ATLANTIS PRESS

Atlantis Studies in Uncertainty Modelling, volume 1

11th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT 2019)

A comparative analysis of tools for visualizing association rules: A proposal for visualising fuzzy association rules

Carlos Fernandez-Basso[a] and M. Dolores Ruiz[a] and Miguel Delgado[a] and Maria J. Martin-Bautista[a]
[a]Computer Science and A.I. Department, CITIC-UGR, University of Granada, Spain
{cjferba,mdruiz,mdelgado,mbautis}@decsai.ugr.es

Abstract

Discovering new trends and co-occurrences in massive data is a key step when analysing social media, data coming from sensors, etc. Although, nowadays Data Mining is very useful and widely used for the industry, business and government, the main problem of application of machine learning or data mining in other fields is the interpretability and complexity of obtained results for non-expert users in computer or data science. For this reason in the KDD process one of the most important phases is the interpretation and evaluation.
In the case of association rules is essential that results are interpretable for experts. One of the most useful tools for this goal is the visualization, because it clarifies the interpretation of results, being easier to understand in order to take a decision or explaining the behaviour of data.
Keywords: Association Rules, Fuzzy Association Rules, Visualization, Frequent itemsets mining, Data Mining

1 Introduction

Many organizations, buildings or social networks generate a large amount of data on a daily basis, such as the tweets about a particular topic, sales in Amazon, or sensors in large buildings such as for instance an airport. All of these kinds of data can be analysed by means of Data Mining techniques. Association rule mining [2] is one of the major techniques to detect and extract useful information from huge datasets such as sensors [8], recommendation systems [19] or social networks [14].
In this kind of algorithms, frequent itemset and association rule mining, are specially interesting to search relationships between events, courses and students, as for instance in MOOC courses [5]. Additionally, frequent itemsets can then be employed for discovering other types of patterns such as sequential patterns[20], gradual dependencies [17] or exception and anomalous rules[12] to discover meaningful and different patterns amongst them.
However, the nature of the data can be diverse and can be described in numerical, categorical, etc. In case of numerical variables a first approximation could be to categorise them so that, for example the value of a temperature sensor may be given by a range to which it belongs as for instance [18º C, 24º C]. However, based on how the interval is defined, the result can change. To avoid this, the use of linguistic labels such as "Warm" represented by a fuzzy set is a good option to represent the temperature of a room, having at the same time a meaningful semantic to the user.
As a consequence of the above, fuzzy association rules are an useful and important technique for extracting knowledge. Nevertheless the results obtained when applying association rule mining are, sometimes, very difficult to understand due to the large amount of rules and itemsets generated. In the literature we can find studies on revisions on tools to visualize association rule [9]. Besides, these kinds of visualizations do not allow the use of fuzzy association rules.
In this paper, we first review existing visualizations tools for frequent itemsets and association rules in the crisp and the fuzzy case, and afterwards we present several proposals to improve the visualization techniques presented when displaying fuzzy association rules.
This paper is structured as follows: in Section 2 we explain the basic concepts about association rules and fuzzy association rules. In Section 3, we study the different tools in the literature to visualize association rules. In Section 4, we propose a new way to visualize fuzzy association rules. Finally, we present some conclusions and future works.

2 Preliminary concepts

This section introduces all the major foundational concepts required to understand the concepts around crisp and fuzzy association rules presented in this paper.

2.1 Association rules

Association rules were formally defined for the first time by Agrawal et al. [1].

The problem consists in discovering implications of the form $A \to B$ where A, B are subsets of items from $I = \{i_1, i_2, ..., i_m\}$ fulfilling that $A \cap B = \emptyset$ in a database formed by a set of n transactions $D = \{t_1, t_2, ..., t_n\}$ each of them containing subsets of items from I. A is usually referred as the antecedent and B as the consequent of the rule.
The most commonly used measures to extract frequent itemsets and association rules are:

- The *support* [6] is the measure of the frequency with which an item appears in the database. In general, the most interesting association rules are those with a high support value.

$$Supp_D(X) = \frac{|t_i \in D : X \subseteq t_i|}{|D|} \tag{1}$$

- Given the itemsets X and Y, and the database D, the confidence of the rule $X \to Y$ [6], represented as $Conf_D(X \to Y)$, is the conditional probability of Y appearing in those transactions in D that contain X.

$$Conf_D(X \to Y) = \frac{Supp_D(X \cup Y)}{Supp_D(X)} \tag{2}$$

The problem of uncovering association rules is usually developed in two steps:

- Step 1: Finding all the itemsets above the minimum support threshold. These itemsets are known as frequent itemsets.
- Step 2: Using the frequent itemsets, association rules are discovered by imposing a minimum threshold for an assessment measure such as confidence.

