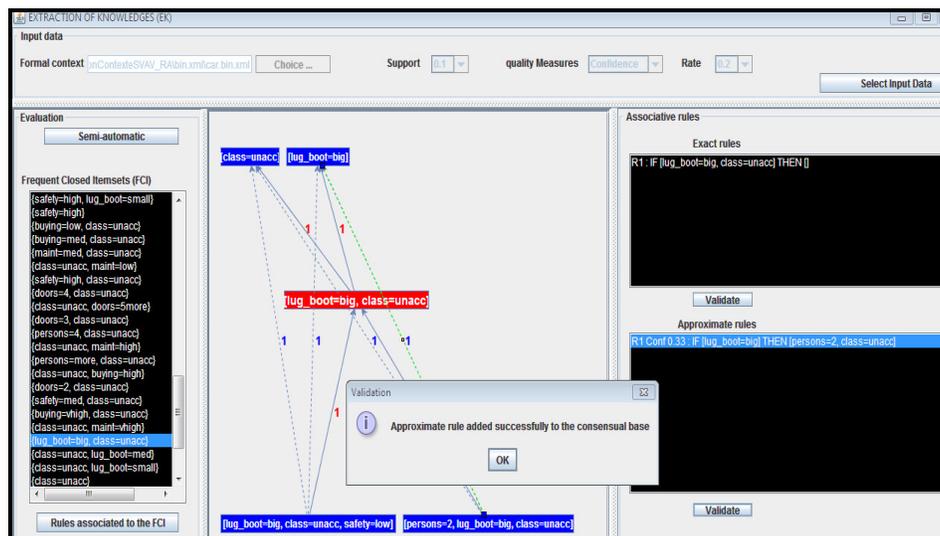


Figure 5. Validation of association rules

He could also navigate in this graph for choosing any FCI which is a children or a parent of the current FCI. The graph associated to the chosen one is displayed.

Figure 6 represent the graph and the exact rules associated to one of the current FCI children. The child has the intent {persons=2, lug_boot=big, class=unacc}.

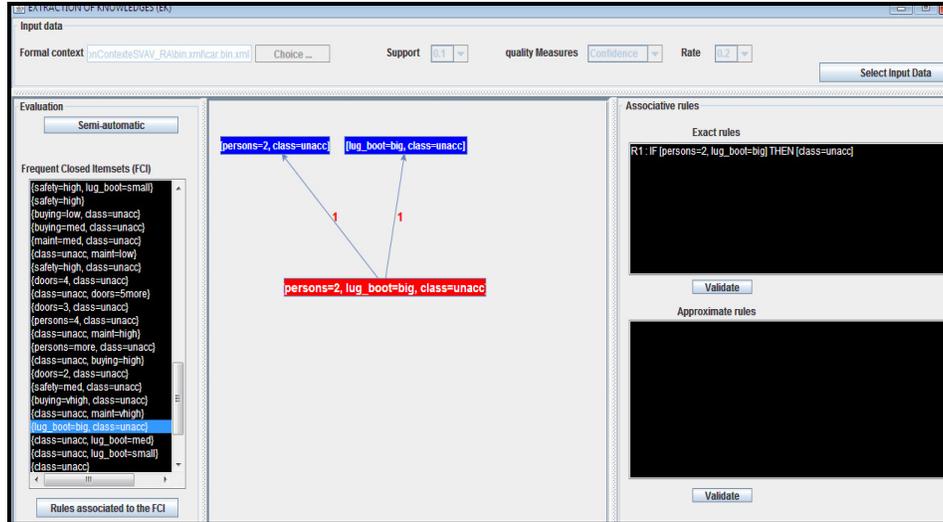


Figure 6. Graph associated to one of the current FCI children

Furthermore, he could visualize the list of generated rules by the system and those that were validated by him. The system provides to the expert the precision of exact and approximate association rules in the objective to assist him during the phase of validation (see Figure 7).

