An Approach to Visualize Implications

Pablo Cordero, Manuel Enciso, Ángel Mora, Pablo Gómez González

Universidad de Málaga (Spain)
pcordero@uma.es, enciso@lcc.uma.es, amora@ctima.uma.es

Abstract. Visualizing implications has become a hot topic, providing new solutions to reveal the knowledge contained in the sets of rules. In Big Data applications, it is even more fundamental since their datasets usually produce a huge number of rules. Data visualization can be considered as a tool to illuminate the information, guiding the search for important rules or significant attributes. Usually, the user do not want to exhaustively examine all the implications, but rather to analyze the relevant knowledge in the rules.

In Formal Concept Analysis (FCA), some well-known tools allow to visualize the concept lattice and, even more, the implications. They focus on how to present these two visions of the same information, but they do not extract a further knowledge. Here, we present a new visualization model for implications, oriented to display and infer some interesting insights from the set of implications.

Keywords: Data visualization · implications · formal concept analysis

1 Introduction

Big Data has quickly become a popular topic and this situation has inspired many researchers to provide new methods, techniques and solutions to address some of its challenges. In [6] the authors make a review of Big Data applications, challenges, techniques and technologies. They remark that it "*also arises with many challenges, such as difficulties in data capture, data storage, data analysis and data visualization*". In this work we focus on this last challenge, which is closely tied, in our opinion, with the data analysis issue. In big data, where the high volume does not only involve the amount of data but also its complexity, data analysis has to be improved with some way to visualize the information, providing the human interaction too.

In [25], the authors summarize the power of data visualization, outstanding its benefits in different issues, adding a percentage measuring their impact. In the top three of this study they have highlighted the following: Improved decision-making (77%), Better ad-hoc data analysis (43%) and Improved collaboration/information sharing (41%).

Data visualization has been traditionally conceived to be the last step of data analysis. However, when data analysis refers to big data, data visualization allows the user to discover more insights by himself, joining the power of big data techniques and the human interaction [22]. This view of data visualization has been strongly emphasized by some authors [6]: *"The interactive analysis processes the data in an interactive environment, allowing users to undertake their own analysis of information. The user is directly connected to the computer and hence can interact with it in real time. The data can be reviewed, compared and analyzed in tabular or graphic format or both at the same time."*

Data visualization is not a new trend in FCA. Usually, the Concept Lattice is represented as a Hasse diagram where some elements has been added to enrich its semantics, adopting the so-called line diagram [10]. Since the very beginning, it was used to represent the concept lattice in a proper and direct way, by using a natural description that communicates the knowledge in a natural way [7]. This lattice can be efficiently represented as a labelled directed acyclic graph where vertices are formal concepts and its adjacency relation is the transitive reduction of the ordering relation. The line diagram has a strong point: it enclose several views of the information (concept order relation, implications, intent/extent connection, etc) in just one –and simple– representation. However, in our opinion, behind its apparent simplicity, the user has to be very familiarized with its interpretation to properly understand all the information. In [8] the line diagram is presented as a key element for some FCA topics: concept hierachy, attributes partition, etc. In [21] the authors study this diagram and make emphasis on three issues to improve this data visualization approach: reduction, layout and interaction. They also have reviewed three tools to visualize concept lattices, developed by the Defense Science and Technology Organisation [20]: Carve, SORTeD and DAnCe. In [14] the author presents an approach enriching the traditional representation of Concept Lattices by providing new visual elements and providing some interactive features, updating the classical representation to the new trends in visualization. He also studies the situation of the diagram when some changes are performed in the information, for instance when a new attribute/object are introduced or deleted. In addition, he also presents various strategies for pruning a formal concept lattice, to gain a clearer structure or to emphasize on interesting parts. These strategies provide more readable diagrams than the classical one, but visualization models are not significantly changed. In [17] the authors introduce a significant change, simplifying the Hasse diagram by a tree-like approach. They suggest that this variation allows a better understanding of the diagram, since the simplification from graphs to trees preserves all lattices entities and allows the user to focus on some structures and patterns, better guiding the search of insights. In addition, they enriched the visual vocabulary by adding colors and size, both in the nodes and in the lines, to add some semantic information to the new model. We end this review with the work presented in [1], where the authors mainly use Hasse diagrams to visualize concept lattice but with a strong root on the FCA framework. Hence, they provide the visualization of Pattern Structures and AOC-posets, concept annotation, filtering concept lattice and pattern concept lattice based on several criteria. Filtering allows the representation of sublattices and superlattices of interesting concepts, providing a way to easily discover new insights.

