Finding Patterns in Temporal Data

Krist Wongsuphasawat[1,2], Taowei David Wang[1,2], Catherine Plaisant[2], Ben Shneiderman[1,2]
Department of Computer Science[1], Human-Computer Interaction Lab[2]
Contact: {kristw, tw7, plaisant, ben}@cs.umd.edu
www.cs.umd.edu/hcil/temporalviz

Temporal categorical data is a collection of records that contain *timestamped events*, such as (11/24/2009,Flu), (11/30/2009, Stroke), etc. Temporal categorical data is increasingly being collected by many organizations. Health organizations have Electronic Health Record (EHR) databases containing millions of records of patient histories. Each patient history contains hospital admissions, patients' symptoms and treatments. Transportation systems generate logs of incidents (notifications and arrivals of each unit on the scene). Temporal categorical data also includes student reports, web logs, financial histories, market baskets and data in many other domains.

These temporal categorical databases contain many patterns. For example, using EHR database that keeps track of patient transfers in the hospital, our physician partner wants to find "bounce backs" patients, which are patients in Intensive Care Unit (ICU) who seemed to get better so the physicians transferred them to normal rooms, or "Floor". However, they were transferred back to the ICU within 48 hours. These situations impact the quality of care and are undesired by the hospital. Therefore, the physician wants to find these patients to analyze what might have been the reasons that led to those transfers. There has been a long history of visualizing temporal categorical data, from visualizing intensive details of a single record to visualizing and searching for multiple records. Last year we presented two interactive search and visualization tools, LifeLines2 and Similan, and have been using these tools to support the physician in finding patients with specified patterns.

By working on these case studies, we discovered a need to explore the sequential patterns in the data. For example, a physician might ask, "What usually happened to the patient after they arrived the hospital?" Many visualizations do not support this task well due to a scalability issue. When there is a large number of records, users cannot see all records on

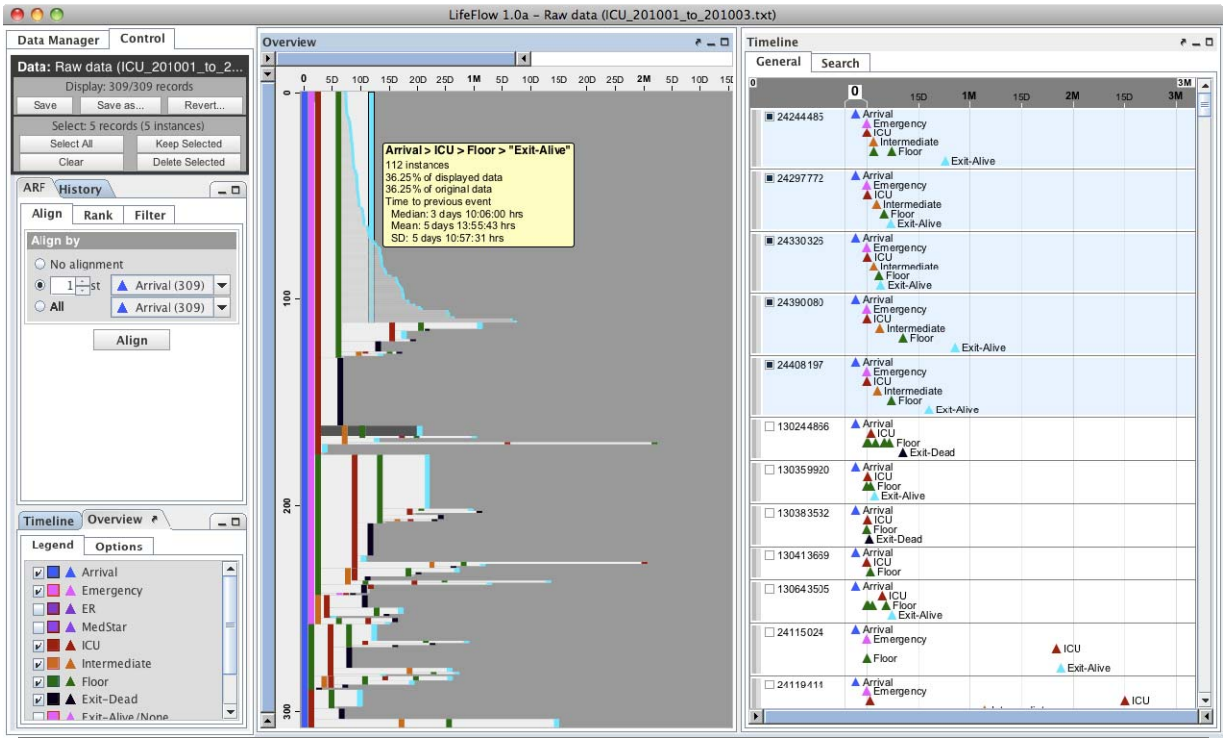

Figure 1. A working prototype, LifeFlow 1.0a –The control panel is on the left. In the middle is the LifeFlow visualization, showing patient transfer patterns from the time of arrival. On the right is a view inherited from LifeLines2, showing each patient in details. The selected path on LifeFlow is highlighted in dark gray and all records in that path is selected and highlighted in light blue on the LifeLines2 view.

the screen at once, which makes it difficult to recognize any pattern. A few scalable visualizations were later developed, but these techniques aggregate the records in a way that the sequential information is obscured. According to the information visualization mantra: "overview, zoom and filter, then details-on-demand", an overview of sequential patterns is needed.

