EuroVis Workshop on Visual Analytics (2014)
M. Pohl and J. Roberts (Editors)

Interactively Visualizing Summaries of Rules and Exceptions

Geetika Sharma[1], Gautam Shroff[1], Aditeya Pandey[1], Puneet Agarwal[1] and Ashwin Srinivasan[2]

[1]TCS Research
[2]Indraprastha Institute of Information Technology Delhi

Abstract

Rules along with their exceptions have been used to explain large data sets in a comprehensible manner. In this paper we describe an interactive visualization scheme for rules and their exceptions. Our visual encoding is based on principles for creating perceptually effective visualizations from literature. Our visualization scheme presents an overview first, allows semantic zooming and then shows details on demand using established principles of interactive visualization. We assume that rules and exceptions have been mined and summarized using available techniques; however our visualization is applicable for more general rule hierarchies as well. We illustrate our visualization using rules and exceptions extracted from real customer surveys as well as on rule sets derived from past literature.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—Line and curve generation I.3.8 [Computer Graphics]: Application—Visualization

1. Introduction

While a number of visualization techniques have been proposed for exploratory mining of rules from large datasets, we were motivated by the need for perceptually effective visualizations to aid in comprehending association rules and exceptions and communicating these to end users. We present an interactive layout for visualizing rule and exceptions mined and summarized using techniques such as [YCHX] and [LHH00]. Our layout can be used to visualize rule hierarchies as well as general rules and exceptions.

Association rules are amongst the widely used techniques for determining frequently occurring patterns in large datasets. A well-known difficulty with their use is the tendency to identify a very large number of rules from which the interesting ones have to be discovered, often through human intervention. Further, data gets fragmented into multiple overlapping patterns which are not easy to comprehend.

Many visualization techniques for association rules, such as [BGB03], [CHYN07] and [Yan05], have been proposed to address the problem of visualizing large sets of rules. However, the number of output rules is usually very large and visualizations often get too cluttered to aid significantly. In this paper, we focus instead on the problem of visualization of association rules with the purpose of understanding and communicating the dominant set of rules. For this we will draw on work on discovering rules with exceptions as techniques for summarizing large sets of association rules by a small subset of rules along with their *exceptions*. While the work on learning rules with one or more levels of exceptions has largely focussed on obtaining theories for data, they have not been adapted to the problem of visualizing association rules in meaningful hierarchies.

We propose a new interactive layout for rules based on the visual information seeking 'mantra': "Overview first, zoom and filter, then details on demand" [Shn96]. At first glance, our visualization shows the dominant rules. For each rule, its exceptions and their properties are highlighted. Our zoom works at a semantic level, allowing the user to navigate through different exception levels. A simple interface for obtaining additional details is provided. Our visual encoding is defined using well-known principles for creating perceptually efficient visualizations and we show visualizations of multiple rule sets. The rest of the paper is organized as follows. We discuss related work in Section 2 and present our visualization in Section 3. We show visualizations of different datasets in Section 4 and conclude in Section 5.

2. Related Work

Association rules were first defined in [AIS93]. A rule is defined as an implication $X \implies Y$, where X and Y are unique

DOI: 10.2312/eurova.20141146

attributes of a large dataset if X and Y occur together in a large number of transactions. X is called the antecedent of the rule and Y the consequent. The support of a rule is the proportion of records in which the rule occurs. The confidence of the rule $X \implies Y$ is the ratio of the number of records in which X and Y both occur and the number of records in which X appears; confidence ranges from 0 when X and Y never occur together, to 1 when they always occur together. Rules with confidence less than 1 have exceptions which do not imply the consequent Y. The lift of a rule is the ratio of confidence and the support of the consequent, indicating how 'interesting' the rule is; rules with lifts differing significantly from 1 are more interesting (or unexpected) than those with near unit lift. An exception has all the antecedents of the main rule and at least one more and does not imply the consequent of the main rule.

Most visualizations of association rules have been designed with the aim of discovering useful rules. While a survey of visualizations such as 2D and 3D matrices, two-key plots and double-decker plots can be found in [BD08].

A 3D visual tool, ARVis, for rummaging through rules is described in [BGB03]. Rules are drawn on a 3D information landscape, with more important rules placed in the foreground and less important ones in the background. Each rule is represented with a sphere whose area represents support and a cone whose height represents confidence. The use of 3D visualization in this case seems unjustified as (a) the data being represented is not inherently 3D, (b) the interface is more complex and harder to learn, (c) occlusion may hide relevant data points at certain viewing angles, (d) perspective projection may distort sizes of objects and (e) displaying and reading text is harder.