Unlike the others works we have just previously mentioned in this review, this work also includes a former visualization of implications. Traditionally, FCA literature has accepted this two fold views of the same knowledge, designing methods to represent and navigate in the concept lattice and other ones to manage and infer new implications. Recently, some authors have proposed the use of data visualization to represent implications and association rules. This issue is far to be finally addressed. In this work we propose a new step to approach the implication visualization and exploration. Our starting point is a set of implications, no matter how it has been extracted from the data set. We refer the reader interested in this area to the recent works [3,19].

The work is organized as follows: in Section 2 we make a review of the tools that have been presented to visualize the implications, with a particular emphasis on the R package arulesViz presented in [11]. In Section 3 we present the contribution of our work: a new model to visualize the implications. The model has been guided by some semantics characteristics of the attributes, in particular by the role that the attribute plays in the system. We end the paper with a Conclusions and Future Works section.

2 Visualization of Implications

In our work, we focus on the implications appearing in Formal Concept Analysis framework. As we have said, in [1] the authors propose an interesting tool to visualize the concept lattice and they also add some visualization of the implications associated to the concepts. This work can be considered as a first step in FCA. Unlike the authors put their attention on the concept lattice, a first approach regarding implications is presented and a basic visualization of implications with some direct information is showed. The authors also make a brief reflection on the need to delve deeper into this issue [2]. They include in LatViz [1] two preliminar visualizations:

- One plot with the ID of the each implication and the attributes appearing in their left hand side (LHS) and the right hand side (RHS).
- One matrix where in the rows appear the LHS and in the columns the RHS, so that the matrix is a kind of relationship where each element corresponds to an implication. Some colors are used to show the support and lift.

The authors conclude that this approach is not suitable for large sets of implications. They also provide a third diagram, a scatter plot where each dot corresponds to an implication. The x-axis is the increasing support and the y-axis is the increasing lift.

[1] The plots described here can be accesible in the cited works, but unfortunately they cannot be generated directly with the tool (https://latviz.loria.fr/).

In the literature, there exist other tools to display the information about rules: WiFisViz [15] and VisAR [23]. In our opinion, both of them present a very similar approach that the LatViz attribute matrix already described.

We now review the R package arulesViz [12,13], which can be considered as a proper tool to visualize association rules. It is oriented to present the information so that visual analysis can be further developed. As M. Hahsler says in [12] *"mining association rules often results in a vast number of found rules, leaving the analyst with the task to go through a large set of rules to identify interesting ones. Visualization and especially interactive visualization has a long history of making large amounts of data better accessible"*.

arulesViz incorporates the usual elements to visualize rules, already used in the tools we have mentioned above, and it also adds some clustering techniques. It allows to show the rules in some flexible ways as follows:

- A detailed information is visualized by means of a scatter plot called "two-key plot" [24] where support and confidence are used for the x and y-axes and the color of the points is used to indicate the number of attributes contained in the rule (the order).
- As a second level in the visualization, a grouping technique is used to design a new method, called grouped matrix-based visualization, which is based on a method of clustering rules.
- Using a graph-based plot to connect the attributes appearing in the association rules.

One of the interesting features of this tool is the incorporation of some customization to perform some kind of visual analysis. Thus, it is possible to change the information in the axes and it also allows us to show a third or fourth parameter by means of other visualization elements like color or shadow. These elements provide a good balance between simplicity of the plots and quantity of information [9].

