Journal of Biomedical Informatics 56 (2015) 369–378

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Mining and exploring care pathways from electronic medical records with visual analytics

Adam Perer [a,*], Fei Wang [b], Jianying Hu [a]

[a] IBM T.J. Watson Research Center, 1101 Kitchawan Road, P.O. Box 218, Yorktown Heights, NY 10598, USA
[b] University of Connecticut, Storrs, CT, USA

ARTICLE INFO

Article history:
Received 3 December 2014
Revised 13 May 2015
Accepted 26 June 2015
Available online 2 July 2015

Keywords:
Visual analytics
Temporal event visualization
Frequent sequence mining
Data-driven care plans

ABSTRACT

Objective: In order to derive data-driven insights, we develop *Care Pathway Explorer*, a system that mines and visualizes a set of frequent event sequences from patient EMR data. The goal is to utilize historical EMR data to extract common sequences of medical events such as diagnoses and treatments, and investigate how these sequences correlate with patient outcome.
Materials and methods: The Care Pathway Explorer uses a frequent sequence mining algorithm adapted to handle the real-world properties of EMR data, including techniques for handling event concurrency, multiple levels-of-detail, temporal context, and outcome. The mined patterns are then visualized in an interactive user interface consisting of novel overview and flow visualizations.
Results: We use the proposed system to analyze the diagnoses and treatments of a cohort of hyperlipidemic patients with hypertension and diabetes pre-conditions, and demonstrate the clinical relevance of patterns mined from EMR data. The patterns that were identified corresponded to clinical and published knowledge, some of it unknown to the physician at the time of discovery.
Conclusion: Care Pathway Explorer, which combines frequent sequence mining techniques with advanced visualizations supports the integration of data-driven insights into care pathway discovery.

1. Introduction

Advances in data collection and storage technologies have led to ubiquitous yet complex Electronic Medical Records (EMRs). Because patient EMRs reflect the temporal nature of patient care, a patient's sequence of symptoms and diagnoses often correlates with their medications and procedures. These observed events may unlock the ability to analyze disease progression pathways and identify temporal patterns [33]. We believe such patterns may provide important insights into how diseases evolve over time and the effects of implemented interventions.

However, despite the fact that temporal knowledge discovery and pattern mining is no longer an unaddressed problem in data mining [1], it is still not easy to directly apply or adapt existing technologies to medical data for a number of reasons:

- There are many different event types encoded in EMRs. For example, there are thousands of distinct diagnosis codes, lab tests and drugs. Typically, large numbers of distinct event types can adversely affect the computational efficiency of temporal pattern mining techniques.
- EMRs may contain millions of patients over decades, and such voluminous data poses a great computational challenge to conventional methods.
- In medical scenarios, there are typically outcomes associated with each patient, such as the diagnosis of a disease or hospitalization. Clinicians are not only interested in the temporal patterns, but also in the correlations between such patterns and the patients' outcomes. Most existing pattern mining techniques lack the capability to elucidate such correlations.

In this paper, we propose *Care Pathway Explorer*, an interactive hierarchical information exploration system for analyzing longitudinal medical records. Our system provides a visual overview of frequent patterns mined from EMR patient traces. Instead of mining and visualizing all details at once, the interface supports interactive exploration for researchers to examine the level-of-detail relevant to user tasks by leveraging event hierarchies.

* Corresponding author. Tel.: +1 914 945 2681.
E-mail address: adam.perer@us.ibm.com (A. Perer).

http://dx.doi.org/10.1016/j.jbi.2015.06.020
1532-0464/

There are several major components to the *Care Pathway Explorer*:

- The Event Database, which stores patient electronic medical records with multiple event types, patient outcomes, and the event hierarchy for each event type.
- The Data Preprocessor, which constructs patient traces from the event database that can be directly fed into the Frequent Pattern Analytics engine.
- The Frequent Pattern Analytics, which mines frequent patterns from patient traces obtained by the Data Preprocessor and analyzes how these mined patterns correlate with outcomes.
- The Visual Interface, which provides visualizations of the frequent pathway events, with which users can interact to explore details of interest and generate insights.

These visual analytic technologies combine to support the mining and visualization of care pathway patterns. The goal is to provide insights into which practices lead to desirable patient outcomes, so clinicians can interpret meaningful patterns and customize care plans for complex patients. As established clinical practice guidelines typically only cover a single disease condition for average patients, such customization tools are critical in order to tailor care plans to the specific needs of real world patients who often have multiple complex comorbidities.

