Demo: Peeking into Patterns of Clinical Event Sequences with Peekquence

Bum Chul Kwon[1], Janu Verma[1], Adam Perer[1]
[1] IBM T.J. Watson Research Center, Yorktown Heights, NY, USA

1 Introduction

Finding temporal patterns in longitudinal event sequences is a challenging task, as the volume and variety of events often make it difficult to extract salient patterns. In response to this challenge, data scientists have turned to machine learning, known as frequent sequence mining (FSM) techniques, to automatically detect the most common sequences of events to unearth interesting patterns. For instance, Frequence [4], Care Pathway Explorer [5], and TimeStitch [6] all use frequent sequence mining techniques to find frequent sequences of events.

However, these algorithms often require users to specify a support threshold that, if too high, will yield only a few patterns, or if too low, will yield numerous patterns that may be difficult for data scientists to determine the interesting sequences from the mundane. In this demo, we aim to make the results of frequent sequence mining algorithms more interpretable by giving end-users powerful ways to explore the data.

Our novel visual analytics system, Peekquence [1], integrates several new techniques that include: 1) powerful ways to navigate the patterns by sorting with metrics relevant to users (variability, correlation to outcome, etc), 2) integration of patterns with patient timelines, so users can understand where the patterns occur in the actual data, and 3) overviews the summarize the most common events in the patterns.

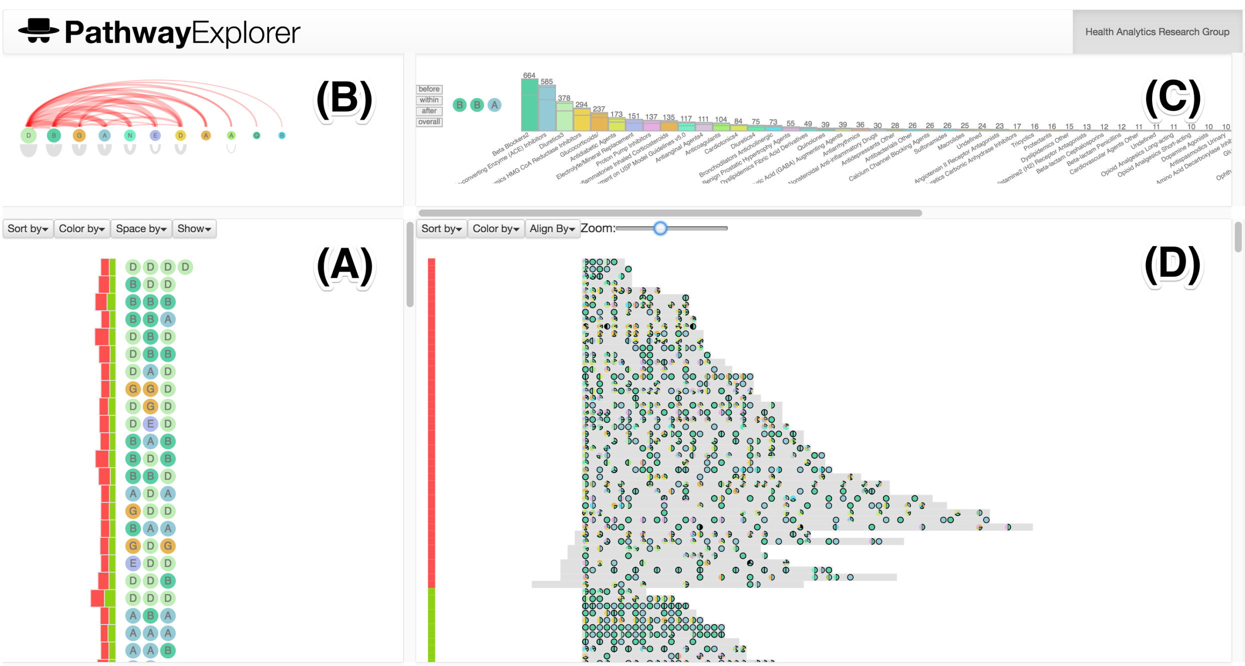

Figure 1: Peekquence consists of four views: (A) the pattern list view showing patterns mined from SPAM with event sequences (colored circles with letters) as well as bars of patients with the ratio of case and control labels (diagnosis of a disease); (B) the sequence network view showing the frequency of event type co-occurrences within mined patterns; (C) the event co-occurrence histogram view showing the frequency of events co-occurring for a selected pattern; (D) the patient timeline view showing patients' event sequences that match the selected pattern.