We refer the reader to [13] for further details of these features. In the rest of the section, we summarize how arulesViz works to extract interesting knowledge.

The starting point is to use the R language to generate the rules [11] using the apriori algorithm [2]. Then, these rules are directly depicted in a scatter plot (see Figure 1) where each rule is a dot, having two measures on the axes (support and confidence) and a third one (lift) by using color intensity [3]. This plot provides a great amount of information, in a very intuitive way. However, the information displayed is centered in the rule as an atomic element, which must to be improved for a better knowledge discovery.

In the next plot (see Figure 2), the idea is to show some semantic information: the role played by the attributes in the rules. Thus, the different itemsets in each

[2] As we previously mentioned, we use the extraction method included in arulesViz, but any other method can be used istead.

[3] In this section we are visualizing the set of rules of the Adults dataset in the UCI Repository (http://archive.ics.uci.edu/ml/datasets/adult) where the threshold support is set to 0.3 and the confidence to 0.55. This set has 563 rules.

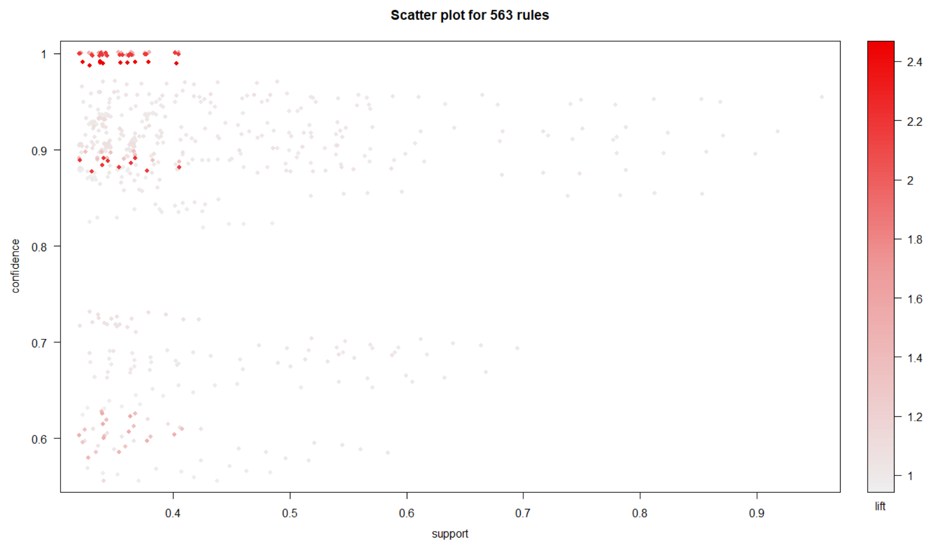

Fig. 1. Basic arulesViz scatter plot

antecedent and consequent are depicted in the X and Y axes respectively. Both index sets are independent, thus the same number do not represent the same itemset. An interactivity feature is included in arulesViz: when a dot is clicked in the plot, the full description of the rule, including the name of the attributes in the LHS and RHS, is showed. Rules can be colored according to the value of a third feature (in this case, we have selected their lifts). The problem with this plot is that for a huge number of rules, it would be impossible to discover some useful insights. Even for a mid size volume of rules, the connection between the premise and the conclusion of the rules is not easily identified in the plot.

A third plot, named the Two-key plot [24] is also included in the arulesViz package (see Figure 3). In the axes, support and confidence are showed and the color of the points is used to indicate the number of attributes in the rule. This plot not only shows the usual parameters of the rules (support, confidence, etc.) but it also incorporates some properties that can help to identify the most important rules when a huge number of rules is visualized.