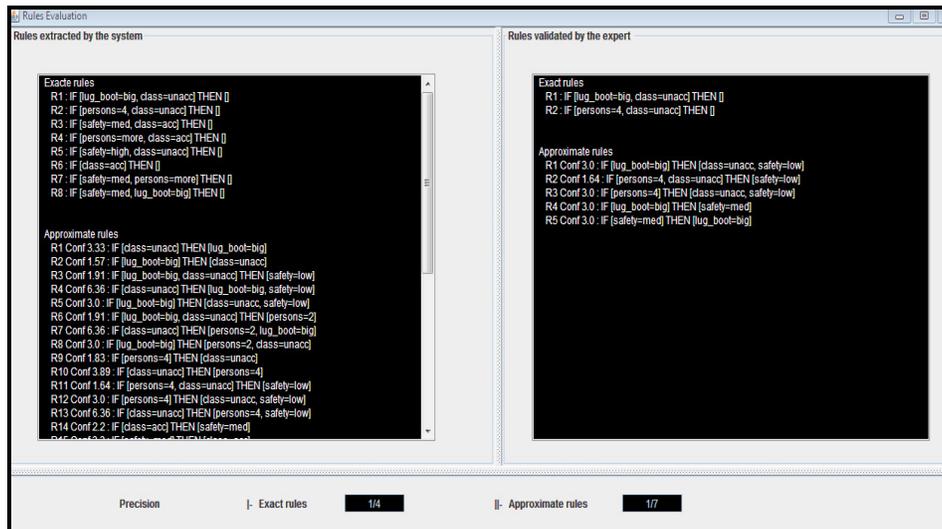


Figure 7. Rules extracted by the system vs rules validated by the expert.

5. CONCLUSION AND PERSPECTIVES

In this paper we presented a state of the art related to quality measures and some visualization systems used in the rule mining task. We observe that the lack of a semantic representation of association rules for the expert during the step of validation stills crucial. In order to assist the expert during the validation of association rules task, we propose in this paper a new method for synthetic representation visualization based on the notion of Formal Conceptual Graph. This method allows to the expert to obtain a synthetic representation related to his area of interest thanks to the visualization of semantic links between nodes of FCG. Thus, he could validate the

association rules extracted in an easy way. As a future work, some experimentations using a representative datasets will be done. We plane also to extend our method for semi-automatic evaluation of extracted association rules by submitting them as a classifier in some rule-based classification systems.

REFERENCES

- [1] Amdouni, H. & Gammoudi, M. M. (2011) “CondClose: A new algorithm of association rules extraction”, *IJCSI (International Journal of Computer Science Issues)*, Volume 8, Issue 4.
- [2] Ben Yahia, S. & Mephu Nguifo, E. (2006) “Visualisation des règles associations : vers une approche méta-cognitive”, Dans *Actes conférences INFORSID*, Hammamet, pp 735-750.
- [3] Bertin, J. (1967) “Sémiologie graphique”, Gauthier-Villars, (3e édition en 1999 aux Editions de l’Ecole des Hautes Etudes en Sciences Sociales).
- [4] Blanchard, J. (2005) “Un système de visualisation pour l’extraction, l’évaluation, et l’exploration interactives des règles d’association”, Thèse de doctorat à l’Ecole Polytechnique de l’Université de Nantes, soutenue le 24 novembre.
- [5] Blanchard, J., Guillet, F. & Briand, H. (2007) “Interactive visual exploration of association rules with rule-focusing methodology”, *Knowledge and Information Systems*, vol. 13, num. 1, p. 43-75, Springer-Verlag.
- [6] Braga, D., Campi, A., Klemettinen, M. & Lanzin, P. L. (2002) “Mining association rules from XML data”, in *Proceedings of the fourth international conference on data warehousing and knowledge discovery (DaWaK 2002)*, Lecture Notes in Computer Science, vol. 2454, , p. 21–30, Springer-Verlag.
- [7] Boutari, A.M., Carpineto, C. & Nicolussi, R. (2010) “Evaluating term concept association measures for short text expansion: two case studies of classification and clustering”, in: *Proceedings of the Seventh International Conference on Concept Lattices and their Applications (CLA 2010)*.
- [8] Briand, H., Sebag, M., Gras, R. & Guillet, F. (2004) “Mesures de qualité pour la fouille de données”, Cepaduès Editions, numéro spécial de la Revue des Nouvelles Technologies de l’Information.
- [9] Falk I. & Gardent C. (2011) “Combining formal concept analysis and translation to assign frames and thematic grids to French verbs”, in: Amedeo Napoli, V.V. (ed.) *International Conference on Concept Lattices and Their Applications CLA 2011*. pp. 223-228. INRIA Nancy - Grand Est and LORIA.
- [10] Fernandes, L.A.F. & Garcia, A.C.B. (2012) “Association Rule Visualization and Pruning through Response-Style Data Organization and Clustering”, J. Pavon et al. (Eds.): *IBERAMIA 2012*, LNAI 7637, pp. 71–80, 2012. Springer-Verlag Berlin Heidelberg.
- [11] Freitas, A. A. (1998) “On objective measures of rule surprisingness”, in *Proceedings of the second European conference on principles of data mining and knowledge discovery (PKDD’98)* (J. Zytkow& M. Quafafou, éd.), Lecture Notes in Artificial Intelligence, vol. 1510, p. 1–9, Springer-Verlag.
- [12] Fule, P. & Roddick, J. F. (2004) “Experiences in building a tool for navigating association rule result sets”, in *CRPIT’04: Proceedings of the second Australasian workshop on information security, data mining, web intelligence, and software internationalization* (J. Hogan, P. Montague, M. Purvis & C. Stekete, éd.), Australian Computer Society, Inc., p. 103–108.
- [13] Gammoudi, M. M., Mendes, J. D. & Pinto, W. S. (1997) “Inheritance Graph Hierarchy Construction Using Rectangular Decomposition of a Binary Relation and Designer Feedback”, *ER 1997*, 285-298.
- [14] Gasmi, G., Ben Yahia, S., Nguifo, E. M. & Slimani, Y. (2005) “IGB: A new informative generic base of association rules”, *Proceedings of the Intl. Ninth Paci_c-Asia Conference on Knowledge Data Discovery (PAKDD’05)*, LNAI 3518, Hanoi, Vietnam, Springer-Verlag, p. 81-90.
- [15] Godin, R., Mili, H., Arfi, A., Mineau, G. W. & Missaoui, R. (1994) “A Tool for Building and Evaluating Class Hierarchies Based on a Concept Formation Approach”, *Proceedings of the OOPSLA 94 Workshop on Artificial Intelligence for Object-Oriented Software Engineering*.
- [16] Hahsler, M. & Chelluboina, S. (2011) “Visualizing association rules in hierarchical groups”, *42nd Symposium on the Interface: Statistical, Machine Learning and Visualization Algorithms (Interface 2011)*, The Interface Foundation of North America.
- [17] Han, J., Fu, Y., Wang, W., Koperski, K. & Zaiane, O. (1996) “DMQL: a data mining query language for relational databases”, in *Proceedings of the 1996 SIGMOD workshop on research issues on data mining and knowledge discovery (DMKD)*.