CbVAR [CHYN07] is a tool for extracting and visualizing clusters of rules from a dataset. The tool's purpose is to provide a global view of all the clusters discovered. Clusters are represented by a rule chosen on the basis of an interestingness measure such as lift or support using a 2D matrix of antecedents and consequents. A fish eye view of a cluster is provided by interacting with the tool which shows details of the rules in the cluster in a 2D or 3D representation.

All of these techniques have been designed to visualize a large number of rules. However, even if they were used to display a small number of rules, it is not clear whether they could be extended to represent exceptions effectively.

Rule mining to discover general rules and their exceptions is proposed in [LHH00]. They also show visualizations of rules with two antecedents represented by x and y axes and their ranges partitioning a 2D plane. However, this approach cannot be used when a rule or exception has more than 3 antecedents or have non-quantitative domains.

3. Visualizing Rules and Exceptions

Creating a visualization entails defining a mapping of data attributes to visual attributes of graphical elements that make up the visualization. Data attributes can be abstracted into categories: nominal, ordered and quantitative; these are then mapped to the visual attributes of position, size, colour, texture, intensity and so on. It is well-known that certain data types are better represented by particular visual variables aiding perceptual judgements about their values [CM84], [Mac86]. For example, nominal attributes are well represented by colour, shape and texture while quantitative variables are well represented by length, position, area, angle and volume. We use the above principles to motivate our visual encoding.

The interactive elements of our visualization are based on the mantra: "Overview first, zoom and filter..." from [Shn96]. We set semantic zoom levels based on the number of exception levels for a rule. At the highest level, a user sees one main rule and all its exceptions as shown in Figure 1. We use a circle as the mark to represent a rule and map the radius of the circle to the support of the rule. Since support is the portion of data for which the rule holds and radius controls the area of a circle, larger the support, larger the circle. Further, area is a perceptually effective indicator of quantitative variables, such as support, making judgements about their relative values easier.

The confidence of the rule is mapped to the circle's fill colour opacity, label 1 in the figure. The syntax of the main rule and, its statistics such as support and confidence are additionally printed in a text box above the circle, label 2 in the figure.

Exceptions to a rule are represented using the same visual encoding as for rules, but are contained within the main rule's circle. The first level exception circles have a fill colour contrasting that of the main rule. The representation of confidence by fill colour opacity is motivated by the fact that exceptions with higher confidence will have a stronger visual presence as they are more opaque than those with lower confidence. This intuitively concurs with the definition of confidence of a rule.

The next level of exceptions imply the consequent of the main rule, being exceptions to exceptions. They are visualized using the same encoding recursively but have the same fill colour as the main rule.

When the semantic zoom slider is moved to level two, the first level of exceptions come in focus. We treat antecedents as nominal variables and represent them by hue, label 3 in the figure. The exception circles transition into a pie with equally sized slices representing the antecedents. This visually indicates the number of antecedents that make up the exception. Hovering the mouse over an exception circle brings up a text box with its syntax and statistics, label 4 in the figure. Further, the value of an antecedent variable may lie in a

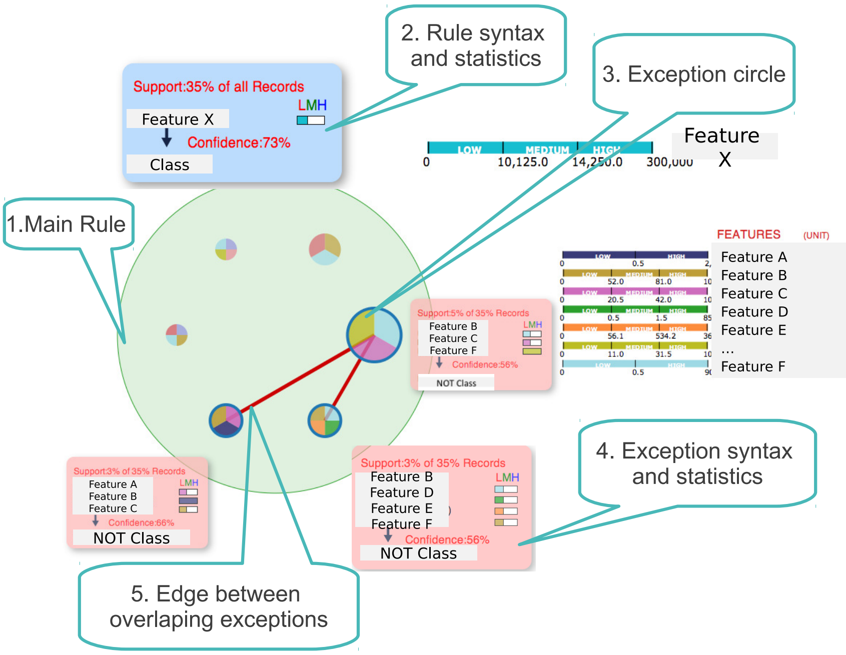
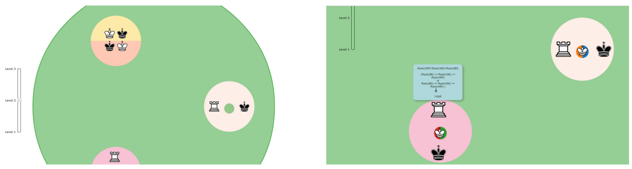