As we mentioned in the Introduction, the current trends in data visualization includes some kind of user interaction [22] so that it provides not only a way to communicate the data analysis we have already did, but also to make some further exploration by means of the visual analysis. In arulesViz, some interactive features for scatter plots have been considered, and you can do zoom into a plot, panning, and hovering over points to obtain a deeper information about the rule. In addition, it also allows rule inspection by selecting a region or a point in the

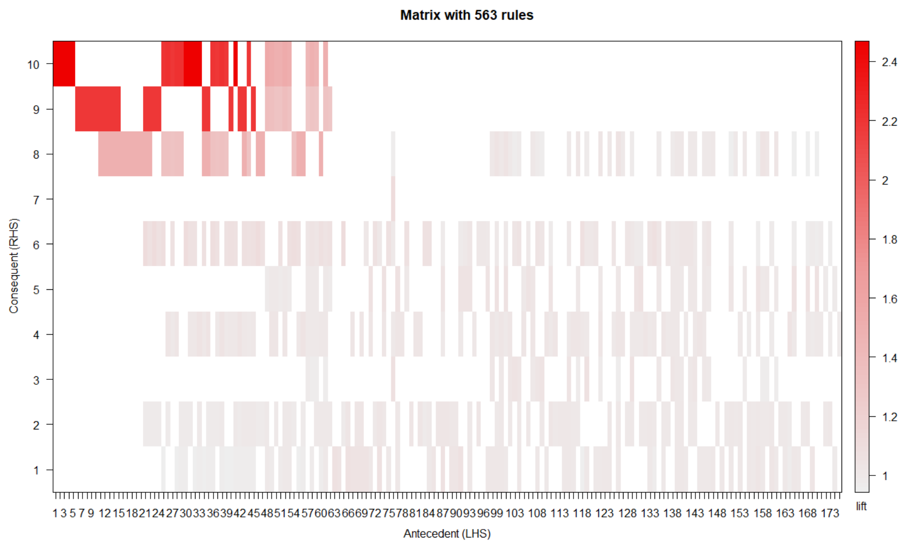

Fig. 2. Representation of LHS and RHS itemset matrix in arulesViz

plot, filtering rules, etc. To end the review of arulesViz, we remark that the author consider a limit of 1,000 rules to be displayed [13].

As a general conclusion to this section, in our opinion the visualization models designed for implications are extremely centered on the implications itself. Thus, they are focussed on displaying the rules and their natural characteristics: support, confidence, lift, attributes in the RHS or LHS. However, when the number of implications grows, this information is difficult to be visually analyzed and some other elements need to be displayed. In our opinion, this new elements have to be defined so that interesting insights allows to infer the role that the attributes play in the system. This knowledge seems to be difficult to extract from the parameters usually included in the visualizations. Thus, we present a novel approach in this line in the following section.

3 Our Approach to Data Visualization for Implications

As we have explained, strategies presented in literature were strongly based on the use of some basic information that could be interesting when the set of implications has a small or medium size. However, when its size grows this information does not provide interesting insights to make visual analysis. In this situation, the visualization has to propose new models based on some semantics information or patterns to discover interesting knowledge.

In this work, we propose a three-step visualization model to develop a visual analysis process. We approach this model by designing two interactive plots and

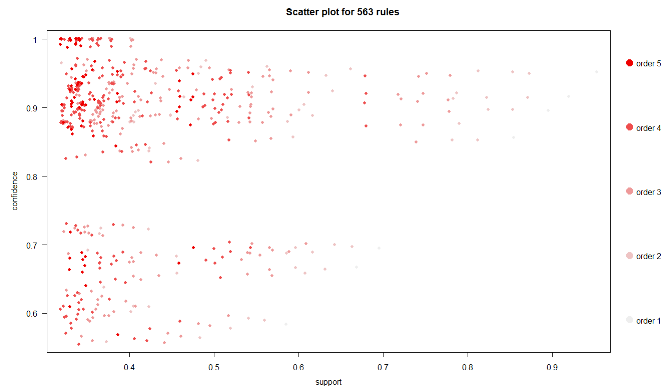

Fig. 3. The two-key plot in arulesViz

a final data table. This process will help the user to move from a general set of implications to a smaller subset. The search is guided by the role that the attributes play in the system.