1.1. Background and significance

The exploration of temporal knowledge from longitudinal EMRs with data mining techniques is an important problem that has been the focus of study of much medical informatics research. In general, there are two types of studies: *holistic* and *localized*.

The goal of a holistic study is to exploit knowledge that can describe the overall event traces of the patient population. A typical technique that falls into this category is Business Process Management (BPM), which is a holistic management approach focused on aligning all aspects of an organization with the desires and needs of clients. In the healthcare domain, BPM technologies are mainly used for analyzing clinical pathways [2–4], which are standardized and normalized treatment patterns. However, applying BPM techniques to real patient data (e.g., for designing personalized clinical pathways) results in very complex and chaotic graphs.

To avoid the clutter caused by holistic studies, localized studies focus more on exploring the local characteristics of the patient event traces. For example, Norén et al. [5] propose a graphical statistical approach for summarizing and visualizing temporal associations between the prescription of a drug and the occurrence of a medical event, where the focus is the time period around the drug description. Chittaro and Combi [6] and Fails et al. [7] propose visual interfaces for constructing database queries to seek temporal patterns in multivariate temporal clinical data; the latter was further used in [8] for searching temporal patterns in patient histories. However, the system requires user specification of the structure of the pattern to constrain the database queries. Mörchen and Ultsch [9] propose a method called *Time Series Knowledge Mining* (TSKM) for uncovering local temporal relationships in multivariate data, but requires predefined temporal grammar and logic with prior knowledge.

Another methodology that is relevant to sequential pattern mining is *Temporal Abstraction* [10,11,35]. However, this technology generally requires an interval-based representation, which needs to know the duration of each event. In real-world EMR systems, duration information is often not captured, so we choose to use techniques that do not require this information. However, it is also possible to abstract point-based data by applying temporal knowledge which results in a more abstract representation of the data, in the form of symbolic time intervals Batal et al. provide several pattern mining techniques that uses a time interval-related representation of a sequence, which requires either the events have continuous values that can be quantized or the duration of every event is available [36,42]. Moskovitch et al. provide several approaches for discretizing continuous event values to derive more discriminative time-interval related patterns [40,41]. Patel et al. also provide a technique for mining interval-based events [43]. KNAVE II [12], VISITORS [13] and ViTA-Lab [34] are visual interfaces to interactively explore the temporal abstraction process in single and multiple patients, respectively. Other interval-based approaches include MuTIny [14] that discovers multi-time interval patterns, and MEMURY [15]. As most EMR data contains point events, and not interval events, our method aims to mine patterns from sequences of point events. Our work is different from these approaches in the sense that we only focus on point-based event sequences, although we propose a scheme for multiple levels-of-detail that could be applied to any type of pattern mining algorithm. We further note that pattern explosion can happen for either type of algorithm, whether it is for point-based event sequences or time-interval event sequences. Our scheme for multiple levels-of-detail, as well as our visual user interface, can be applied to these other type of pattern mining algorithms.

The *Care Pathway Explorer* system presented in this paper falls into the category of localized studies of EMRs. *Care Pathway Explorer* mines frequent patterns from patient traces and then illustrates them in a visually comprehensible and interactive user interface. There have recently been significant advances in the visualization community toward designing techniques for temporal event sequences of electronic health records. CareCruiser is a visualization system to compare EMR data to medical protocols [16]. LifeFlow [17] introduced a way to aggregate multiple event sequences into a tree, and EventFlow [18] later extended this approach to support both point-based and interval-based events. Outflow [19] designed a way to aggregate events into a graph, as well as integrating statistics. CoCo [37] is a tool for comparing event sequences at a cohort level.

Most recently, Frequence [20] is a user interface that integrates data mining and visualization in an interactive information exploration system for finding frequent patterns from longitudinal event sequences. The work described in this paper is an extension and adaptation of Frequence to support the use cases of medical informatics more directly, including a customized Event Database and Data Pre-processor designed for patient EMRs. Furthermore, an additional visualization was created to support an overall view of all events found in the patterns supporting a use case requested by physicians. In addition, Care Pathway Explorer has been integrated with a care plan template authoring tool, to support an end-to-end workflow from data-driven insights to institutional implementation.

2. Materials and methods

In this section, we introduce the *Care Pathway Explorer* system in detail, which supports the following flow of exploration:

1. The system shows an overview of the frequent patterns mined from patient event traces at the coarsest level, featuring statistics that indicate their correlations with outcomes.
2. The physician examines the frequent patterns and interactively selects specific patterns of interest for more detail.