2 Peekquence

The core visual unit of Peekquence are mined patterns, rather than events. As patterns may contain many different event types and be composed of long event sequences, visualization techniques based on sankey diagrams (a la CareFlow [3]) or aggregated vertical bars (a la EventFlow [2]) tend to suffer from visual complexity without user-controlled filters based on domain expertise. Instead, we opted for a simpler visualization technique: a list of patterns, made up of *event glyphs* that visually encode each event type in the pattern. The event glyphs are visually encoded as circles, colored according to an categorical ontology, and labeled with an abbreviation of the event type's name. All of the four views in Peekquence, shown in Figure 1, use this glyph as the common visual element. In addition to a list of patterns (Figure 1A), there is an overview of common event types in the patterns (Figure 1B), histograms that summarize event types that co-occur with the patterns (Figure 1C), and a coordinated view to the actual patient timelines to understand how the mined patterns manifest in the actual data (Figure 1D).

In Figure 1, Peekquence is demonstrated on mined patterns from a cohort of patients with diagnoses of both congestive heart failure (CHF) and chronic obstructive pulmonary disease (COPD). Of these patients, some are cases that were hospitalized and the remaining are matched controls who have the disease but were not hospitalized. The goal is to use Peekquence to understand if the patients with CHF and COPD that were hospitalized have any distinct patterns of treatments compared to those who were not hospitalizd. For these patients, one year of data is mined after their diagnoses of CHF. In this figure, only treatment events are mined, but the system is capable of merging multiple types of events (e.g. diagnoses, procedures, and labs).

Peekquence has led to interesting discoveries of the benefits and problems by relying on mined patterns as the main unit of visualization. There was no data curation done to the event types loaded into the user interface, but the algorithm was able to surface highly relevant types due to their prominence among patients with CHF and COPD.

3 Conclusion

In this paper, we presented Peekquence, a visual analytics system which aims to increase the interpretability of frequent sequence mining algorithms. The four views combined with interactions provide useful functionalities for users to make sense of patterns as well as their occurrences within patients' records.

References

[1] B. C. Kwon, J. Verma, and A. Perer. Peekquence: Visual analytics for event sequence data. *ACM SIGKDD Workshop on Interactive Data Exploration and Analytics (IDEA 2016)*, 2016.

[2] M. Monroe, R. Lan, H. Lee, C. Plaisant, and B. Shneiderman. Temporal Event Sequence Simplification. *IEEE Transactions on Visualization and Computer Graphics*, 19, 2013.

[3] A. Perer and D. Gotz. Data-driven Exploration of Care Plans for Patients. In *ACM Human Factors in Computing Systems*, pages 439–444, 2013.

[4] A. Perer and F. Wang. Frequence: Interactive mining and visualization of temporal frequent event sequences. In *Proceedings of the 19th International Conference on Intelligent User Interfaces*, pages 153–162, New York, NY, USA, 2014. ACM.

[5] A. Perer, F. Wang, and J. Hu. Mining and exploring care pathways from electronic medical records with visual analytics. *Journal of Biomedical Informatics*, 56(C):369–378, Aug. 2015.

[6] P. J. Polack Jr, S.-T. Chen, M. Kahng, M. Sharmin, and D. H. Chau. Timestitch: Interactive multi-focus cohort discovery and comparison. In *IEEE Proceedings of the Visual Analytics Science and Technology (VAST)*, pages 209–210. IEEE, 2015.