In our approach we focus on implications, but not in association rules. We use the basic elements of association rules that also make sense for implications: LHS and RHS. In addition, we have also designed specific parameters to enrich the visualization model, providing a deeper knowledge of the system. The key elements used in our visual model are the following:

- Attribute closure. We are interested in examining the full semantic power of the premises, thus we transform the original set of implications by including in the RHS the maximum attribute set for each premise, i.e. given a set of implications Σ and an implication $A \to B \in \Sigma$ we built $A \to A^+ - A$ where A^+ is the syntactic closure of A with respect to Σ. This syntactic closure can be approached by using Armstrong Axioms [4] or Simplification Logic [18]. This transformation can be done in an efficient time, since the syntactic closure has a quadratic cost on the set of implications [16].
- RHS and LHS cardinals. We count the number of attributes in the LHS and the RHS since we are interested in the relation between the sizes of both sides of the implication. This parameter shows whether or not the implications are balanced. Depending of the environments, the user can search for implications where a small itemset determines a big one or vice versa.
- Presence in RHS and Presence in LHS. We count how much an attribute appears in all the implications by measuring the percentage of the implications where the attribute appears, respectively, on the right and left hand sides.
- Global presence. Its is defined as the sum of the relative presence in RHS and relative presence in LHS. Thus, we are searching for interesting attributes instead of interesting rules.

– Generator/generated attribute roles. An attribute whose relative presence in LHS is greater than its relative presence in LHS will be called a generator. Otherwise, an attribute presence in the RHS is greater than its corresponding presence in LHS is considered a generated attribute.

In the next three subsections, we describe how these parameters are used to design the plots that are the base of our three step analysis process.

3.1 The Atribute Plot

As we have seen in previous sections, most of the works in the literature center the visualization on the implications. However, we start our analysis with the attributes, by plotting them according to some semantic information. Particularly, we show the role played by each of them: is it a generator or generated attribute?

Thus, we design a plot where the axes represent the presence of the attributes in the LHS (X axis) and the RHS (Y axis) of all the implications (see Figure 4). In this plot, we also draw a straight line which represents the identity function. The points located above the line are attributes playing the role of a generated attribute. Conversely, the points bellow the line are generators. On the one hand, a direct interpretation of this plot is the usual presence of the attribute in the premises or the conclusions of the implications. On the other hand, this plot also helps to figure out how the attributes are located in the concept lattice. Top and bottom attributes in the lattice are the attributes placed close to the axes. Thus, the attributes in the lower right quadrant of the diagram appears in closed sets belonging to the lower levels of the concept lattice whereas the upper left dots corresponds with attributes in the closed sets located in the upper levels of the lattice.

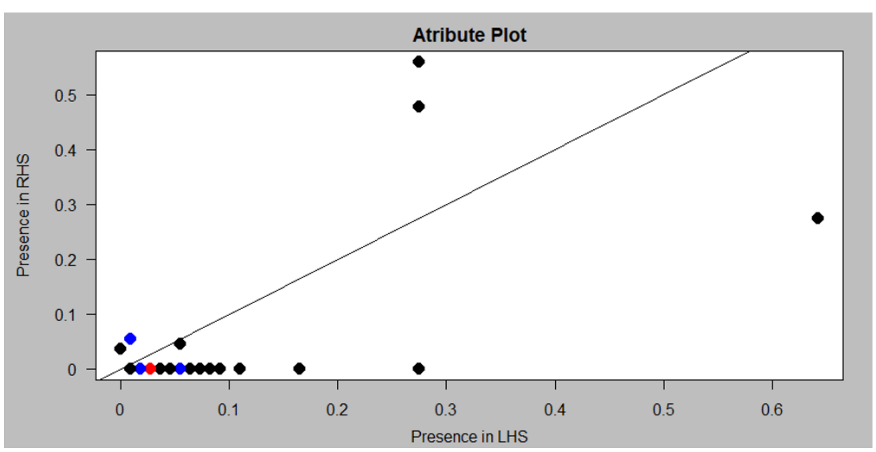

Fig. 4. Atribute Plot with generated/generator attribute roles.