2.2 Fuzzy Association rules

To deal with uncertain and imprecise data we introduce the concept of fuzzy transaction and fuzzy association rule defined in [7, 11].

Definition 1 *Let I be a set of items. A fuzzy transaction, t, is a non-empty fuzzy subset of I in which the membership degree of an item $i \in I$ in t is represented by a number in the range [0, 1] and denoted by $t(i)$.*

By this definition a crisp transaction is a special case of fuzzy transaction. We denote by $\tilde{D}$ a database consisting in a set of fuzzy transactions.

Definition 2 *Let $A \subseteq I$ be an itemset, i.e. a subset of items in I, the degree of membership of A in a fuzzy transaction $t \in \tilde{D}$ is defined as the minimum of the membership degree of all its items:*

$$t(A) = \min_{i \in A} t(i). \tag{3}$$

Definition 3 *Let $A, B \subseteq I$ be itemsets in the fuzzy database $\tilde{D}$. Then, a fuzzy association rule $A \to B$ is satisfied in $\tilde{D}$ if and only if $t(A) \le t(B)\ \forall t \in \tilde{D}$, that is, the degree of satisfiability of B in $\tilde{D}$ is greater than or equal to the degree of satisfiability of A for all fuzzy transactions t in in $\tilde{D}$.*

The support and confidence measures are then defined using a semantic approach based on the evaluation of quantified sentences as proposed in [7, 11] using the GD-method [11] and the quantifier $Q_M(x) = x$,which represents the quantifier *"the majority"*.

Definition 4 *The support of a fuzzy rule $A \to B$ is defined as:*

$$FSupp(A \to B) = \sum_{\alpha_i \in \Lambda} (\alpha_i - \alpha_{i+1}) \frac{|\{t \in \tilde{D} : t(A) \ge \alpha_i \text{ and } t(B) \ge \alpha_i\}|}{|\tilde{D}|} \tag{4}$$

where $\Lambda = \{\alpha_1, \alpha_2, \ldots, \alpha_p\}$ with $\alpha_i > \alpha_{i+1}$ and $\alpha_{p+1} = 0$ is a set of α-cuts.

Definition 5 *The confidence of a fuzzy rule $A \to B$ is defined as:*

$$FConf(A \to B) = \sum_{\alpha_i \in \Lambda} (\alpha_i - \alpha_{i+1}) \frac{|\{t \in \tilde{D} : t(A) \ge \alpha_i \text{ and } t(B) \ge \alpha_i\}|}{|\{t \in \tilde{D} : t(A) \ge \alpha_i\}|} \tag{5}$$

where $\Lambda = \{\alpha_1, \alpha_2, \ldots, \alpha_p\}$ with $\alpha_i > \alpha_{i+1}$ and $\alpha_{p+1} = 0$ is a set of α-cuts.

3 Association rules visualization. Previous works and comparison

In its first subsection, basic concepts and terminology from visualization and data structure models, and its second subsection is devoted to the previous works and applications developed till the date for visualizing association rules. Afterwards, we also explain the principal advantages and drawback of these tools and

we analyse the interpretability of each of them.
For visualizing association rules it can be used different ways. However we have to take into account the structure of obtained results and what we want to see in the graphic representation in order to help help the user to discover hidden insights or interpreting the results. These techniques can be classified in different subgroups so that it is clarified the different ways and structures used to represent association rules.
Consequently, we analysed the tools for visualizing association rules attending to different categories, such as data structure or kind of visualization. To sum up, in the next sections, we divide the available visualization tools in three big groups: table representation, visualization via charts like a matrix or scatter plots and graph-based methods.

