Dynamic Topic Analysis using VisualDTA

Andrew Kurtz
ajkurtz (at) indiana.edu

Abstract

Within a conversation, even when there are only two participants, the discussion topic changes over time. It is interesting to analyze the evolution of topics within different types of conversation-such as IRC, IM, and message board conversations-to observe how they are similar or different. Herring (2003) has created a technique for analyzing the coherence of conversation within computer-mediated discourse. This technique is called Dynamic Topic Analysis (DTA) and it provides a way to quantify and visualize the structure of the topic flow within a conversation. A coding scheme is used to quantify the topic drift within a conversation and a visualization based on the coding may be created to make it easier to see the flow of the topics within the conversation.

This presentation focuses on the development and use of an interactive Java application, called VisualDTA, which is used to automatically create a visualization of a DTA coding. In addition to displaying the visualization, VisualDTA provides interactive tools to enhance the topic analysis process. This research is being conducted by Andrew Kurtz and Susan Herring and is based on the visualization described in Dynamic Topic Analysis of Synchronous Chat (Herring, 2003).

Utilizing VisualDTA to analyze a conversation involves creating a DTA coding. Typically the coding would be created in Microsoft Excel and exported to a TAB delimited file. The coding file is loaded into VisualDTA and the visualization is created and displayed. The display is a tree with the root at the top and the children flowing down and to the right. The passing of time in the conversation is represented going down the y-axis. Moving right on the x-axis represents the semantic distance of how off-topic the proposition is from the previous proposition. The propositions are represented by a letter that indicates the relationship between the current proposition and the proposition to which it is responding, called the move (T for "on-topic", P for "parallel shift", E for "explanation", M for "metatalk", and B for "break"). Propositions that are in reply to a previous proposition are connected to the proposition to which it is responding using a line. A dotted line is used when the connection is tenuous.

When the visualization has been displayed, many interactive options are available.

- The visualization may be displayed step-by-step as the conversation unfolds.
- A short description for each proposition may be displayed.
- Hovering over a proposition with the mouse will display information about that proposition, including any attributes coded such as speaker, gender, and role.
- Basic statistics can be shown that includes the average semantic distance and counts for proposition attributes, such as move type, speaker, and gender.
- Custom statistics can be shown to relate two attributes. Such as showing gender by role.
- Clicking on a proposition allows the user to highlight all of the propositions that match an attribute for the selected proposition. For example, all the propositions by a specific speaker can be highlighted.

The added interactivity provided by the VisualDTA program enhances the analysis that may be performed using the DTA technique. We are continuing this research and are including more analysis tools within the application. One of the challenges for the future is to refine the visualization to improve the display of the relationships between propositions. Currently, the lines connecting propositions may overlap making it difficult to determine the proposition to which it is responding.

References
