Journal of Biomedical Informatics 86 (2018) 33–48

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

Probabilistic modeling personalized treatment pathways using electronic health records

Zhengxing Huang[a,*], Zhenxiao Ge[a], Wei Dong[b], Kunlun He[b], Huilong Duan[a]

[a] College of Biomedical Engineering and Instrument Science, Zhejiang University, China
[b] Department of Cardiology, Chinese PLA General Hospital, China

ARTICLE INFO

Keywords:
Personalized Treatment Pathway
Process Mining
Hidden Markov Model
Electronic Health Record

ABSTRACT

Background: Modeling personalized treatment pathways plays an important role in understanding essential/critical treatment behaviors performed on patients during their hospitalizations and thus provides the opportunity for the improvement of better health service delivery in treatment pathways.
Objective: Unlike traditional business process mining, modeling personalized treatment pathways is more challenging because they are typically case-specific. Although several studies have been devoted to modeling patient treatment pathways, limited efforts have been made on the extraction of latent semantics and their transitions behind patient treatment pathways, which are often ambiguous and poorly understood.
Methods: In this article, we propose an extension of the Hidden Markov Model to mine and model personalized treatment pathways by extracting latent treatment topics and identifying their sequential dependencies in pathways, in the form of probabilistic distributions and transitions of patients' raw Electronic Health Record (EHR) data.
Results: We evaluated the proposed model on 48,024 patients with cardiovascular diseases. A total of 15 treatment topics and their typical transition routes were discovered from EHR data that contained 1,391,251 treatment events with 2786 types of interventions and that were evaluated by ten clinicians manually. The obtained *p-values* are 0.000146 and 0.009106 in comparison with both Latent Dirichlet Allocation and Sequent Naïve Bayes models, respectively; this outcome indicate that our approach achieves a better understanding of human evaluators on modeling personalized treatment pathway than that of benchmark models.
Conclusion: The experimental results on a real-world data set clearly suggest that the proposed model has efficiency in mining and modeling personalized treatment pathways. We argue that the discovered treatment topics and their transition routes, as actionable knowledge that represents the practice of treating individual patients in their clinical pathways, can be further exploited to help physicians better understand their specialty and learn from previous experiences for treatment analysis and improvement.

1. Background and significance

Healthcare organizations face a significant challenge in how best to manage the efficiencies and effects of treating patients with costly health service requirements, and this challenge has increased the need to understand how health services are delivered for obtaining cost-effective improvements in clinical practice [1,2]. In this context, clinical guidelines or treatment pathways (TP), that aims to improve the continuity and co-ordination of treatment behaviors across different disciplines and sectors, are developed to manage the quality of health services and ensure cost-effectiveness [2–6]. However, these guidelines and pathways often bear no relation to the ideal as envisaged by the designers of TP in heterogeneous and extremely complex contexts, and they inevitably result in continued variations in actual patient TP [5,7].

To address this problem, recent studies have begun to look at medical cost controlling and health outcomes of TP to inform investment and practice alteration decisions, by utilizing a large volume of Electronic Health Record (EHR) data [8–10], which are regularly provided by various hospital information systems. However, most of these techniques and tools look at aggregated EHR data seen from an external perspective, e.g., length of stay (LOS), medical cost, bed utilization, and health service levels, and can only answer relatively simple questions such as what the average LOS or medical cost is for a group of patients [11]. While healthcare organizations typically have an oversimplified and incorrect view of the actual situations in TP, it is important to gain

* Corresponding author.
E-mail address: zhengxinghuang@zju.edu.cn (Z. Huang).

https://doi.org/10.1016/j.jbi.2018.08.004
Received 9 February 2018; Received in revised form 26 July 2018; Accepted 6 August 2018
Available online 21 August 2018
1532-0464/ © 2018 Elsevier Inc. All rights reserved.

insights into personalized TP to extract crucial and potential information, such as essential/critical treatment behaviors and their causal dependencies in TP, to assist clinicians and managers to substantially improve their health service quality in pathways at a very refined level [3,7,8].

To accomplish goal, the methods that use data mining and machine learning techniques to analyze TP based on associated EHR data have been receiving gradual attention in medical informatics [12–21]. These techniques, also called process mining, have been widely applied in business process management and have achieved emerging attention in clinical settings [8]. In general, process mining techniques are used to mine actual behaviors in business processes and discover hidden business process patterns/models from event logs [12]. Shifting to clinical settings, one fundamental aspect of process mining is to discover critical/essential treatment topics and their transitions in individual patient pathways from EHR data, because it is not biased by perceptions or normative behaviors [5,19,22]. Extracted treatment topics and their causal dependencies, as the backbone of a personalized patient pathway, can help clinical analysts locate treatment information of interest, and provide an overview of TP in an easy and quick manner [21].