3.1 Table visualization

Firstly, one simple method for visualizing association rules is to use a table. Although it is a simple representation, it is useful for small sets of rules since it enables the expert to observed the sets of rules ordered by antecedent, consequent take into account their measurements, such as support, confidence or lift for example. In Figure 1, we can see an example


Viewer Zoom

Show 10 entries   Search:

	LHS	RHS	support	confidence	lift	count
	All	All	All	All	All	All
[53]	{Instant food products,soda}	{hamburger meat}	0.001	0.632	18.996	12.000
[37]	{soda,popcorn}	{salty snack}	0.001	0.632	16.698	12.000
[444]	{flour,baking powder}	{sugar}	0.001	0.556	16.408	10.000
[327]	{ham,processed cheese}	{white bread}	0.002	0.633	15.045	19.000
		{hamburger				

Showing 1 to 10 of 5,668 entries   Previous 1 2 3 4 5 ... 567 Next

Figure 1: Table visualization

of table visualization for association rules, using the ArulesViz package. In this case, the method used allows to display the rules and rank them according to their antecedent and consequent values.

Advantages

The main advantage of this visualization is its simplicity, it is very easy to add and remove new assessment measures and it is very intuitive for ranking by their value. In addition, it is easily adaptable for fuzzy association rules.

Disadvantages

The greatest disadvantage of table representation is that it is not suitable for interpreting and understanding big set of rules, because it is very difficult to observe all rules and relationships between itemsets at the same time.

3.2 Matrix-based visualization

In a second place, we have grouped some techniques similar to charts. In these cases the visualization approaches use different charts for representing the association rules. In Figure 2 we can see the different ways for visualizing rules using scatter plot [4] or a Matrix in 2D [16] or 3D.
In these techniques the information to be visualized can be organized in two ways. Firstly, the rules can be displayed by antecedent and consequent in one of the axis (Figure 2a, 2c). The interest measure can be shown at the intersection of the antecedent and the consequent of rules. Another way is to display only the measures (e.g. support and confidence) in the axis and to use an interactive box [21] to know which rules corresponds to that pair of values. An example is shown in Figure 2b.
In Figure 2a we can see an example of the matrix technique where we distribute the antecedents and consequents in the axes of the graph. This tool has an interactive component and by positioning ourselves in the intersection that represents a rule we can see the different measures stored for that rule.
Finally, Scatter plot allows us to visualize the rules by two measures of the rule that are represented in the axes. In this case we can determine which rules have similar values and study the set of general rules.

Advantages

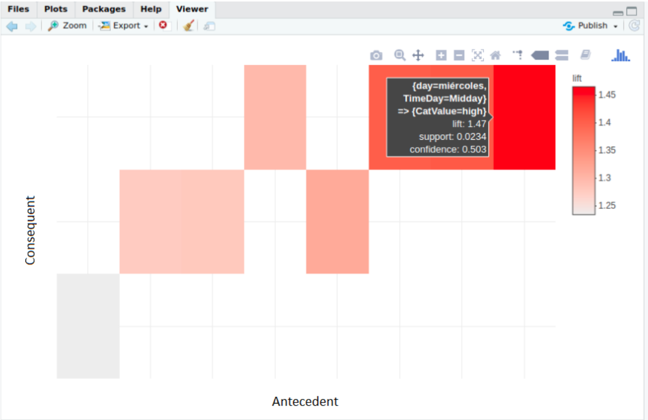
One of the main advantages is that we can easily know which itemsets are more frequent, or which itemsets appear in more rules. In addition, these methods display this information joint with the interest measures for improving the comprehension of results. These techniques can be turned in different ways using different colours, shapes or lengths.

Disadvantages

However, this kind of displaying have some drawbacks. Firstly, it is very difficult to see the relationships between items and rules. In addition, we can not display results obtained from large datasets because it is difficult to display a high number of items or rules with this type of charts.

