Healthcare Analytics for Clinical and Non-Clinical Settings

Adam Perer
IBM T.J Watson Research Center
Hawthorne, NY USA
adam.perer@us.ibm.com

CHI'12, May 5–10, 2012, Austin, Texas, USA.
ACM 978-1-4503-1016-1/12/05.

Abstract
The growing use of Electronic Health Records and Personal Heath Records yields a massive amount of long-term data about patients, diseases, and providers. In order to make sense of this data, many novel healthcare analytics and visualization technologies are being developed to help practitioners understand and analyze these longitudinal records. However, to date, many analytic technologies solely focus on clinical data that comes from doctors, hospitals, and healthcare providers. However, analytics that include non-clinical data can potentially be even more effective for enhancing treatments for patients. This workshop is an important event for fostering a research agenda to accomplish this motivation.

Author Keywords
Healthcare analytics

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Our Background
Adam Perer is a research scientist in the Healthcare Analytics group at IBM Research. Our group is made up of physicians, statisticians, health economists, and computer scientists to design, develop, and deploy

novel healthcare analytics and visualization technologies.

Currently, our analytics research focuses on 5 categories:

Descriptive Analytics, to understand the patient population and performance of treatments. One of our projects in this category is Utilization Pattern Analysis, which helps practitioners understand how large groups of patients are treated. Such analytics provide insight to help guide the design of disease management programs [1].

Predictive Analytics, to identify high-risk, high-impact patients. We have developed predictive models and analytics from analyzing patient's longitudinal medical records to provide early detection of congestive heart failure [2].

Prescriptive Analytics, to identify possible actions and care delivery pathways. Analyzing disease progression pathways in terms of these observed events can provide important insights into how diseases evolve over time. We have developed visualizations that connect these pathways to the eventual outcomes of patients so practitioners can understand how certain progression paths may lead to better or worse outcomes [3].

Patient Similarity Analytics, to identify similar patients. Our analytics identify patients who are similar to an index patient for decision support and Comparative Effectiveness Research (CER) analysis [4].

Visual analytics, to support pattern discovery with interactive information visualizations. One of our visual analytics projects is SolarMap, which visualizes topical clusters to help understand how certain diseases are related to each other [5].

Our Motivation

To date, most of the analytics that we develop focus on data that comes from traditional clinical settings. However, we believe that analytics that take into account data from non-clinical settings can be extremely valuable for understanding and monitoring successful treatments of patients. We believe attending this workshop can help us broaden our analytics across the continuum of care. We hope the workshop will foster collaboration with experts who have experience in non-clinical settings, as well as brainstorm how our clinical technologies can be migrated to non-clinical data.

Example: Congestive Heart Failure

One example of our recent work is to provide visual interfaces to help the analysis of disease progression to potentially improve disease diagnoses and treatments.

Congestive heart failure (CHF) occurs when the heart cannot supply the necessary blood flow to meet the needs of the body. This condition is potentially fatal and affects about 2% of adults in developed countries. Despite its widespread occurrence, the disease is difficult to manage and there are currently no systematic diagnostic criteria. However, one commonly used approach comes from the Framingham study, which yielded 18 distinct so-called *Framingham symptoms* for CHF. While these symptoms are used regularly to diagnose CHF, our medical collaborators are interested in understanding how the various symptoms and their order of onset correlate with patient outcome.

To examine this problem, my team built a visual analytics tool, OutFlow, which mines data from EMR and visualizes the progression of Framingham symptoms in patients. It uses a simple visual encoding, as shown in Figure 1. The analytics are seeded with a patient, such as one who is currently being treated. The system gathers data about that patient as well as similar patients found using our Patient Similarity Analytics. In this example, all of these patients had a set of symptoms (labeled as State C in Figure 1). However, there are two types of patients that reached this state. The majority of patients arrived at this state by having a set of symptoms A, while also some arrived having a different set of symptoms B.