However, the diversity of medical behaviors in TP is far higher than that of common business processes [7,12]. TP is typically case-specific and loosely-structured, i.e., an individual TP is described as a series of clinical epochs/stages in the patient's hospitalization and each clinical epoch is typically composed of various and heterogeneous treatment events that can occur arbitrarily without a restricted order [23,24]; additionally, the composition of treatment events has a large variability depending on case-specific decisions that are made by interpreting patient-specific statuses [22]. In addition, latent semantic meanings (i.e., treatment topics) behind clinical epochs are always related to other adjacent epochs, and they are often ambiguous and poorly understood. Thus, it is a crucial challenge to determine how to extract latent topics and identify their dependency characteristics from a loosely-structured patient's TP. In the literature, although there are some previous studies on mining TP using EHR data, most of them attempt to modify the existing algorithms to handle structured business process data [7,16,17], which often generates spaghetti-like TP patterns that are incomprehensive to clinical experts, and thus hinder the continuous improvement of TPs [3].

To accomplish the crucial job of filling this void, we present a novel graphical model for mining personalized TPs from a large volume of EHR data. Specifically, we present an extension of a well-known Hidden Markov Model (HMM) to incorporate the prior knowledge of treatment behaviors into mining latent treatment topics and their specific transitions for each patient's TP from the associated EHR. The proposed model incorporates the medical behavior dependency that is observed in the patient pathway and transforms the observed treatment behaviors into a probabilistic space, to configure latent treatment topics and their transition routes in the pathway. The proposed model facilitates the analysis and improvement of TPs, and can answer questions such as what latent treatment topics exist in patient pathways, and how treatments shift in the execution of patient TPs. We conduct a case study on a clinical dataset that was collected from the cardiology department of Chinese PLA General Hospital, to validate the effectiveness and efficiency of the proposed model for mining personalized TP of patients with cardiovascular disease (CVD). The experimental results indicate that our model can assist clinicians in discovering latent treatment topics and the transitions that are worthwhile to note, and it can identify target patients for specific and vital care in their TP.

The remainder of this article is organized as follows. In Section 2, we introduce related work. Section 3 presents our approach of modeling personalized TP using EHR. Section 4 shows the experimental results on a clinical data set with 48,024 CVD patients, and finally, Section 5 concludes the work.

2. Related work

Process mining is a research discipline that focuses on providing evidence-based analysis for effective business process management [8,25,26]. Shifting to clinical settings, applications that employ process mining techniques to routinely collected clinical data can enable healthcare stakeholders to empirically investigate treatment behaviors as they are delivered by different health providers [7]. In [8], Rojas et al systematically reviewed the literature on process mining in healthcare and they concluded that the adoption of three main types of process mining techniques (i.e., process discovery [12–14,19–21,24], conformance checking [18,27,28], and enhancement [29–31]) in the field of TP not only ensures treatment behaviors be firmly understood but also can generate benefits that are associated with the efficiency of health service delivery [8].

For example, Mans et al., proposed using process mining to discover treatment process models for both stroke and gynecological oncology [16,17]. Perer et al., developed a system to extract and visualize a set of frequent treatment event sequences from patient EHR data, and they investigated how these sequences correlate with patient outcomes [22]. Bouarfa and Dankelman proposed a conformance checking algorithm to detect outliers in TPs [18]. Rovani et al. reported a case study that shows how process mining can be applied to mediate between event data that reflect the clinical reality and TP models that describe best-practices in medicine [24]. Lakshmanan et al. presented a hybrid approach for extracting correlations between TP and patient outcomes, which is based on which TP can be further enhanced [29]. Antonelli et al., proposed a new process mining framework based on generalized association rules to discover multiple-level correlations among medical treatments [35]. In [48], Zhang and Padman proposed a data-driven approach, illustrated in the context of chronic kidney disease, to develop clinical pathways of care delivery from EHR D data. By summarizing longitudinal information from EHRs into clusters of common sequences of patient visits, their approach can assist in the efficient review of current practices and identifying potential innovations in the care delivery process [48]. In our previous work [13], we constructed concise and comprehensive summaries of TP by segmenting the observed time period of patient treatment journeys into continuous and overlapping time intervals, and we discovered frequent treatment behaviors in each specific time interval from the data. To model complex and diverse TP, we developed a probabilistic topic model to link patient features and treatment events together to extract latent treatment topics and patterns hidden in EHR data [14,21]. Work that is closely related to ours is presented in [44], in which Lin et al. present an HMM-based approach for discovering clinical pathway patterns. Their work has clearly proven its robustness and scalability in modeling clinical pathways using HMMs. In our work, we propose to leverage HMMs for modeling sequential EHR data. Different from the work presented in [44], we propose to use Bayesian methods to estimate HMM parameters from the posterior distribution. Instead of searching for an optimal set of parameter values using the traditional Maximum-Likelihood-Estimation-based HMM learning solution, Bayesian HMM can utilize the observed data to directly maximize the probability of the hidden variables by integrating the overall possible parameter values [45], which can provide a more scalable and stable process of parameter estimation for an HMM.

From a more general perspective of business process management, the most common approach to business process analysis is to group similar process instances based on their edit-distance of control-flow graphs [32,33]. For example, in [32], a rather simple graph edit-distance measure has been proposed and adopted for similarity assessment in business process change reuse. In [33], Ma et al., proposed a graph distance-based approach for measuring the similarity between data-oriented workflows with variable time constraints. However, Qiao et al., showed that traditional graph edit-distance-based methods might not be appropriate for mining real-world business processes, which tend

Download English Version:

https://daneshyari.com/en/article/8953886

Download Persian Version:

https://daneshyari.com/article/8953886

Daneshyari.com