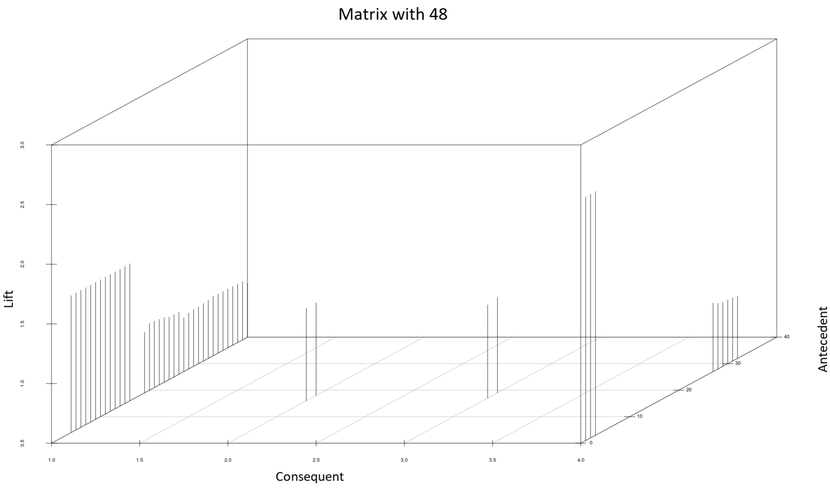
3.3 Grouped Matrix-based visualization

In this section, we explain a mixed technique named the grouped matrix-based visualization. In the previous case, Matrix-based, were limited to the number of



(a) 2D Matrix plot

(b) Scatter plot



(c) 3D Matrix plot

Figure 2: Matrix-based examples

rules it can be visualized effectively and it presented a difficult understanding of relationships between items and rules [22]. Using the grouped version we can explain better the relation between rules and itemsets. In Figure 3 we can see an example of this type of visualization using the ArulesViz package [15].
The representation in three dimensions is similar to the one explained for two dimensions for figure 2a, but the third axis is used to visualize a measure of the rule. With this, it gains inperpretability but it is even more complex to determine which itemsets belong to a visualized rule.

Advantages

Its principal advantage is that it enables the display of large sets of rules. It is a good technique for analysing rule by antecedent, since rules having the same antecedent can be grouped together.

Disadvantages

However although it is a useful technique, its main weakness is that it cannot be easily adaptable to display rules with more than one consequent. In addition to that, this technique has some problems when visualizing very frequent itemsets. In this situation this kind of tools are not very useful since very frequent rules filled almost all the available space.

3.4 Graph-based visualizations

Finally, the last type of techniques are those based in a graph representation. In this way, we need to transform association rules into a graph. In the literature we can find different ways to transform rules in to graphs [24, 13]. One way is to represent each item