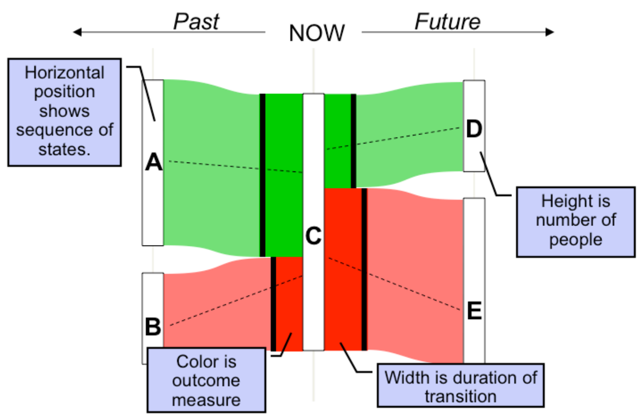

Figure 1. The visual encoding used by Outflow to encode events visually.

As color determines the outcome of the patients in our interface, patients that were in state A (green) have a much better chance of surviving than patients that began with state B (red). Using this tool, the doctors can also look into the future with seeing that patients in state D had a much better chance of survival than those in state E. This can help the doctor better prescribe care pathways that will help the patient land in a healthier state. However, the progression of diseases often produces a much more complicated structure. A real example of a set of similar patients is shown in Figure 2.

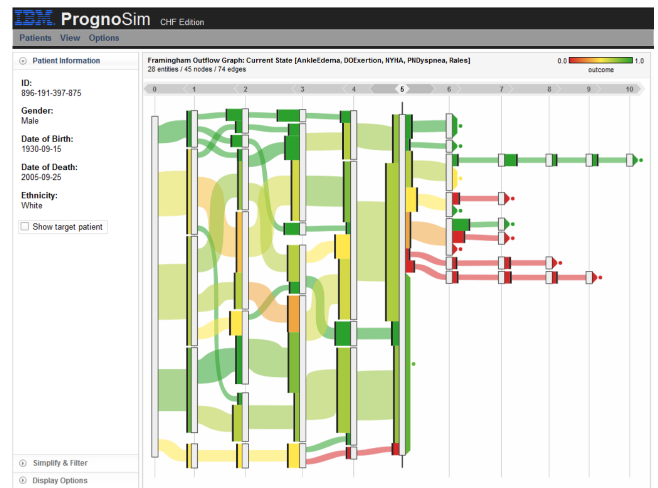

Figure 2. Example of disease progression of patients with CHF in Outflow.

Currently, the system only takes into consideration data that is extracted from clinical records. However, we suspect that our analytics would become even more powerful if we took into consideration non-clinical data supplied by patients. This would provide a more complete picture of the patients' health and potentially allow better similarly and diagnosis analytics by taking into consideration the activities that the patient does outside of the hospital. We believe participation in this workshop would allow us to help shape an agenda of how to leverage of the benefits of analytics and visualizations that use both clinical and non-clinical data from patients.

References

[1] N Lee, A F Laine, J Hu, F Wang, J Sun, S Ebadollahi. Mining electronic medical records to explore the linkage between healthcare resource utilization and disease severity in diabetic patients. IEEE Healthcare Informatics, Imaging and Systems Biology (HISB). 2011.

[2] S. R. Steinhubl, B. A. Williams, J. Sun, R. J. Byrd, Z. Daar, S. Ebadollahi, W. F. Stewart. Text and Data Mining of Longitudinal Electronic Health Records (EHRs) in a Primary Care Population Can Identify Heart Failure (HF) Patients Months to Years Prior to Formal Diagnosis Using the Framingham Criteria. American Heart Association Scientific Sessions. 2011.

[3] Krist Wongsuphasawat and David Gotz**.** Outflow: Visualizing Patient Flow by Symptoms and Outcome. IEEE VisWeek Workshop on Visual Analytics in Healthcare. 2011.

[4] Fei Wang, Jimeng Sun, Shahram Ebadollahi. Integrating Distance Metrics Learned from Multiple Experts and its Application in Inter-Patient Similarity Assessment. SIAM Conference on Data Mining. 2011.

[5] Nan Cao, David Gotz, Jimeng Sun, Yu-Ru Lin, and Huamin Qu. SolarMap: Multifaceted Visual Analytics for Topic Exploration. IEEE International Conference on Data Mining (ICDM). 2011