- [18] Han, J. & Cercone, N. (2000) "Ruleviz: A model for visualizing knowledge discovery process", Proceedings of 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Boston, MA, USA, p. 244-253.
- [19] Hofmann, H., Siebes, A.P.J.M. & Wilhelm, A.F.X. (2000) "Visualizing association rules with interactive mosaic plots", in 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 227-235, Boston, MA, USA.
- [20] Imielinski, T. & Virmani, A. (1999) "MSQL: a query language for database mining", Data Mining and Knowledge Discovery 3, no. 4, p. 373-408.
- [21] Klemettinen, M., Mannila, H., Ronkainen, P., Toivonen, H. & Verkamo, A. I. (1994) "Finding interesting rules from large sets of discovered association rules", in Proceedings of the third international conference on information and knowledge management (CIKM 1994), ACM Press, p. 401-407.
- [22] Lenca, P., Meyer, P. & Vaillant, B. (2004) "Evaluation et analyse multicritère des mesures de qualité des règles d'associations", National Journal of Information Technologies (RNTI), France, pp.219-246.
- [23] Liu, B., Hsu, W. & Ma, Y. (1999) "Pruning and summarizing the discovered associations", in KDD'99: Proceedings of the fifth ACM SIGKDD international conference on knowledge discovery and data mining, ACM Press, p. 125-134.
- [24] Ma, Y., Liu, B. & Wong, C. K. (2000) "Web for data mining: organizing and interpreting the discovered rules using the web", SIGKDD Explorations 2, no. 1, p. 16-23.
- [25] Maddouri M. & Gammoudi J. (2007) "On Semantic Properties of Interestingness Measures for Extracting Rules from Data", B. Beliczynski et al. (Eds.): ICANNGA 2007, Part I, LNCS 4431, pp. 148-158, Springer-Verlag Berlin Heidelberg.
- [26] Marinica, C. & Guillet, F. (2009) "Improving Post-Mining of Association Rules with Ontologies", Proceeding of the XIIIth International Conference Applied Stochastic Models and Data Analysis (ASMDA-2009), p. 76-80.
- [27] Meo, R., Psaila, G. & Ceri, S. (1998) "An extension to SQL for mining association rules", Data Mining and Knowledge Discovery 2, no. 2, p. 195-224.
- [28] Montgomery, H. (1983) "Decision rules and the search for a dominance structure: towards a process model of decision making", in Analysing and aiding decision processes (P.Humphreys, O. Svenson & A. Vari, eds.), Elsevier Science Publishers, p. 343-369.
- [29] Meštrović, A. (2012) "Semantic Matching Using Concept Lattice", Concept Discovery in Unstructured Data / Ignatov, Dmitry I.; Kuznetsov, Sergei O.; Poelmans, Jonas (ed). - Leuven, Belgija: Katholieke Universiteit Leuven, p.p. 49-58.
- [30] Nishisato, S. (1993) "Elements of Dual Scaling – An Introduction to Practical Data Analysis", Psychology Press.
- [31] Old, J. (2006) "Homograph disambiguation using formal concept analysis", in: Missaoui, R., Schmid, J. (eds.) 4th International Conference on Formal Concept Analysis, Lecture Notes in Computer Science. vol. 3874, pp. 221-232. Springer-Verlag.
- [32] Potemkin, S. (2011) "Concept lattice implementation in semantic structuring of adjectives", in: Ignatov, D., Kuznetsov, S., Poelmans, J. (eds.) Concept Discovery in Unstructured Data. pp. 63-70.
- [33] Priss, U.E. (1998) "The formalization of wordnet by methods of relational concept analysis", in: WordNet: An Electronic Lexical Database and Some of its Applications. pp. 179-196. MIT Press.
- [34] Shneiderman, B. (1996) "The eyes have it: A task by data type taxonomy for information visualization", PRESS I. C. S., Ed., Proceedings IEEE Symposium on Visual Languages, Boulder, Colorado, 336-343.
- [35] Tuzhilin, A. & Adomavicius, G. (2002) "Handling very large numbers of association rules in the analysis of microarray data". In KDD'02: Proceedings of the eighth ACM SIGKDD international conference on knowledge discovery and data mining, ACM Press, p. 396-404.
- [36] Vaillant, B., Menou, S., Moga, S., Lenca, P. & Lallich S. (2007) "Qualité des règles d'association : étude de données d'entreprise", in Atelier Data mining dans la banque, l'assurance et la finance (associé à la conférence Extraction et Gestion des Connaissances (2007)), pp.45-54, Namur, Belgique.
- [37] Wong, P. C. & Bergeron, R. D. (1997) "30 Years of Multidimensional Multivariate Visualization", in Gregory M. Nielson, Hans Hagan, and Heinrich Muller, editors, Scientific Visualization - Overviews, Methodologies and Techniques, pages 3-33, Los Alamitos, CA, IEEE Computer Society Press.

