Temporal Pattern Discovery Using Lifelines2

Taowei David Wang, Catherine Plaisant, Ben Shneiderman

Human-Computer Interaction Lab, University of Maryland, College Park, MD

ABSTRACT

Temporal data often reveals interesting patterns. In the past two years, we have applied our open-source temporal categorical data visualization system Lifelines2 to eleven different case studies to assess its usefulness and usability in real life work of the domain experts. We report the different levels of its success in the following five stories.

KEYWORDS: Temporal visualization, Lifelines2

INDEX TERMS: H.5.2 [User Interfaces]

1 INTRODUCTION

Temporal data are interesting to analysts for a couple of reasons: (1) Temporal ordering of events may reveal potential cause-and-effect relationships. (2) The existence of certain temporal pattern of events is often indicative of deviation from normal process, and can be a cause for concern. While many tools facilitate the logging, storage, retrieval, and sharing of temporal data, novel visual analytic techniques are not as prevalent. We proposed Lifelines2 to help analysts study time-stamped event data in personal histories [4]. In the last two years, we have partnered with physicians, hospital administrators, digital library administrators, student academic progress reviewers, and law enforcement officers to study their data using Lifelines2. This paper is about their stories.

2 VISUALIZATION AND INTERACTION DESIGN

Lifelines2 is designed to visualize multiple records of temporal categorical data. It supports three main tasks. First, the main time line visualization and the *alignment* operator support examination of temporal ordering of events [4]. Analysts can use alignment to change the focus of the visualization, bringing events central to their interest to the center and placing all other events relative to this focus. As a consequence, analysts can better see what events tend to occur before and after with respect to this central event across multiple records.

Secondly, analysts can use *sequence filters* to find records with the specified temporal patterns [5]. Search results can then be saved as groups for later reference. Thirdly, *temporal summaries* support discovery of event trends over time and comparison of saved groups of records [6].

We present how these features help our collaborators gain insights that would otherwise be difficult to obtain, improve their current analysis process, and help them accomplish goals in their own field.

3 DISCOVERING INCONSISTENT DATA

One of the most common results of applying visualization to real world data that had never been visualized before is the discovery of inconsistent data. While this type of discovery usually does not yield scientifically important outcome, it reveals the messiness of real world data. By reviewing the visualization, analysts can often identify the inconsistencies and decide how to deal with them. It is often imperative to clean or remove these inconsistencies, otherwise any automatic analysis methods – such as statistical or data mining techniques – would yield incorrect and misleading results.

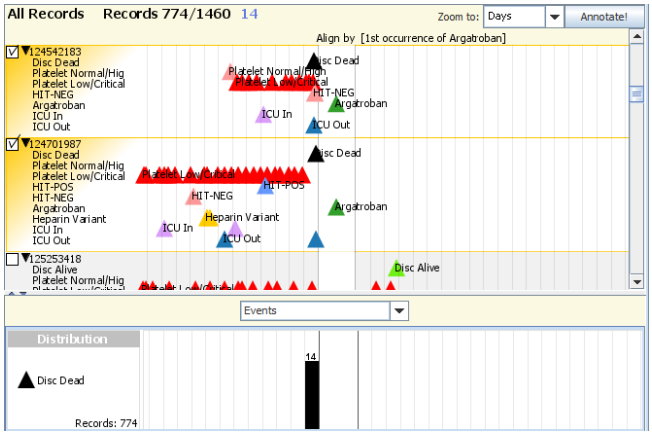

Figure 1. After aligning by `Argatroban` (dark green), the visualization shows two records with `Argatroban` within one day after their `Disc Dead` event (black). The temporal summary at the bottom shows all 14 such occurrences. The data is de-identified to preserve privacy.