In addition, we add some interactive capabilities to this diagram. The user can click on any point, coloring the attribute red, and triggering a new diagram which will be described in the following subsection. Moreover, since in some cases the attributes come from an original many-valued formal context [5] that were converted into a classical one by multiplication of the values. If applicable, we add some element to visualize this situation. So, we consider that all the attributes that corresponds with different values of an original attribute form a category. When a dot is clicked, all the attributes belonging to the same category are highlighted in blue color. As an illustration of this mode, in Figure 4, the red point represents "occupation: transport moving", while the blue ones represent the other occupations: "occupation: Armed forces","occupation: tech-support" and "occupation: craft repair".

3.2 The Implication Plot

Once you have selected one attribute in the attribute plot, another diagram is created. Thus, this new plot presents the information filtered according to the selection made in the first step. In this second diagram, we shift from the attribute to the implication view. Our goal now is to visualize the information of the implications following the same idea behind the attribute plot: the balance on the number of attributes between the RHS and the LHS. Thus, this new plot shows the implications gathered by the cardinal of their LHS and RHS. Thus, our idea is to group in the same dot all the implications following the same structural schema. It also adds some additional information like the number of implications in the group and the combination of cardinals in both sides. An example of this diagram is showed in Figure 5. It follows the following visualization model:

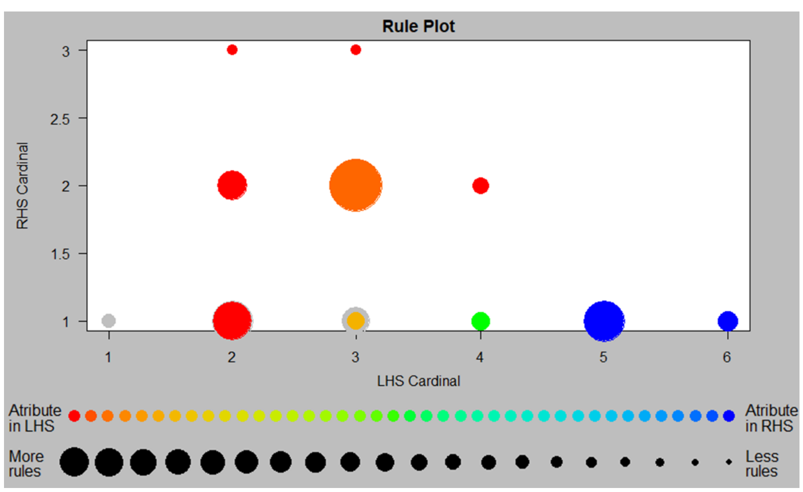

Fig. 5. Implication diagram with information of the selected attribute.

- The X axis is the increasing LHS cardinal while the Y axis is the increasing RHS cardinal. As we mentioned, each plot group all the implications with the corresponding cardinals.
- The size of each dot represents the number of implications with the same pair of values in both sides. The size is proportional to the number of rules for the selected point.
- Each dot embebed an inner point representing the percentage of the implications having a given LHS and RHS size where the selected attribute appears.
- Finally, the dot color represents the main role that the attribute plays in the implications with a given LHS and RHS. The generator character is highlighted in red color whereas the blue color is used for the generated one. The intermediate situations are colored in a color determined by the range between the red and blue colors.

This diagram allows to identify if the attribute plays the generated/generator role in an implication schema, according with their size, LHS and RHS cardinals. Once we have discovered some insights, then we can finally get the information of all the implications that have been grouped in each dot by clicking on the corresponding dot. This will be showed in a table explained in the following subsection. In the future, we plan to add to this table some kind of visual representation, following the models presented in previous works, as a help to ease the exploration when the size of the output implication set increases.