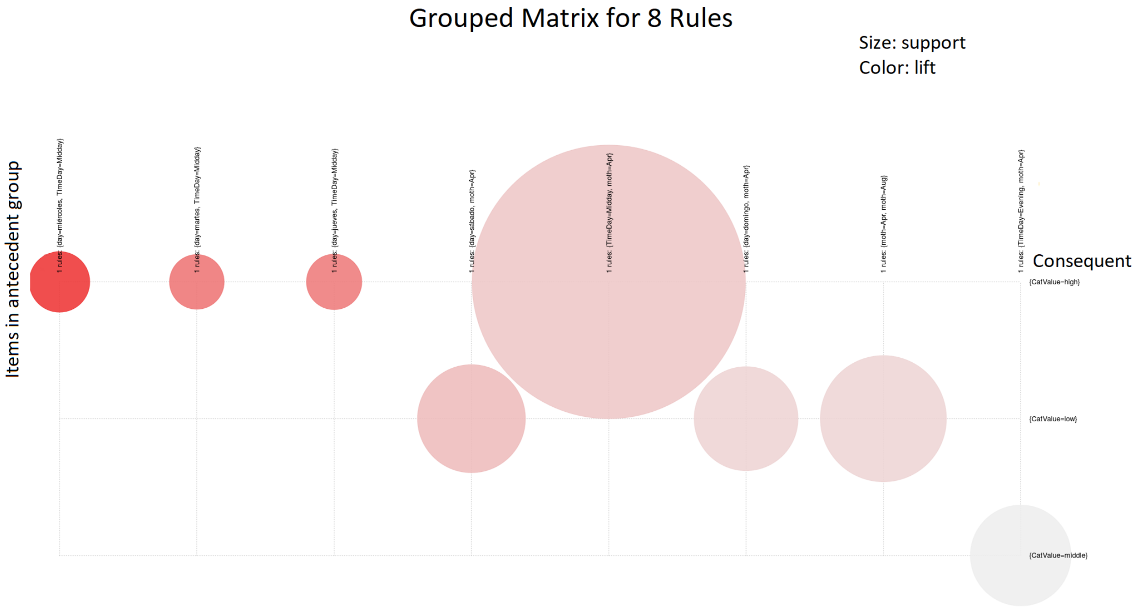


Figure 3: Grouped matrix-based visualization by ArulesViz

as a node and for each rule adding an edge connecting the items belonging to the rule.

We can observe different kinds of graphs in Figure 4. They can be divided into two groups attending to the type of layout. Within the first group we can distinguish the paracoord (parallel coordinates)[3] the WiFIsViz [10] and VisAR [23] (see the Figures 4c, 4b, 4d respectively).

In the case of WiFIsViz, we can see that this tool is very complex to understand because the representation divides the rule into antecedent and consequent difficulting to visualize the rule together. VisAR and Paracoord are very similar and they allow to understand very good the relation between itemsets in the rules, but the drawback is when we have a large set of rules or items. In these cases the tools do not have mechanisms to show all rules.

The second group contains graph without layouts restrictions (see Figure 4a and 4b). Different layout can be used and customized like the nodes colour, shapes, etc.

Advantages

As a result of using a graph structure, we can improve the visualization of association rules customizing the available layout. This can improve the representation and increase the interoperability when visualizing large sets of rules. Besides, the graph representation is suitable for visualizing rules with several consequents and antecedents.

Disadvantages

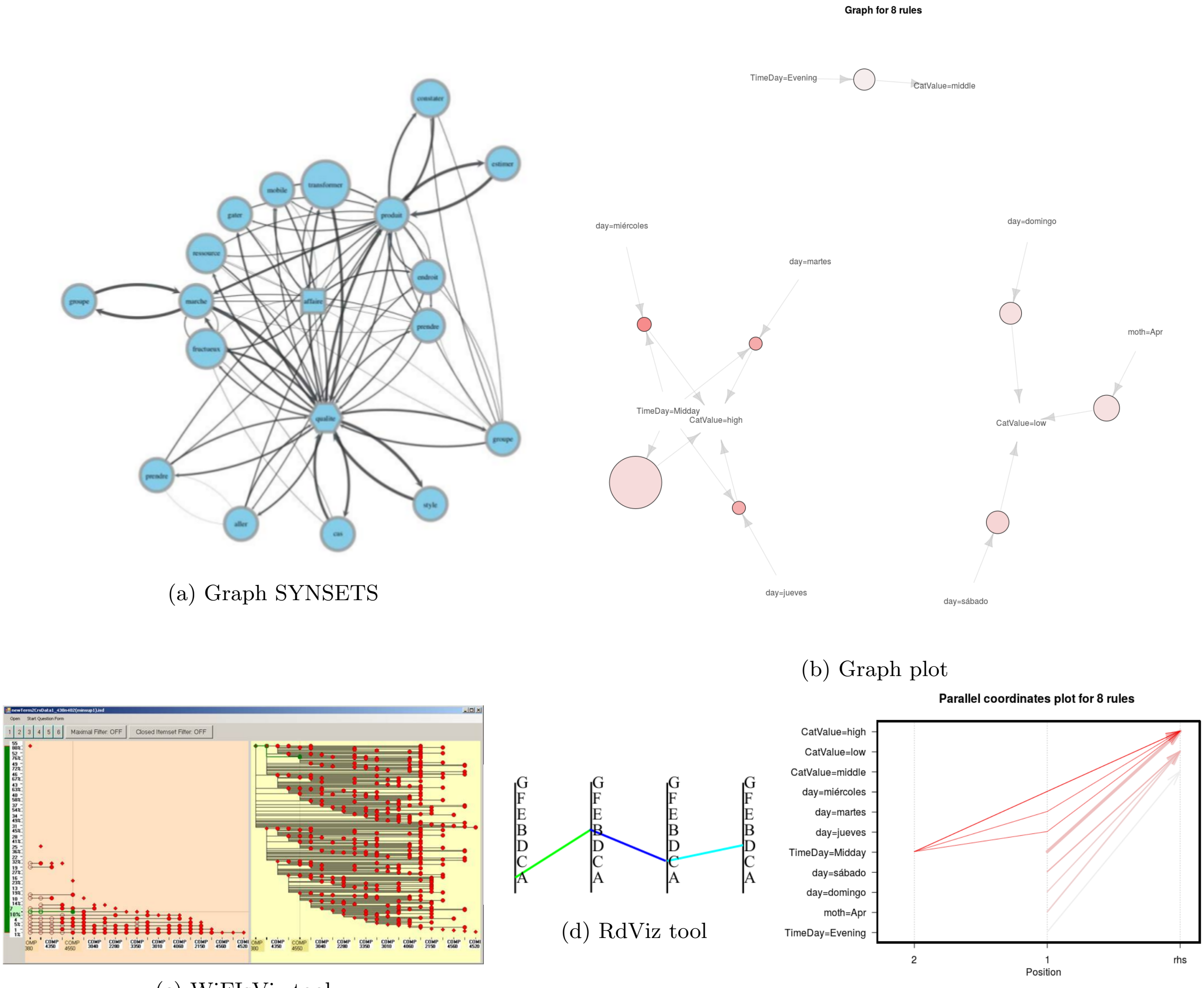
Nevertheless, the graph-based visualizations have some drawbacks when working with large sets because depending on the layout used, some rules are lost behind other rules. In addition, this type of visualization is not prepared to handle with different measures at the same time, since using more than two layouts (e.g. size and colour intensity) to represent them may cause some confussion to the user. Additionally, they are not prepared to represent fuzzy association rules.

