Applying Sequence Alignment for Analyzing User Interactions with Geovisualizations

Amy L. Griffin
School of Physical, Environmental and Mathematical Sciences
University of New South Wales@ADFA

Introduction

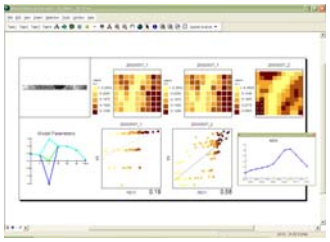
Since the late 1980s, GIScientists have directed attention to developing interactive geographic representation and analysis tools. Proponents of such tools have promoted their ability to support scientific investigations of a variety of phenomena. However, there have been few empirical studies that demonstrate the mechanisms through which, and the degree to which, interactivity facilitates visual thinking and knowledge construction, and little attention has been paid to how interactive tools aid in hypothesis generation and simulation model refinement. Two important questions are:

1. Do the representations a model user sees have an impact upon his or her conceptualization of the modeling problem?
2. Do the displays that the model user has seen in the past (which may have been influenced by his or her training) influence the types of representations s/he chooses for viewing the model results?

This study begins to provide evidence towards answering these questions by characterizing how experts use interactive data-display devices within the context of simulation modelling.

Test Instrument and Data

The test instrument was a process-based model of Hantavirus Pulmonary Syndrome (HPS) risk that was implemented in ArcGIS. Users could interact with data inputs and outputs through three types of visual display: linked maps and scatterplots, time series graphs and the model parameters graph, which is a representation of the model parameters that were used to generate the results that are currently loaded in the display.



The study used a modified version of verbal protocol analysis that asked participants to make self-evaluative statements about their work with the tools as well as to verbalize what they were attending to and thinking about.

The model use session included practice thinking aloud about an unrelated problem, a series of three structured tasks that introduced participants to three methods of approaching the model exploration process, and 45 to 90 minute session of free-form exploration of the case study problem using the model and data display devices.

Research Questions:

The larger project that this research fits into seeks to answer four questions:

- What are the patterns of use for different system components?
- What kinds of information to users attend to in the visual information display devices?
- How do participants obtain information from the system and how is this information used?
- What kinds of hypotheses are generated?

To answer these questions, I identified quotations from each of the transcripts of the verbal protocols from the model use sessions and coded them using four coding schemes designed to capture information that could help to answer the four research questions.

The coding process, then, produces multiple sequences of actions, attention, cognitive operation and hypothesis type, which is itself broken down into four attributes (complexity, basis, goal and relationship hypothesized). These sequences, as a whole, provide a rich record of the user's interactions with the system. The challenge, then, it to make sense of the complexity.

Method:

Sequence alignment is a method that was developed in the 1980s by bioinformaticians for analyzing DNA sequences. Since this time, the methodology has been adapted for social science and geovisualization applications (e.g., Wilson 1998; Shoval and Isaacson 2007; Fabrikant et al. 2008). In contrast to step-by-step sequence analysis methods (e.g., time series, Markov chain or survival analysis), whose emphasis is on trying to elucidate a causal chain of events or transitions between events, sequence alignment is essentially concerned with finding patterns among whole sequences or subsequences. In other words, it is concerned with grouping sequences that have similar patterns of events. In the context of this research, these groupings can be considered to be groupings of geovisualization users who have common patterns of behavior or thought.

Sequence alignment works by calculating a similarity metric for every pair of sequences being examined. This similarity metric is a measure of how many operations have to be carried out to transform one sequence into another. The basic operations that are used to transform sequences are: insertion, deletion and substitution.

The example below shows an alignment between the characters in the words 'map' and 'graph'. This alignment involves two insertions and one substitution.

```
map
graph
-map-
```

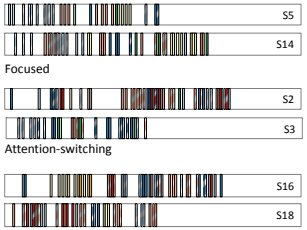
In the example above, it takes three operations to transform 'graph' into 'map': 1 substitution ('m' for 'r'), and two insertions of gaps at the beginning and end of the word.

