

SubSect — An Interactive Itemset Visualization*

Joey De Pauw¹, Sandy Moens¹, and Bart Goethals^{1,2}

¹ University of Antwerp, Belgium

² Monash University, Australia

1 Introduction

Itemsets and association rules are among the most simple and intuitive patterns that are used to explore transaction datasets. However, they lack meaning without both context and domain knowledge. Typically a user has to sift through hundreds of these patterns before finding an interesting one, losing sight of the forest for the trees. Furthermore, *interestingness* is a subjective measure that can only be approximated by objective metrics or features [3].

In previous work this problem has been tackled for instance by sorting and filtering patterns based on different metrics [3] or by trying to minimize the number of reported patterns to the most descriptive subset [1]. Another approach is to represent patterns in informative visualizations and rely on the end user to find what is interesting in their respective domain [2].

We propose a novel itemset and association rule visualization that makes it possible to inspect, assess, and compare patterns at a glance. This can not only save time and effort, but also reduce errors introduced by misconceptions. Our visualization is based on the double decker plot from Hofmann et al. [2] and exploits the monotonicity property, which states that itemsets have a lower or equal support compared to the support of their subsets.

2 Visualization

Consider the example in Figure 1a. Every item in the itemset is represented in the center. The arcs around the center items show three levels of itemsets that can be formed from these items. For example, the blue full circle near the center includes all four items A , B , C and D , and has a frequency of 0.2 as indicated by the label and its radius. The other segments represent subsets, like for example the cyan arc which spans items A , B and C . In correspondence with the higher frequency of this itemset (0.25), its arc also has a proportionally larger radius.

In every image only the most interesting and informative subsets are rendered: for a k -itemset these are the $k-1$ -itemsets and the 1-itemsets. Together this combination of subsets provides the most useful information: the 1-itemsets give a global context and the $k-1$ -itemsets place the k -itemset in a local context. Itemsets larger than one are given a unique color, making it easier to link multiple instances of the visualization that have items in common.

* Copyright © 2019 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

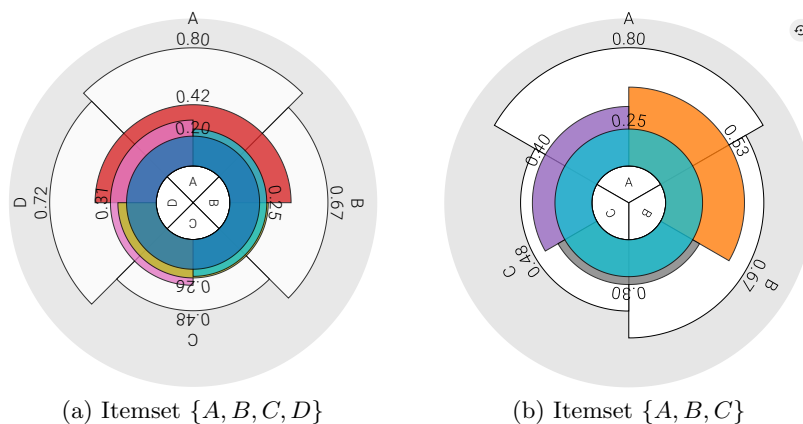


Fig. 1. Our visualization for the arbitrary itemset $\{A, B, C, D\}$ (a) and for one of its subsets (b). Each arc represents a set of items and shows its respective support.

Furthermore, the visualization is equipped with two interactions for maximal usability: *dive deeper* and α -*conditional view*³. Animations like hover highlighting indicate the presence of these interactions and gradual transition animations ease the transition between “states” of the visualization, making the effect of the interactions more clear. Clicking on the cyan arc for example will *dive into* its respective itemset $\{A, B, C\}$. An animation shows that item D is removed from the center and the cyan arc becomes a full circle. Three new subsets are now visible. The result is shown in Figure 1b. Naturally this action can be repeated from the new view to dive deeper or the user can choose to go back to the top level with the *reset* button that just became available.

Similar to the interaction for selecting an itemset to dig deeper, it is also possible to click a single item (in the center or on the outer edges) and add it to the α set or the “*scope*”. In this α -conditional view, the scope is visible on a smaller visualization to the left. On the right-hand side, we see the remaining items and itemsets, but now with their frequencies relative to the scope.

References

1. Calders, T., Goethals, B.: Non-derivable itemset mining. *Data Mining and Knowledge Discovery* 14(1), 171–206 (2007)
2. Hofmann, H., Siebes, A.P., Wilhelm, A.F.: Visualizing association rules with interactive mosaic plots. In: *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*. pp. 227–235. ACM (2000)
3. Tan, P.N., Kumar, V., Srivastava, J.: Selecting the right interestingness measure for association patterns. In: *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*. pp. 32–41. ACM (2002)

³ A live version with examples can be found on <https://joeydp.github.io/SubSect/>.