- [38] Wong, P., Whitney, P. & Thomas, J. (1999) “Visualizing Association Rules for Text Mining”, Proceedings of the IEEE Symposium on Information Visualization, San Francisco, CA, October 1999, p. 120.123.
- [39] Yang, L. (2003) “Visualizing Frequent Itemsets, Association Rules and Sequential Patterns in Parallel Coordinates”, in: Kumar, V., Gavrilova, M.L., Tan, C.J.K., L’Ecuyer, P. (eds.) ICCSA 2003. LNCS, vol. 2667, pp. 21–30. Springer, Heidelberg.
- [40] Zaki M. & Phoophakdee B. (2003) “MIRAGE: A framework for mining, exploring and visualizing minimal association rules”, Technical report, July 2003, Rensselaer Polytechnic Institute, Computer Sciences Department, USA.

Authors

Hamida Amdouni is currently an Assistant at Manouba University, ESENE. She received her Master degree in Computer Science at FST-Tunisia in 2005. Now, she prepares her PhD at the Faculty of Sciences of Tunis and she is supervised by Pr. Gammoudi. Her main research contributions concern: data mining, Formal Concept Analysis (FCA) and Customer Relation Management. She is member of SCO-ECRI in Research Laboratory RIADI



Mohamed Mohsen Gammoudi is currently a full Professor at ISAM, University of Manouba, Tunisia. He is responsible of SCO-ECRI team in Research Laboratory RIADI. He obtained his habilitation to Supervise research in 2005 at the Faculty of Sciences of Tunis. He got his PhD in September 1993 in Sophia Antipolis Laboratory I3S/CNRS. Professor Gammoudi’s professional work experience began in 1992 when he was assigned as an assistant at the Technical University of Nice. Then he was hired as a visiting professor between 1993 and 1997 at Federal University of Maranhao, Brazil. Since, he supervised several PhD and master thesis.