3.3 The Rules Data Table

The final step is to show all the implications with the selected LHS and RHS cardinals. They are placed in a data table with some classical estimators so that the user can examine the set according to the search he has performed by using the visual tool. This data table have been built by the user in just two actions, the selection of one attribute and one implication schema, allowing the user to develop visual analytics in a few steps. Following the example of the previous section, if the points with the biggest RHS are selected –the (2,3) and (3,3) pairs– the implications collected in these points are showed in Figure 6. This table also has some interactive skills, allowing the user to sort it by the desired parameter.

4 Conclusion and Future Works

In the work we focus on the need to visualize the implications, which are key elements in FCA and have not attracted as much attention as concept lattice in this issue. We present a new visualization model designed to better discover insights from the set of implications that was not included in other previous methods. Our model is guided by the idea of identifying the role that the attributes play in the implications. We have described some elements where the model is based on and the diagrams and tables used to visually present the information.

As future works, we plan to connect this model with the visualization of the concept lattice so that the two-fold representation that was originally proposed in the FCA framework can be also handle in the visualization issue.

Show 10 entries

	LHS	RHS	support	lift	count	right.support	left.support
[1]	{relationship=Husband,capital-loss=High}	{marital-status=Married-civ-spouse,sex=Male,capital-gain=None}	0.015	2.454	732.000	0.407	0.015
[2]	{relationship=Husband,capital-gain=High}	{marital-status=Married-civ-spouse,sex=Male,capital-loss=None}	0.022	2.454	1,082.000	0.407	0.022
[3]	{relationship=Husband,capital-gain=Low,hours-per-week=Full-time}	{marital-status=Married-civ-spouse,sex=Male,capital-loss=None}	0.014	2.182	704.000	0.458	0.014
[4]	{workclass=Private,relationship=Husband,capital-gain=Low}	{marital-status=Married-civ-spouse,sex=Male,capital-loss=None}	0.017	2.182	835.000	0.458	0.017

Showing 1 to 4 of 4 entries

Fig. 6. Information of the implications associated with the selected attribute.

Acknowledgment

This work has been partially supported by the project TIN2017-89023-P of the Science and Innovation Ministry of Spain, co-funded by the European Regional Development Fund (ERDF).

References

1. Alam, M., Le, T.N.N., Napoli, A.: Latviz: A new practical tool for performing interactive exploration over concept lattices. In: Huchard, M., Kuznetsov, S. (eds.) Proceedings of the Thirteenth International Conference on Concept Lattices and Their Applications, Moscow, Russia, July 18-22, 2016. CEUR Workshop Proceedings, vol. 1624, pp. 9–20. CEUR-WS.org (2016), `http://ceur-ws.org/Vol-1624/paper1.pdf`
2. Alam, M., Le, T.N.N., Napoli, A.: Steps towards interactive formal concept analysis with latviz. In: Kuznetsov, S.O., Napoli, A., Rudolph, S. (eds.) Proceedings of the 5th International Workshop "What can FCA do for Artificial Intelligence"? co-located with the European Conference on Artificial Intelligence, FCA4AI@ECAI 2016, The Hague, the Netherlands, August 30, 2016. CEUR Workshop Proceedings, vol. 1703, pp. 51–62. CEUR-WS.org (2016), `http://ceur-ws.org/Vol-1703/paper7.pdf`
3. Andrews, S.: In-close, a fast algorithm for computing formal concepts. In: International Conference on Conceptual Structures (ICCS) (2009)
4. Armstrong, W.W.: Dependency structures of data base relationships. In: IFIP Congress. pp. 580–583 (1974)
5. Belohlavek, R., Vychodil, V.: What is a fuzzy concept lattice? In: Belohlavek, R., Snasel, V. (eds.) Proceedings of the CLA 2005 International Workshop on Concept Lattices and their Applications, Olomouc, Czech Republic, September 7-9, 2005.