In one of our first case studies, physicians from the Washington Hospital Center were looking to find patients with heparin-induced thrombocytopenia conditions. In one of the related queries, the physicians applied Lifelines2 to search for patients who were given the drug argatroban in relation to their heparin-induced thrombocytopenia conditions. Dr. Greg Marchand, Dr. Vikramjit Mukherjee, and Dr. Mark Smith discovered that some patients were given argatroban *after* they have been discharged dead. By aligning by `Argatroban` and displaying the distribution of `Discharged Dead`, the physicians were able to identify that there were 14 patients who exhibited this unlikely sequence in the 2008 calendar year, and that in all these cases, the drug was given within 24 hours after death (Figure 1). This eventually led the physicians to conclude that there was a systematic delay in the medication database, and it might be caused by the fact that the medication database needed to be routed through the billing database first, which, in turn, may involve health insurance carriers. This discovery led the physicians to not use the argatroban events in their subsequent analyses reported in [6].

4 VISUAL DISCOVERY OF PATTERNS

The International Children's Digital Library is a collection of freely available online library of over 4,000 children's books in over 50 languages. The administrators of the library website (Anne Rose and Dr. Ben Bederson) and a visiting scholar (Rong

The authors are with the Human-Computer Interaction Lab, 2117 Hornbake Building, South Wing, University of Maryland, College Park, MD 20742. E-mail: {tw7, plaisant, ben}@cs.umd.edu

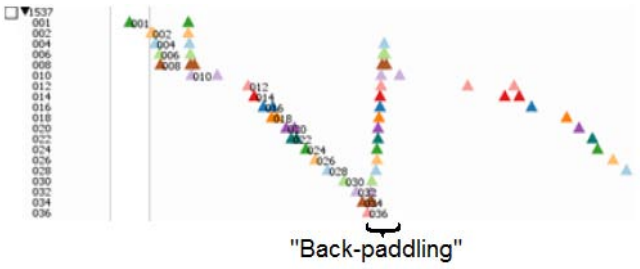

Figure 2. The visitor exhibits the "back-paddling" pattern before reading forward again.

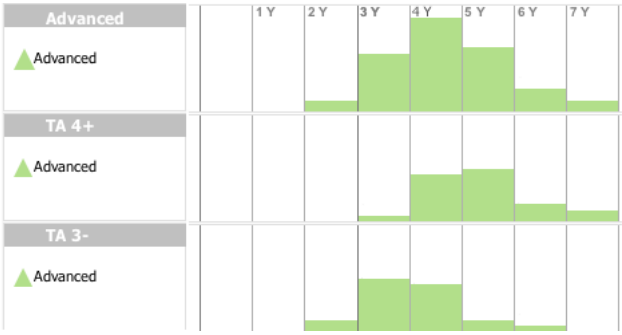

Figure 3. Advancement to candidacy since admission. Those who TA 4+ semesters tend to take on average 1 year longer.

Chen) decided to analyze visitors' reading behavior using web server logs. We helped them map each visit to a record in Lifelines2 and each visit to a book page to an event. In this preliminary study, they focused only on one of the most popular books in one week in October 2008. There were over 23,000 unique visitors and 336,000 page visits in that period. Each visitor's page-reading sequence is visualized in Lifelines2 (Figure 2). Each category represents a page, and the visits to these pages are plotted on the linear time line. Only even pages are used because each web page shows a spread of two book pages.

The analysts immediately noticed a few interesting, unexpected patterns. The most prominent one is the "back-paddling" pattern (Figure 2). This pattern shows the visitor successively and quickly pressing the "back" button to traverse backwards in the book. This prompted an interesting question to the web site designers – why do users prefer "back-paddling" instead of using an always-accessible book overview that shows all pages and provides quick access to any desired page? Was it a question of lack of affordances or lack or training?

In addition to potential design questions that arose out of the visual inspection of data, the log analysts also found that many visitors, surprisingly, do not begin reading from the first page. Instead, many visitors skipped to the first page with text. This is surprising because children's books are full of rich illustrations in almost every page, offering enjoyable content even when it is without text. These visual discoveries helped the log analysts define what a Book Reading Session is. The analysts then used the definition to perform subsequent analysis, which was later published in the European Conference on Research and Advanced Technology for Digital Libraries '09 [1].