3.5 Comparison of tools

In this section, we compare and analyse some characteristics of the reviewed visualization techniques. In table 1 the reader can find a summary of such characteristics.

Firstly, the basic representation as a table, is useful to observe a fragment of the complete set of rules, coming, for instance, from a query but not for a complete view for all of the discovered rules. For a broader observation of the complete set of rules it can be used the Scatter or the Matrix plot. These two methods allow, in a second step, to focus on determined rules for a detailed inspection. These visualizations are very useful to know the overall result of the mining process. In addition, it is very easy to inspect what are the main itemsets and rules together with their measures. Besides that, it can be used the grouped matrix visualization for grouping the rules by itemsets and knowing the number of rules found for each consequent and antecedent.

Lastly, it can be used the graph-based visualizations



(a) Graph SYNSETS

(b) Graph plot

(c) WiFIsViz tool

(d) RdViz tool

(e) Paracoord plot

Figure 4: Graph-based examples

Library	Method	Type	Focus	# of measures that can be displayed	# rules
arulesViz [15]	Table	2D	Measures and rule	*	1000
arulesViz [15]	Scatter plot	2D	Measures and set of rules	3(interactive information)	1000
arulesViz [15]	Tow-key plot	2D	Rule length	2+order	< 1000
arulesViz [15]	Matrix	2D and 3D	Antecedent and Consequent	1(interactive information) and 1	100000
arulesViz [15]	Grouped Matrix	2D	Antecedent and Consequent	2	1000
arulesViz [15]	Graph	2D	Items	2	100
FpVAT [18]	RDViz(Raw data visualization module)	2D	Rules relationship	3	1000
WiFIsViz [10]	Orthogonal Graph	2D	Rules and frequent itemsets	3	1000
VisAR [23]	Orthogonal Graph	2D	Rules and frequent itemsets	3	1000

Table 1: Comparative table of tools for visualizing association rules.

for more complex and complete displaying of rules. In this way, they can be used for more complex set of rules, enabling the visualization of different groups of rules related by itemsets. Besides, they enable to use more than two measures due to the use of different layouts configuration.

As a summary, we have found different tools for visualizing association rules, but not all of them are easily adaptable for visualizing fuzzy association rules. For this reason in next section we propose a new way for displaying fuzzy association rules.

4 A Proposal for Fuzzy Association Rules visualization

After the state of the art about the visualization tools for association rule made, we found that not all of the tools are suitable for displaying fuzzy association rules. For this reason we propose a new way to display this kind of rules.