Figure 1: *Rules and Exceptions mined from Real Data*



(a) Level 2, Illegal Positions (b) Level 3, Legal Positions

Figure 2: *King-Rook-King Chess Endgame Rules and Exceptions*

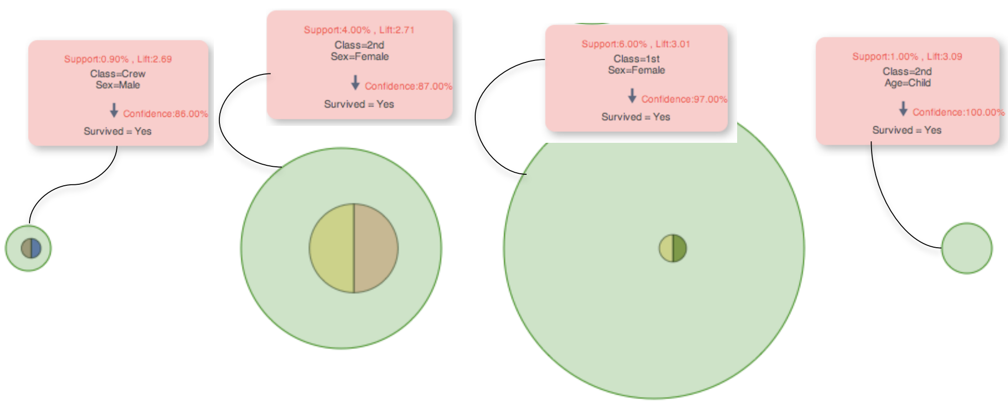


Figure 3: *Survival Rules and Exceptions from Titanic Data*

numerical range. This is shown by a horizontal bar filled up at the appropriate range interval.

Two exceptions can share antecedents as there may be records that satisfy both exceptions. Shared antecedents are easily identified by the same colour in different pies. Record overlap is represented by lines connecting exception circles with the width of the line representing the degree of overlap, label 5 in the figure.

4. Visualizations of Real and Illustrative Datasets

In this section we describe our visualization for four different rule sets, three from literature and one from real-world data.

Real-World Data Set The rule and exceptions shown in Figure 1 have been mined from the data of a real world survey for a product with 6443 rows of records. The attribute names have been obfuscated as the data is proprietary. We used association rule mining followed by summarization based on rule clustering similar to [YCHX] but that produces exceptions as well. The technique, whose details are out of the scope of this paper, reduces 177,094 rules mined to 3 high level rules with multiple exceptions. This set of rules and exceptions illustrates the complexity that can arise in real data.

Our visualization was presented to our primary end-user, a product engineer: First, the fact that our rule summarization engine had whittled down the multitude of rules and exceptions to just a few most interesting ones made the analysis appreciably tangible. Second, regarding the visualization, the appearance of exceptions as 'holes' in the main rule brought out their significance and intuitive semantics. Third, the self-controlled pace at which the user could obtain information about the exceptions made relationships between antecedents more apparent.

King-Rook-King Chess Endgame The next rule set we visualize is from the well-known King-Rook-King chess endgame having two levels of exceptions. The rule set describes which positions of White King, White Rook and Black King are illegal when White has to move next. A position is called illegal if it cannot be reached as the Black King would have been mated in the previous move [SMB92].

The main rule says all positions are legal, except when the White Rook and Black King are in a (1) horizontal or (2) vertical line or (3) when the two kings are adjacent. Exceptions to the rules (1) and (2) occur when the White King is between the White Rook and Black King, so the Black King is not threatened by the White Rook. Exception levels 2 and 3 are shown in Figure 2 (a) and (b).