5 Finding Support for Hypotheses

Sometimes analysts already have a hypothesis in mind, and would like to see if there is credible support in the data before additional analysis such as hypothesis-testing statistical tests are performed. We describe how Lifelines2 helps analysts to deal with this type of directed discovery.

5.1 Monitoring Academic Progress

Graduate students undergo review by faculty members to assess their progress annually in the Department of Computer Science, University of Maryland. This process seeks to identify students who show signs of falling behind, and once identified, faculty would recommend the students and their academic advisors to take appropriate actions. Completion of coursework, advancing to candidacy, and publications are some of the metrics our faculty uses. However, reviewers are also interested in developing more metrics. After reviewing students for a few years, Dr. Neil Spring developed a new hypothesis. He hypothesized that students who served as a teaching assistant (TA) for four or more semesters takes longer to advance to candidacy than those who serve three or fewer. While it sounds plausible, no one had investigated whether our data supported the hypothesis, and that if there were support, no one knew how long the delay was.

Using Lifelines2, Dr. Spring separated students who had advanced to candidacy into 2 groups: TA 4+ (those who TA'ed for four or more semesters), and TA 3- (the complement group). He then aligned by the students' admission time, and plotted histograms of their candidacy advancement in the original group and the two subgroups (Figure 3 (counts removed for privacy reasons)). Our data supported his hypothesis. Students tended to advance to candidacy one year later if they had TA'ed four or more semesters. The faculty said that while he could not infer causal relationship here, it demonstrates that TA assignment is one potentially useful predictor of student progress that he would consider using in the future.

5.2 Insufficient Triage Incidents

Dr. Phuong Ho from the Washington Hospital Center heard anecdotal evidence from nurses that the number of "step-up" cases have increased. "Step-ups" describe patients who are admitted to the hospital in a lower-level care room and then escalate to a higher-level care room within 24 hours. This may be indicative of insufficient initial triage. Using Lifelines2, Dr. Ho used sequence filters, alignment, and temporal summaries to search for these step-up cases. He applied the analysis on a quarterly basis for three year's worth of data (2007-2009). He and his colleagues were able to obtain accurate count of step-up cases when it was difficult to do so using their current electronic health record system. After obtaining and analyzing the results, the physicians concluded that the number of "step-up" cases has indeed been rising, but it is because of the increase of hospital admissions. The percentage of "step-up" cases has actually been in decline. This analysis helped the physicians perform a historical analysis quickly, refute a hypothesis, and avoid a potentially costly protocol change.

Because the speed and ease at which the physicians were able to formulate their query and execute them in this case study (and a similarly structured case "bounce-back", Dr. Ho recommended the use of Lifelines2 over existing hospital methods. The idea is that if the analysis can be performed quickly and flexibly, more quality assurance metrics can be developed.

We are currently assisting the physicians to perform these analyses on a quarter-by-quarter basis. We are comparing the Lifelines2 results with the hospital's current method's results to make sure that there are no overlooked issues. The goal is to eventually phase out the current methods and rely on Lifelines2 only.

5.3 Hospital Protocol Change and Its Impact

In November of 2009, Washington Hospital Center's protocol that provides guidance on when bi-level positive airway pressure (BiPAP) devices are used on pulmonary patients was changed. Dr. Ho and his colleagues wanted to see if the protocol change indeed impacted care of patients positively. They came up with a series of patterns that describe patients with escalating condition due to a BiPAP application, and applied the search in Lifelines2. By comparing the patient population before and after the protocol change, they found that the escalation rate was reduced from 20% to 12.9%. This very preliminary result showed that the new protocol was effective.

Because there are no existing methods to perform this analysis, the only way to verify the correctness of the analysis using Lifelines2 is by having someone perform a manual review of the patient records. There are over 600 patient records in this case study, and each patient record takes at minimum three minutes to review, it is expected to take a long time before that process ends. Dr. Ho and we are currently waiting on that result to move forward.