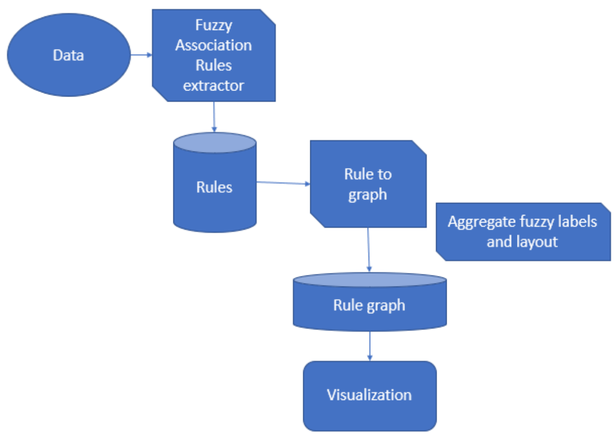


Figure 5: Pipeline proposed for fuzzy association rules visualization

The procedure (that can be seen in Figure 5) is describe in more attention to the transformation of itemsets and rules into nodes and edges respectively. Consequently, we fit the transformation by adding a new layer representing the same type of fuzzy-label as it can be seen in blue and green areas in Figure 6 like a new kind of node. Then, by means of these layers in the graph structure we can visualize the rules grouping them attending to the same type of label. For instance, if we represent the temperature, the labels "cold", "warm", and "hot" belong to the same group. We can see an example of this proposal in Figure 6, where the nodes coloured in the same cluster belong to the same type of attribute.

5 Conclusions and future works

To sum up we have reviewed different kinds of techniques to visualize association rules. Some of these techniques can be adapted for fuzzy association rules visualization.
The new proposal uses a Graph-Based visualization with some improvements to adapt the graph structure to represent fuzzy association rules. In this way, fuzzy association rules can be visualized by a graph, representing in addition some interesting features like the type of attributes.
As regards future research, we want to make some experimentations with large sets of fuzzy association

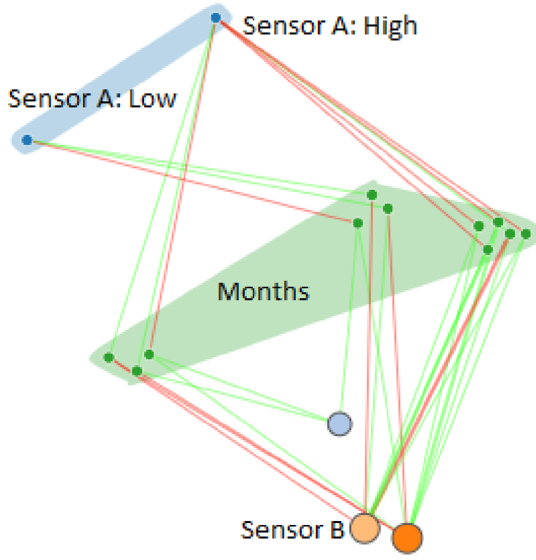


Figure 6: An example of graph visualization for fuzzy association rules.

rules. Additionally, we plan to improve the proposal made in this paper for fuzzy association rules combining different layers and/or layouts.

Acknowledgement

This research paper is part of the COPKIT project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 786687 and by the Energy IN TIME project funded from the European Union under Grant agreement No. 608981, and of the project grant TIN2015-64776-C3-1-R supported by the Spanish Ministry for Economy and Competitiveness.

References

[1] R. Agrawal, T. Imielinski, A. Swami, Mining associations between sets of items in large databases, in: ACM-SIGMOD International Conference on Data, 1993, pp. 207–216.

[2] R. Agrawal, R. Srikant, Fast algorithms for mining association rules in large databases, in: Proceedings of the 20th International Conference on Very Large Data Bases, VLDB '94, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1994, pp. 487–499.

[3] B. Almende, B. Thieurmel, T. Robert, visnetwork: network visualization using 'vis. js'

library. 2017, URL https://CRAN. R-project. org/package= visNetwork. R package version 2 (1) (2017) 172.

[4] R. J. Bayardo Jr, R. Agrawal, Mining the most interesting rules, in: Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, Citeseer, 1999, pp. 145–154.

