Visual Exploration of Temporal Data in Electronic Medical Records

Josua Krause[1], Narges Razavian, PhD[1], Enrico Bertini, PhD[1], David Sontag, PhD[1]
[1]New York University, New York, NY

Introduction and Background

Electronic medical records (EMRs) and administrative data contain a large number of distinct events happening in the life of a patient, including diagnoses, procedures, medications, and lab test results over time. Visualizing this data would be useful for clinicians to quickly comprehend a patient's past medical history, and for researchers to design predictive models, better understand chronic diseases, comorbidity patterns, and contributors of health care cost.

Methods

For the visual representation of a patient's history we display every event as a dot (see Figure 1a), similar to *LifeLines*[1]. The color of the dot represents the group of the event, i.e. a diagnosis, a performed procedure, a lab-test result, a prescribed medication, etc. The horizontal position of a dot shows the day, whereas the vertical position distinguishes between different types of events. A significant contribution is to sort the types by the first occurrence in the history of the patient. This order is helpful for using the data to form causal hypotheses. Additionally, it helps with identifying major incidents (indicated by a steep rise) and allows for the visualization of a larger number of types than in previous work. Since the space for labeling events by type is limited, we use this fact in combination with the cost of an event (indicating the importance) to select which labels to display. Additionally, a user can use an interactive "lens" to hover over events of interest to show their labels (see Figure 1b). To reduce vertical space requirements we also provide different levels of aggregation according to the CCS hierarchy for diagnosis and procedures. The software is web based using a flexible format allowing for multiple datasets as input. We developed the system using health claims from a private insurer, and also the semi-synthetic data set from CMS[2].

Results and Discussion

The tool proved to be helpful for analyzing the output of machine learning algorithms in predictive healthcare analysis. It provides a quick way to identify the problems of an automatically generated model by looking at patients that were classified wrong or turn out to be outliers in the cohort. Additionally, the tool enables clinicians to quickly "tell the story" of the patient using just the administrative data often found in insurance claims.
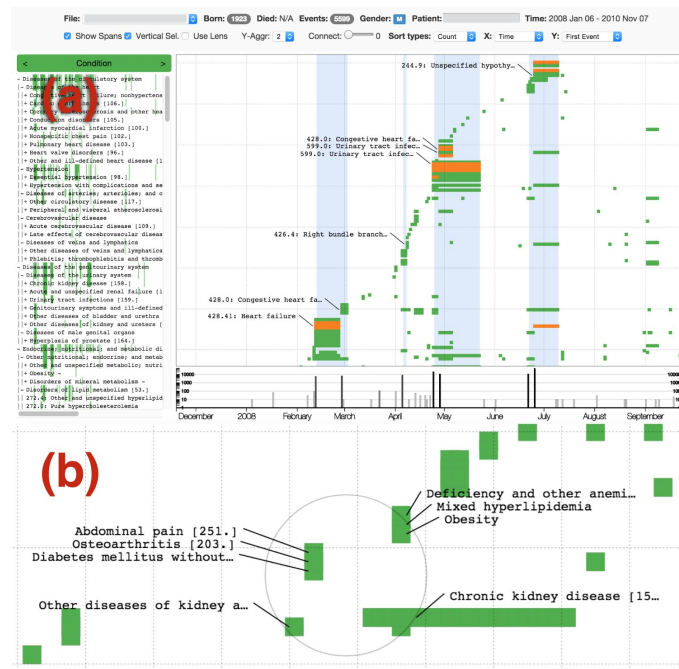

Figure 1a: Overview of the system. At the top general information about the patient is displayed. The main view shows events of the patient's history. The view area can be zoomed into providing a more detailed view. Blue stripes indicate hospital stays. On the left the CCS hierarchy of events happening in the timeline are shown. To give an impression about the frequency and importance of a type its occurrences along the timeline are also encoded here, shown as colored lines for each occurrence. At the bottom the sum of all costs of one day are shown as histograms.

Figure 1b: A user can use a lens to get information about events she is interested in. The horizontal granularity of single events is by day and the vertical position is determined by the first occurrence of the particular event in the history of the patient.

References

1. Plaisant, C., Milash, B., Rose, A., Widoff, S., Shneiderman, B. LifeLines: visualizing personal histories. ACM CHI 1996 Conference Proc.;4:221-227
2. CMS medicare claims synthetic public use files 2008–2010. http://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/SynPUFs/