UNM
Department of
Computer Science

Visualizing Dempster-Shafer Possibility Trees: *a research tool for intelligence analysts*

Nathan Gauntt, Victor Vergara, Steve Smith, Roshan Rammohan, Thomas Caudell

Abstract

Intelligence analysts try to predict behavior of entities under surveillance and estimate the likelihood of certain events occurring, such as terrorist attack. In the real world, it is often difficult or impossible to extract fixed probabilities from evidence gathered by sources, as the source may be unreliable or hostile. Also, analysts must focus their attention on the most probable attack scenarios derived from a combinatorically huge universe of scenarios.

Dempster-Shafer (DS) theory is a method of introducing uncertainty into probabilistic models, combining sequences of uncertain priors into probabilities over sets of events. Logic models, such as logic gate models (LGM), provide a way of compactly representing very large universes of logically related event sequences. In the literature there is active research combining DS theory with LGMs to automate and formalize sequence likelihood estimation. Our research concerns taking these computational tools and ideas from the data mining community, and creating a real-time graphics interface that allows for efficient interactive exploration of the LGM model and the universe of possible event sequences.

Background

Intelligence agencies, such as the Department of Homeland Security, receive disparate pieces of evidence from field sources, of varying credibility and reliability. There are many ad-hoc methods of analyzing this evidence, and the popular commercial toolkits i2 Analyst's Notebook [5] and Sentinel Visualizer [4] overlay functions of evidence, such as 'A is the leader of organization B' or 'B funds C', on top of these social networks as color and geometry changes (fig 1).

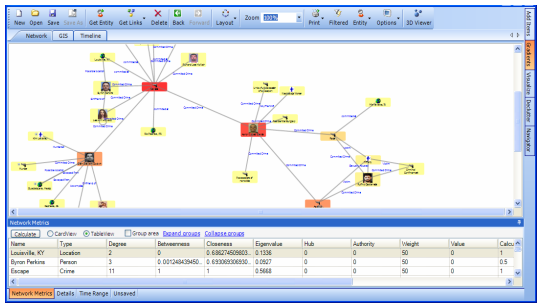

figure 1: Scalar data overlay on social network, from [4]

This technique has significant drawbacks, including:

- Large networks become hard to read without zooming
- Logic of decision making hidden, only output is color

LGM's address the second point, as they make explicit the logic reasoning behind, in our case, the calculation of an interval $(Belief, Plausibility)$, where X is an event and: $Belief(X) \leq Probability(X) \leq Plausibility(X)$.

Given that our knowledge representation is an event tree instead of a social network graph, we still must address the problem of visual analysis, on trees (and more generally dags) instead of on generalized networks. The literature has used hierarchical tree clustering to analyze complex problems for a long time, such as are used in biology (fig 2).

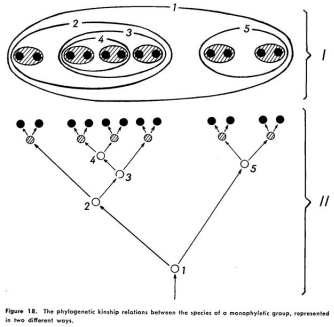

figure 2: Hennig cluster tree (bottom) and ellipse clustering (top), from [6]

More modern tree visualization techniques, such as circle trees [7] and treemaps [1,2] use computational layout algorithms to use screen space efficiently, in order to display representations of very large data sets (figure 3). In the case of polygon treemaps [2], one can make theoretical guarantees on the ratio of representation area displayed versus (normalized) data dimension weight.

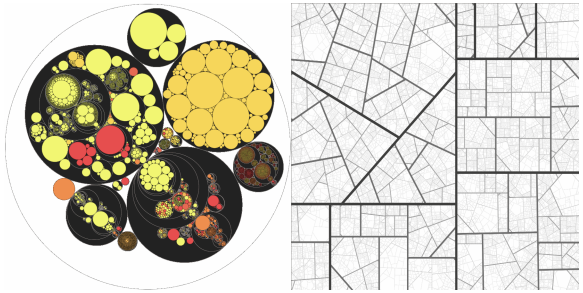

figure 3: circle trees (left), and polygonal treemap (right), from [7,2]

These techniques fare better for display of large, nested data such as data in our event trees, and are also amenable to zooming in a way that preserves global structure while looking in detail at a piece of local substructure. More recently, in [8], dendrograms are being used to investigate data exploration and compression using group theory and Haar wavelet analysis, indicating the ongoing usefulness and mathematical depth of this metaphor.

Results and Discussion

The current prototype interface (fig 4) has black and grey bars representing belief and plausibility of a chain of events happening. In probability, this would be a product of event probabilities, if the events are independent. In DS theory, the operation is more involved, and produces a range.

figure 4: event chains before non-linear scaling (left) and after (right)

Non-linear scaling is applied when the cursor hovers over an area with densely packed (and possibly overlapping) ranges. The space distortion is greater near the cursor and zero for a threshold distance away (orange box).

I hope to have motivated the use of dendrograms as robust and flexible visual metaphors. Future work will include an OpenGL implementation of a real-time interactive interface. Also, the ability to sort event chain scenarios by metrics other than tree topology will be explored, such as sort by actor, method, intent, and consequence.

References

[1] "Tree Visualization with Treemaps: a 2-D Space-filling Approach", Shneiderman, B; *ACM Transactions on Graphics, vol 11, 92-99*, January 1992

[2] "Circular Partitions with Applications to Visualizations and Embeddings", Onak, K; Sidiropoulos, A; *24th Annual Symposium on Comp. Geom.*, 2008

[4] "Sentinel Visualizer Website", *http://www.fmsasg.com/products/SentinelVisualizer*, (accessed 2010)

[5] "What's New in i2 Analyst's Notebook 6.0.55 White Paper", *http://hill-assoc.com/content/New in i2 Analyst's Notebook 6.0.55 (UK).pdf*, 2005 (accessed May 2010)

[6] "Phylogenetic Systematics", Hennig, W; *Univ. of Illinois Press*, (1966)

[7] "CTreemap project", *http://lip.sourceforge.net/ctreemap.html*, (accessed 2010)

[8] "The Haar Wavelet Transform of a Dendrogram", Murtagh,F.; *Journal of Classification, 24, 3-32* 2007

*This work was funded in part by a grant from the Department of Homeland Security