[5] G. Bello-Orgaz, J. J. Jung, D. Camacho, Social big data: Recent achievements and new challenges, Information Fusion 28 (2016) 45–59.

[6] F. Berzal, I. Blanco, D. Sánchez, M.-A. Vila, A new framework to assess association rules, in: Advances in Intelligent Data Analysis, Springer, 2001, pp. 95–104.

[7] F. Berzal, M. Delgado, D. Sánchez, M. Vila, Measuring accuracy and interest of association rules: A new framework, Intelligent Data Analysis 6 (3) (2002) 221–235.

[8] A. Boukerche, S. Samarah, A novel algorithm for mining association rules in wireless ad hoc sensor networks, IEEE Transactions on Parallel and Distributed Systems 19 (7) (2008) 865–877.

[9] W. Castillo-Rojas, A. Peralta, C. Vargas, Visualización exploratoria e interactiva de modelos de reglas de asociación, Ingeniare. Revista chilena de ingeniería 23 (2015) 505–513.

[10] W. Chen, C. Xie, P. Shang, Q. Peng, Visual analysis of user-driven association rule mining, Journal of Visual Languages & Computing 42 (2017) 76–85.

[11] M. Delgado, N. Marín, D. Sánchez, M. Vila, Fuzzy association rules: General model and applications, IEEE Transactions on Fuzzy Systems 11 (2) (2003) 214–225.

[12] M. Delgado, M. Ruiz, D. Sánchez, New approaches for discovering exception and anomalous rules, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 19 (2) (2011) 361–399.

[13] W. Fan, X. Wang, Y. Wu, J. Xu, Association rules with graph patterns, Proceedings of the VLDB Endowment 8 (12) (2015) 1502–1513.

[14] H. Feng, M.-J. Lesot, M. Detyniecki, Using association rules to discover color-emotion relationships based on social tagging, in: R. Setchi, I. Jordanov, R. J. Howlett, L. C. Jain (Eds.), Knowledge-Based and Intelligent Information and Engineering Systems, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, pp. 544–553.

[15] M. Hahsler, arulesviz: interactive visualization of association rules with r, The R Journal 9 (2) (2017) 1.

[16] H. Hofmann, A. Siebes, A. F. Wilhelm, Visualizing association rules with interactive mosaic plots, in: KDD, Vol. 2000, 2000, pp. 227–235.

[17] E. Hüllermeier, Association rules for expressing gradual dependencies, in: Proc. PKDD 2002 Lecture Notes in Computer Science, 2431, 2002, pp. 200–211.

[18] C. K.-S. Leung, C. L. Carmichael, Fpvat: a visual analytic tool for supporting frequent pattern mining, ACM SIGKDD Explorations Newsletter 11 (2) (2010) 39–48.

[19] H. Li, Y. Wang, D. Zhang, M. Zhang, E. Y. Chang, PFP: parallel fp-growth for query recommendation, in: Proceedings of the 2008 ACM conference on Recommender systems, ACM, 2008, pp. 107–114.

[20] J. Pei, J. Han, B. Mortazavi-Asl, J. Wang, H. Pinto, Q. Chen, U. Dayal, M.-C. Hsu, Mining sequential patterns by pattern-growth: The prefixspan approach, Knowledge and Data Engineering, IEEE Transactions on 16 (11) (2004) 1424–1440.

[21] C. Sievert, C. Parmer, T. Hocking, S. Chamberlain, K. Ram, M. Corvellec, P. Despouy, plotly: Create interactive web graphics, R package version 4 (1) (2017) 1.

[22] A. Unwin, H. Hofmann, K. Bernt, The twokey plot for multiple association rules control, in: European Conference on Principles of Data Mining and Knowledge Discovery, Springer, 2001, pp. 472–483.

[23] L. Yang, Visualizing frequent itemsets, association rules, and sequential patterns in parallel coordinates, in: International Conference on Computational Science and Its Applications, Springer, 2003, pp. 21–30.

[24] S.-J. Yen, A. L. P. Chen, A graph-based approach for discovering various types of association rules, IEEE Transactions on Knowledge and data Engineering 13 (5) (2001) 839–845.