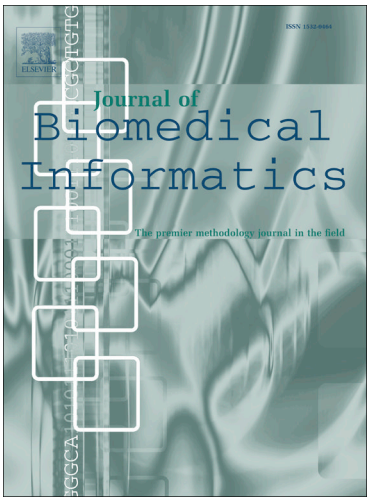

Patient Ranking with Temporally Annotated Data

Luca Bonomi, Xiaoqian Jiang

PII: S1532-0464(17)30277-0
DOI: https://doi.org/10.1016/j.jbi.2017.12.007
Reference: YJBIN 2902

To appear in: *Journal of Biomedical Informatics*

Received Date: 18 January 2017
Revised Date: 23 September 2017
Accepted Date: 13 December 2017

Please cite this article as: Bonomi, L., Jiang, X., Patient Ranking with Temporally Annotated Data, *Journal of Biomedical Informatics* (2017), doi: https://doi.org/10.1016/j.jbi.2017.12.007

Patient Ranking with Temporally Annotated Data

Luca Bonomi[1], Xiaoqian Jiang[2]

Department of Biomedical Informatics
University of California, San Diego

Abstract

Modern medical information systems enable the collection of massive temporal health data. Albeit these data have great potentials for advancing medical research, the data exploration and extraction of useful knowledge present significant challenges. In this work, we develop a new pattern matching technique which aims to facilitates the discovery of clinically useful knowledge from large temporal datasets. Our approach receives in input a set of temporal patterns modeling specific events of interest (e.g., doctor's knowledge, symptoms of diseases) and it returns data instances matching these patterns (e.g., patients exhibiting the specified symptoms). The resulting instances are ranked according to a significance score based on the p-value. Our experimental evaluations on a real-world dataset demonstrate the efficiency and effectiveness of our approach.

1. Introduction

Nowadays, the large amount of temporally annotated data being collected within healthcare institutions holds great promises for many medical applications. For example, the analysis of temporal data associated with patients electronic health records (EHRs) can be used to facilitate early diagnosis and clinical decision-making. While these data may be beneficial for many research studies, the extraction of useful knowledge is often a very challenging task. To this end, a variety of techniques have been proposed over the years. For example, temporal *pattern mining* techniques [1, 2] aim to extract useful information from the data in the form of patterns which model multiple events that occur sequentially within a short period of time. While these approaches have been shown to be effective in many applications, their extraction criterion is solely based on the input data (e.g., frequency, significance) and therefore are unable to take into consideration external domain knowledge (e.g., doctor's knowledge). As a consequence, sequential data that are clinically relevant may still remain hidden. To facilitate the discovery of such information, *pattern matching* approaches (e.g., temporal query languages) have been proposed [3, 4]. These techniques receive in input a set of parameters defining the extraction criteria and return the data instances satisfying them. The works in [5, 6, 7] are examples of pattern matching techniques developed to identify time-series data that match user-defined patterns.

In this work, we propose a novel pattern matching approach that aims to identify temporally annotated sequences that match user-defined patterns of interest and are statistically significant. With our solution, clinicians can use temporal medical data, such as EHRs, to conduct patients similarity search and identify data instances to support their decisions. In general, data driven decisions are central in modern healthcare and have the potential for advancing medical research and saving patient lives. For example, a study by Frankovich et al. [8] discussed how a 13-year-old girl admitted to the hospital with severe inflammation of the kidneys and pancreas was helped by evidence generated by 98 pediatric lupus patients with similar conditions. Furthermore, the retrieval of relevant

[1]lbonomi@ucsd.edu
[2]x1jiang@ucsd.edu