6 Current Impact and Future Work

6.1 Adoption of Lifelines2 in i2b2

Lifelines2 allows users to make unexpected visual discoveries and find support for hypotheses. Although the extent of success reported here is modest, the potential is clear. As a result, the i2b2 clinical research platform from Partners HealthCare, adopted in over 50 research sites [2], has adopted Lifelines2 as a visualization plug-in in its upcoming releases [3]. Institutions that use i2b2 include many national health organizations such as Case Western Reserve Hospital, Harvard Medical School, Kaiser Permanente Health, and several international institutions such as Georges Pompidous Hospital (Paris, France), and University of Seoul (Seoul, South Korea). Lifelines2 will be released as an optional plug-in under the name of Time Align. Users will have full functionality of Lifelines2 in i2b2.

Before Lifelines2 was integrated as a plug-in, the i2b2 research community had found it useful in their research of rheumatoid arthritis. Rheumatoid arthritis is notoriously difficult to diagnose. Because there are many other conditions that show similar symptoms, physicians typically use many tests in many meeting with patients to ascertain that it is indeed the problem. In one phase of the rheumatoid arthritis research project, the researchers needed to recruit patients who had rheumatoid arthritis. Looking at the medical records for rheumatoid arthritis diagnoses was not very fruitful because there were many false positives.

After much investigation, the researchers found that by first *aligning* patients by their first visit to specialized rheumatoid arthritis clinics and restrict only those who have rheumatoid arthritis diagnosis events after, the researchers are much more likely to find true positives. The reason being the diagnoses from these clinics are much more reliable.

The i2b2 research community uses many other commercial tools such as Excel and Spotfire for their analysis. The community is no stranger to information visualization: i2b2 has plug-ins to the original Lifelines and Treemap. However, Lifelines2 brings enough useful techniques to the table that i2b2 welcomes it as an addition.

6.2 Impact to Washington Hospital Center

Lifelines2 is continually being used in collaboration with Washington Hospital Center physicians. The physicians are enthusiastic about using Lifelines2 in developing hospital metrics, quality assurance measures, and health care management. They look forward to a day where every physician has a copy of Lifelines2 to interact with hospital data.

However, despite the success stories already reported here, there are real-world barriers that make direct adoption difficult. There is a difficulty in obtaining data and transform them into Lifelines2 visualization – siloed databases, unexposed database schemas, case-specific mapping of events, etc.

We are helping the physicians to overcome some of the problems by developing tools that they can go from a database search results to Lifelines2 in fewer steps.

6.3 Next Steps

Aside from helping the physicians get around technical issues, we are researching and developing extensions to Lifelines2. We are experimenting with more flexible temporal query methods (similarity-based search), tree-based aggregation and visualization of temporal sequential events, and interplay of temporal and non-temporal data. We hope to integrate these new advances in the successor of Lifelines2 in the near future.

References

[1] R. Chen, A. Rose, and B. Bederson, How people read books online: mining and visualizing web logs for use information, *Proceedings of the European Conference on Research and Advanced Technology for Digital Libraries*, 364-369, 2009.

[2] i2b2 Academic Users' Group, https://www.i2b2.org/work/aug.html

[3] S.N. Murphy, G. Weber, M. Mendis, V. Gainer, H.C. Chueh, S. Churchill, and I. Kohane, Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2), *Journal of American Medical Informatics Association*, 17, 124-130, 2010.

[4] T.D. Wang, C. Plaisant, A.J. Quinn, R. Stanchak, S. Murphy, and B. Shneiderman, Aligning temporal data by sentinel events: discovering patterns in electronic health records, *Proceedings of the Conference on Human Factors in Computing Systems*, 457-466, 2008.

[5] T.D. Wang, A. Deshpand, B. Shneiderman, A temporal pattern search algorithm for personal history event visualization, *IEEE Transactions on Knowledge and Data Engineering*, forthcoming.

[6] T.D. Wang, C. Plaisant, B. Shneiderman, N. Spring, D. Roseman, G. Marchand, V. Mukherjee, and M. Smith, Temporal summaries: supporting temporal categorical searching, aggregation, and comparison, *IEEE Transactions on Visualization and Computer Graphics*, 15(6), 1049-1056, 2009.