Integrating Time-Series Visualizations within Parallel Coordinates for Exploratory Analysis of Incident Databases

Michael Butkiewicz, Thomas Butkiewicz, Remco Chang
Mbutki@gmail.com

Michael Butkiewicz

University of Pittsburgh

Introduction

Traditionally, in a parallel coordinate plot the time dimension is represented as simply another axis. Only for adjacent dimensional axes are the temporal relationships easily perceivable. We present a method of integrating a ThemeRiver type time-series visualization directly between the axes to ensure clarity of trends of relationships between dimensions over time. We demonstrate the usefulness of our technique for exploratory analysis of a real-life, large (90,000+) incident report dataset from the FAA.

Background

- *ThemeRiver™: In Search of Trends, Patterns, and Relationships* by Susan Havre et al.

  The advantages of the flowing representation of the ThemeRiver design in showing themes over time is discussed in contrast to the sequential splices of data at set time integrals which favor a more document-centric view typically used in things such as a histogram.

- *Parallel Coordinates: A Tool for Visualizing Multi-Dimensional Geometry* by A. Inselberg and B. Dimsdale.

  A methodology for visualizing high dimensional data is presented using the system of parallel coordinates which includes a non-projective mapping between N-Dimensional and 2-Dimensional sets.

Research

The decision to use a ThemeRiver view in between axes of parallel coordinates as opposed to the standard linear connection has created a unique hybrid. Through its use we seek to address several issues that the visualization community has traditionally struggled with when using parallel coordinates.

1) The amount of visual clutter found in standard parallel coordinates increases exponentially with each dimension and data item added. Figure 1 shows a traditional parallel coordinates plot with 3 axes and only 100-200 data items. Figure 2 demonstrates the superior scalability of ThemeRiver, by legibly displaying over 90,000 data items with 3 axes (in addition to time).
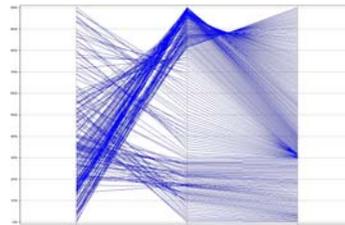

Fig.1. Parallel Coordinates does not scale well for large numbers of data items, as each requires a separate connecting line. Shown here are 100 data items each as a blue line. Notice that occlusion is already becoming an issue.
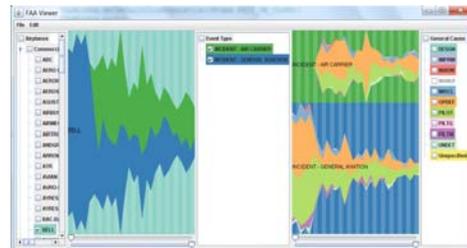

Fig. 2 Our ThemeRiver based view is highly scalable. Here 90,000 data items are displayed without the visualization becoming illegible.

2) Traditional parallel coordinate plots only show the time dimension as a single axis, thus the only temporal relationships visible in the plot are between the time axis and the two adjacent axes. To allow the temporal relationships to exist between each axis, connecting lines have been replaced with ThemeRivers which present the trends between dimensions over time.

3) In order to combine the functionally of the traditional parallel coordinates plot and a ThemeRiver, we implement a two step approach to depicting the river. The first is a "forward flow view" that shows the trends (in frequency) over time of each of the categories-of-interest in the first dimension. This is complimented recursively by the "backward flow view", which illustrates how the categories in the second dimension contribute to the overall trends seen from the first dimension. These two components can be seen in figure 3.
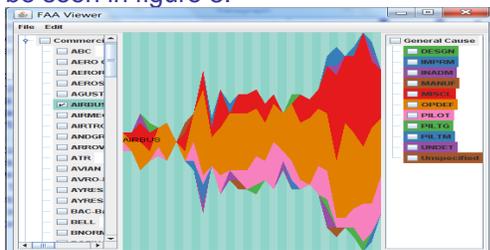
Fig. 3 The 'forward flow' consists of the entire ThemeRiver, showing the number of incidents over time for Airbus (the sole selected category in the 1st dimension.) The 'backward flow' consists of the colored currents within the main river, which show the proportions and trends of categories in the 2nd dimension's in the overall river of the selected category in the 1st dimension.

Impact
The target user of this software is Boeing engineers. This tool allows easy filtering of the incidents to specific Boeing aircrafts, at which point additional axes and plots can be added to explore trends and discover hidden relationships. Making small changes to aircraft maintenance and the manufacturing of airplane related parts has been shown to save companies such as Boeing vast

monetary resources. An example scenario would be to compare trends of landing gear failure between Boeing aircraft used for long duration flights times as opposed to others used to make many short trips. These trends could be used to determine which airplanes require stronger landing equipment to decrease the average time of failure, and to evaluate the effectiveness of such a policy being enacted.

Conclusions
• The ThemeRiver view is far superior for our dataset's large number of data items, due to its inherent scalability.
• Scalability is limited however in terms of how many currents can be displayed within each river. Both due to screen space limits and the number of distinct colors the user can discern. We attempted to overcome these challenges by looping color palettes and allowing the user to select subsets of categories for each dimension.
• The backward & forward flow views are effective for conveying the necessary information; however a method for interactive highlighting of currents is needed to increase inspection capabilities.

Future Work
• Interactive highlighting of elements in the ThemeRivers upon mouse-over of the corresponding names in the adjacent axes. This should greatly improve the backward flow view's legibility and inspection for rivers with excessive numbers of currents.
• Implementation of dual slider bars in between the axes to enable brushing of data items by time. This will allow the range of data moving through to the next axis to be limited to a specified time frame.
• Keyword searches between the axes to filter the data moving through to the next axis, and to display keyword frequency in the ThemeRiver.
• Allowing the user to export the data stream at any point into the IN-SPIRE software to visualize the relationships and clusterings of the free-text narrative sections found in the selected reports.