CEUR Workshop Proceedings, vol. 162, pp. 34–45. CEUR-WS.org (2005), http://ceur-ws.org/Vol-162/
6. Chen, C.P., Zhang, C.Y.: Data-intensive applications, challenges, techniques and technologies: A survey on big data. Information Sciences **275**, 314 – 347 (2014). https://doi.org/https://doi.org/10.1016/j.ins.2014.01.015
7. Eklund, P., Ducrou, J., Brawn, P.: Concept lattices for information visualization: Can novices read line-diagrams? In: Eklund, P. (ed.) International Formal Concept Analysis. pp. 57–73 (2004)
8. Eklund, P., Villerd, J.: A survey of hybrid representations of concept lattices in conceptual knowledge processing. In: Kwuida, L., Sertkaya, B. (eds.) International Formal Concept Analysis. pp. 296–311 (2010)
9. Few, S.: Show Me the Numbers: Designing Tables and Graphs to Enlighten. Analytics Press, second edn. (6 2012)
10. Ganter, B., Obiedkov, S.: Attribute exploration. in: Conceptual exploration (2016)
11. Hahsler, M., Grun, B., Hornik, K.: arules - a computational environment for mining association rules and frequent item sets. J Stat Softw **14**, 1–25 (2005)
12. Hahsler, M.: arulesviz: Interactive visualization of association rules with r. The R Journal **9/2**, 163–175 (2017)
13. Hahsler, M., Karpienko, R.: Visualizing association rules in hierarchical groups. J Bus Econ **87**, 317–335 (2017)
14. Kriegel, F.: Visualization of Conceptual Data with Methods of Formal Concept Analysis. Ph.D. thesis, Technische Universitat Dresden (10 2013)
15. Leung, K.S., Leung, C.K.S., P. Irani, P., L. Carmichael, C.: Wifisviz: Effective visualization of frequent itemsets. 2008 Eighth IEEE International Conference on Data Mining pp. 875 – 880 (2008)
16. Maier, D.: Theory of Relational Databases. Computer Science Pr (1983)
17. Melo, C., Grand, B.L., Aufaure, M., Bezerianos, A.: Extracting and visualising tree-like structures from concept lattices. Proceedings of the International Conference on Information Visualisation pp. 261 – 266 (08 2011). https://doi.org/10.1109/IV.2011.46
18. Mora, A., Cordero, P., Enciso, M., Fortes, I., Aguilera, G.: Closure via functional dependence simplification. Int. J. Comput. Math. **89**(4), 510–526 (2012)
19. Outrata, J., Vychodil, V.: Fast algorithm for computing fixpoints of galois connections induced by object-attribute relational data. Information Sciences **185**(1), 114 – 127 (2012)
20. Pattison, T.: Interactive visualisation of formal concept lattices. CEUR Workshop Proceedings **1244**, 78–89 (01 2014)
21. Pattison, T., Weber, D., Ceglar, A.: Enhancing layout and interaction in formal concept analysis. In: 2014 IEEE Pacific Visualization Symposium. pp. 248–252 (March 2014). https://doi.org/10.1109/PacificVis.2014.21
22. Simon, P.: The Visual Organization: Data Visualization, Big Data, and the Quest for Better Decisions. Harvard Business Review (6 2014)
23. Techapichetvanich, K., Datta, A.: Visar : A new technique for visualizing mined association rules. LNAI **3584**, 88 – 95 (2008)
24. Unwin, A., Hofmann, H., Bernt., K.: The twokey plot for multiple association rules control. PKDD01: Proceedings of the 5th European Conference on Principles of Data Mining and Knowledge Discovery pp. 472–483 (2001)
25. Venkataramana, S., Subash, S.R., Prakash, P.: Visualization of big data: Its tools and challenges. International Journal of Applied Engineering Research **9**, 5277–5290 (01 2014)