The pairwise alignments are then generalized to find the most optimal alignment for all sequences being examined and clustered to find groups of similar sequences.

This research used ClustalTXY, software that was generalized from the molecular biology package ClustalW to accept multivariate sequences that contain more characters than the base pairs found in DNA)as inputs.

Results: Visual Analysis versus Sequence Alignment

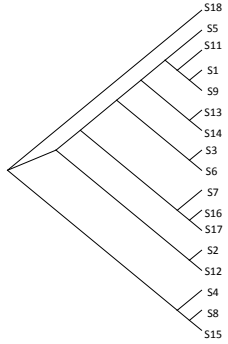
Sequence alignment shows some promise for identifying groups of users who employ similar data or model exploration strategies. Before identifying sequence alignment as a potentially helpful method for analyzing user interaction sequences, I employed a simple visual analysis of the patterns to identify groups of individuals with some sequence similarities. In this analysis I described three groups: focused attention, attention-switching and mixed focus and attention switching.



'Focused attention' users have clustered patterns of attention directed to one display device followed by additional clusters of attention directed to other displays. Overall, they give a particular device more attention before switching devices than users with other attention types.

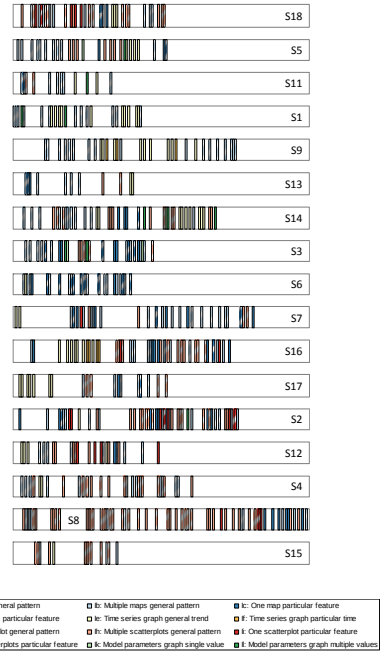
'Attention-switching' is marked by rapid alternation of the user's visual attention between either one or more display devices or aspects of a particular device (e.g., between a general pattern and specific feature within one display type).

Users who I classified as having a 'mixed focused and attention-switching' behavior exhibit characteristics of both of these behaviors, but at distinct periods of the session, often at the beginning or the end (when they have switched behaviors).



The sequence alignment also identified three groups, shown in the cluster-tree above. Although there is some similarity of group membership (e.g., S5 and S14 both fall in the same group), the sequence alignment is able to move beyond just the temporal patterns of code occurrence, and which were the focus of my original visual analysis to also account for and consider the frequency of code occurrence.

The graphic at the right, when compared with the cluster-tree shows that groups are comprised of individuals with similar attention patterns. For example, users S4, S8 and S15 relied more heavily on scatterplots than other representations and did not tend to flip from one device to another at high frequencies, while S18 (at the top), who also relied heavily on scatterplots and is in a different cluster, does exhibit rapid attention-flipping.



Conclusions:

Sequence alignment has the potential to provide a quantitative measure of similarity of behavioral patterns of geovisualization users. By explicitly accounting for both the frequency of a behavior and when in the sequence the behavior occurs, it can provide a more nuanced understanding of the strategies that geovisualization users take to accomplishing tasks. While the interpretation of cluster groupings becomes more difficult as the coding scheme contains a larger number of different types of actions, this method can be very useful for comparing relatively simple sequences.

References:

Fabrikant, S.I., Rebich-Hespanha, S., Andrienko, N., Andrienko, G. and D.R. Montello. (2008). "A Novel Method to Measure Inference Affordance in Static Small-Multiple Map Displays Representing Dynamic Processes." *The Cartographic Journal*, 45(3): 201-15.
Shoval, N. and M. Isaacson. (2007). "Sequence Alignment as a Method for Human Activity Analysis in Space and Time." *Annals of the Association of American Geographers*, 97(2): 282-97.
Wilson, C. (1998). "Activity pattern analysis by means of sequence alignment methods." *Environment and Planning A*, 30: 1017-38.