The number of exception levels and the number of exceptions at a level are apparent from the visualization. To further assist comprehension of the rules, we display instances of data satisfying them using images of chess pieces.

Titanic Data This rule set consists of association rules mined from the survival data set of passengers on the Titanic from [Tit]. There are two sets of rules associating the class, gender and age of passengers with whether they survived or not. Each set has multiple rules implying the same consequent. A sample rule is: $Class = 1$ & $Sex = Female \implies Survived = Yes$, Support: 0.18%, Confidence: 97%. Multiple top-level Titanic rules are shown in Figure 3 for the same consequent ('Survived') ordered by lift.

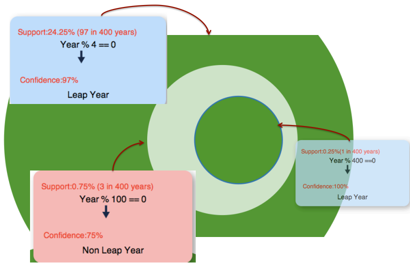


Figure 4: *Gregorian Calendar: 3 Levels*

Gregorian Calendar Rules Finally, we visualize the rules of the Gregorian calendar to decide leap years which have 2 levels of exceptions as explained in [SMB92]: Support and confidence for each rule is based on the occurrence of leap years in a span of 400 years. As the difference in supports of the rules is large, we apply a log scale on support to compute the radii of the circles. The three levels of visualization are shown in figure 4. The number of exception levels and their frequency of occurrence are clearly evident from the visualization.

5. Conclusion

We have proposed a new interactive layout for visualizing rules and their exceptions with the goal of making insights from large data easier to comprehend and communicate. We have illustrated our visualization on multiple rule sets, including those from practice and found it effective in communicating the key messages contained in rule-exception summaries. Future work could include enhancements such as exception circles being ordered based on some interestingness measure. The assignment of colours to antecedents could be used to indicate a property of the data, for example, similar antecedents could be represented by the same hue but different saturation levels to bring out categories of antecedents in the data. We also intend to explore the approach advocated in [Wil99] to connect rule semantics and visual display, including modelling domain-specific depictions as in the chess example.

References

[AIS93] AGRAWAL R., IMIELIÅĎSKI T., SWAMI A.: Mining association rules between sets of items in large databases. In *Proc. ACM SIGMOD International Conference on Management of data* (1993). 1

[BD08] BRUZZESE D., DAVINO C.: Visual mining of association rules. *LNCS Visual Data Mining* (2008), 103–122. 2

[BGB03] BLANCHARD J., GUILLET F., BRIAND H.: A user-driven and quality-oriented visualization for mining association rules. In *Proc. IEEE International Conference on Data Mining* (2003). 1, 2

[CHYN07] COUTURIER O., HAMROUNI T., YAHIA S. B., NGUIFO E. M.: A scalable association rule visualization towards displaying large amounts of knowledge. In *Proc. International Conference on Information Visualization* (2007). 1, 2

[CM84] CLEVELAND W. S., MCGILL R.: Graphical perception: Theory, experimentation, and application to the development of graphical methods. *J. Am. Statistical Assoc. 79* (1984), 531–554. 2

[LHH00] LIU B., HU M., HSU W.: Intuitive representation of decision trees using general rules and exceptions. In *Proc. AAAI* (2000). 1, 2

[Mac86] MACKINLAY J.: Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics 5*, 2 (1986), 110–141. 2

[Shn96] SHNEIDERMAN B.: The eyes have it: A task by data type taxonomy for information visualizations. In *Proc. IEEE Symposium on Visual Languages* (1996). 1, 2

[SMB92] SRINIVASAN A., MUGGLETON S., BAIN M.: Distinguishing exceptions from noise in non-monotonic learning. In *In Proceedings of the 2nd International Workshop on Inductive Logic Programming* (1992), pp. 97–107. 4

[Tit] Titanic data set. `http:/http://www.rdatamining.com/examples/association-rules`. 4

[Wil99] WILKINSON L.: *The Grammar of Graphics*. Springer, New York, 1999. 4

[Yan05] YANG L.: Pruning and visualizing generalized association rules in parallel coordinates. *IEEE Transactions on Knowledge and Data Engineering, 17*, 1 (2005), 60–70. 1

[YCHX] YAN X., CHENG H., HAN J., XIN. D.: Summarizing itemset patterns: A profile-based approach. In *Proc. ACM-SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 314–323. 1, 4