LifeFlow: An Overview Visualization

Therefore, we have recently developed a new visualization called *LifeFlow* to provide a scalable overview of sequential patterns in temporal categorical data. Our technique aggregates multiple records into a tree, based on sequence of events, and visualizes this tree with the LifeFlow visualization.

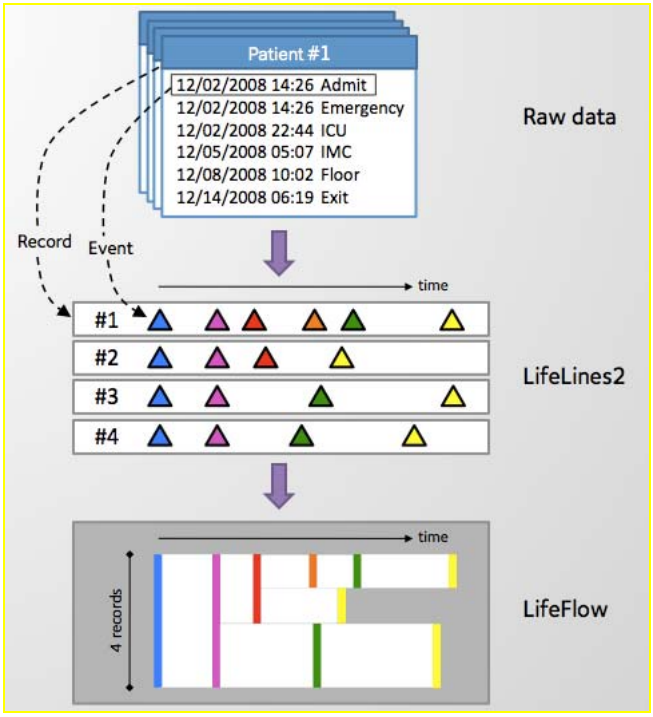

Figure 2. A diagram showing four records of raw data and how they are visualized as LifeLines2 and LifeFlow visualization.

LifeFlow uses a color-coded rectangle, or node, to represent an event. The height of each node represents the number of records. The first node (blue) and second node (pink) in Figure above has 100% height because all records fall into this node. The horizontal axis is used as a timeline. All nodes are placed on a horizontal position according to the representative time gaps between them and their previous nodes. (A mean is currently used by default.) An edge between each pair of nodes contains a distribution of the time gap between the two nodes, which will be overlaid on the nodes when users move their cursors over. A working prototype LifeFlow 1.0a inherited several features from our previous work on LifeLines2, including the Align, Rank and Filter (ARF) Framework and visualization of the records. By connecting these components together, user now can explore the data following the information visualization mantra: see an overview from the LifeFlow visualization; use the ARF Framework and other interactions to zoom and filter; then select any path on LifeFlow to see more details on demand.

Ongoing Case Study

We are conducting a multidimensional in-depth long-term case study with our physician partner at the Washington Hospital Center by helping him analyze patient transfers. An example can be seen from Figure 1. LifeFlow can easily show the common transfer pattern as well as point out the uncommon cases. It also shows the average time that the patients were transferred from one room to another. The preliminary result shows a promising opportunity to use LifeFlow to monitor the Emergency Department (ED) performance over time.

Conclusion and Future Work

We introduced a new visualization called LifeFlow that provides an overview of sequential patterns in temporal categorical data to support users' exploration, and reported an ongoing case study with the physicians. In the mean time, we are working on including additional features, such as displaying statistical information or comparison between two or more data sets. While the examples given and case studies described are medical use cases, we believe our technique can be applied to many other fields, where temporal categorical data is the main focus. We are eager to conduct new case studies with experts in other domains.

ACKNOWLEDGEMENT

We would like to thank the Washington Hospital Center, NIH grant CA147489 and Center for Advanced Transportation Technology (CATT Lab) for their support.

PAPERS

1. Wang, T., Plaisant, C., Quinn, A., Stanchak, R., Shneiderman, B., and Murphy, S., Aligning temporal data by sentinel events: Discovering patterns in electronic health records, *Proc. CHI 2008*, ACM, 457-466
2. Wang, T., Plaisant, C., Shneiderman, B., Spring, N., Roseman, D., Marchand, G., Mukherjee, V., Smith, M., Temporal Summaries: Supporting Temporal Categorical Searching, Aggregation and Comparison, *Proc. IEEE Information Visualization 2009*, IEEE, 1049-1056
3. Wongsuphasawat, K., Shneiderman, B., Finding Comparable Temporal Categorical Records: A Similarity Measure with an Interactive Visualization, Proc. IEEE Symp. Visual Analytics Science and Technology 2009, IEEE, 27-34