J Med Syst (2016) 40: 8
DOI 10.1007/s10916-015-0380-6

SYSTEMS-LEVEL QUALITY IMPROVEMENT

Outcome Prediction in Clinical Treatment Processes

Zhengxing Huang[1] · **Wei Dong**[2] · **Lei Ji**[3] · **Huilong Duan**[1]

Received: 11 May 2015 / Accepted: 9 October 2015 / Published online: 29 October 2015

Abstract Clinical outcome prediction, as strong implications for health service delivery of clinical treatment processes (CTPs), is important for both patients and healthcare providers. Prior studies typically use a priori knowledge, such as demographics or patient physical factors, to estimate clinical outcomes at early stages of CTPs (e.g., admission). They lack the ability to deal with temporal evolution of CTPs. In addition, most of the existing studies employ data mining or machine learning methods to generate a prediction model for a specific type of clinical outcome, however, a mathematical model that predicts multiple clinical outcomes simultaneously, has not yet been established. In this study, a hybrid approach is proposed to provide a continuous predictive monitoring service on multiple clinical outcomes. More specifically, a probabilistic topic model is applied to discover underlying treatment patterns of CTPs from electronic medical records. Then, the learned treatment patterns, as low-dimensional features of CTPs, are exploited for clinical outcome prediction across various stages of CTPs based on multi-label classification. The proposal is evaluated to predict three typical classes of clinical outcomes, i.e., length of stay, readmission time, and the type of discharge, using 3492 pieces of patients' medical records of the unstable angina CTP, extracted from a Chinese hospital. The stable model was characterized by 84.9% accuracy and 6.4% hamming-loss with 3 latent treatment patterns discovered from data, which outperforms the benchmark multi-label classification algorithms for clinical outcome prediction. Our study indicates the proposed approach can potentially improve the quality of clinical outcome prediction, and assist physicians to understand the patient conditions, treatment inventions, and clinical outcomes in an integrated view.

Keywords Clinical treatment process · Electronic medical records · Treatment pattern discovery · Clinical outcome prediction

This article is part of the Topical Collection on *Systems-Level Quality Improvement*

✉ Zhengxing Huang
zhengxinghuang@zju.edu.cn

[1] College of Biomedical Engineering and Instrument Science, Zhejiang University, Zhejiang, China

[2] Cardiology department of Chinese, PLA General Hospital, Beijing, China

[3] IT department of Chinese, PLA General Hospital, Beijing, China

Introduction

Clinical outcomes, e.g., length of stay (LOS), readmission time, discharge type, etc., have been recognized as the critical and essential indicators of medical service delivery in clinical treatment processes (CTPs) [1, 18, 28, 29]. There is a myriad of factors that might influence clinical outcomes of CTPs [2, 3]. For instance, patients may have unexpected complications, and medical resources could not be available at a particular time instant in CTPs [4, 5]. In such a case, specific medical measures must be taken, and may have an important influence on clinical outcomes [1]. Thus, understanding the factors that determine clinical outcomes, or a capacity of clinical outcome prediction, has strong implications for health service delivery, and could promote the development of efficient CTPs [1, 13, 28].

Many variables and factors have been suggested in the literature that might affect clinical outcomes of CTPs [10, 11], and many attempts have been conducted to study these variables to develop prediction models [1, 6–9, 13, 17]. Most

of them adopt data mining or machine learning algorithms to generate a prognosis's model from clinical data [1]. As valuable as these works, there are several limitations:

1. Prior works provide clinical outcome predictions, mainly at the earlier stage of CTPs [1, 6], e.g., admission. Note that CTPs are highly dynamic, context sensitive, knowledge intensive and variants happen frequently in CTPs, which implicitly changes clinical outcomes [2–4, 12]. Existing prognostic models that are mainly built on static patient-specific information such as patient demographics, symptoms, vital signs, etc., might not provide accurate outcome predictions for CTPs.
2. Prior works leverage patient-specific information to predict clinical outcomes considering data mining or machine learning approaches [6–8, 17], whereas alternative prediction methods that measure treatments in CTPs have received little attention [1]. In fact, a large volume of information describing treatments is available [2–4, 14], and thus can be exploited to improve outcome predictions across various stages in CTPs, e.g., admission, pre-surgery, surgery, recovery after the surgery, discharge, etc.
3. While most prior studies build prediction models for a particular type of clinical outcomes [1, 6–9], a mathematical model that predicts multiple outcomes simultaneously has not yet been established to the best of our knowledge.

In this study, we present a new and hybrid approach to overcome the drawbacks of the previous studies, using a large volume of electronic medical records (EMRs), which provides the opportunities to learn from previous cases [29]. In general, a piece of EMR contains heterogeneous medical information, including both patient-specific features with their values, and various treatment interventions with their occurring time stamps, across various stages of CTPs. The different aspects of medical information recorded in EMRs are highly correlated and thus provide huge potential to be exploited to mine the associations between treatments applied to patients and the resulted clinical outcomes, as the main objective of this study.

The proposed approach integrates a treatment analysis model and a prediction model as an effective means to continuously predict multiple clinical outcomes taking into account the latest updates of medical information. In the first phase, we employ a probabilistic model to analyze treatment behaviors. This assists to identify critical patient features and essential treatment interventions that lead to specific treatment outcomes. We argue that the efforts on treatment analysis can generate low-dimensional features of CTPs, i.e., essential and critical treatment patterns underlying CTPs [2, 3, 15], which can be exploited to improve the quality of clinical outcome prediction in the latter phase. Based on the discovered treatment patterns, we employ multi-label classification [16] to develop a prediction model to classify multiple clinical outcomes of CTPs. As a particular learning task, multi-label classification can classify a single patient instance of CTP to several outcome classes at the same time, and thus these outcomes are not mutually exclusive for CTPs. To demonstrate the feasibility of the proposed approach, a case study on a realistic collection of EMRs is conducted, in comparison with typical multi-label classification algorithms without efforts on treatment analysis.

The rest of this paper is organized as follows: Section 2 introduces preliminary knowledge of our approach. Section 3 presents the proposed approach that integrates a probabilistic topic model for latent treatment pattern discovery, and multi-label classification algorithms for multiple clinical outcome predictions. Section 4 presents the experimental results of applying the proposed approach to a collection of EMRs about the unstable angina CTP. Comparisons are made between the proposed approach, and the representative multi-label classification algorithms without efforts on treatment pattern discovery. Finally, conclusions and discussion are given in Section 5.

Preliminaries

In this study, we assume that the users of our method are clinical analysts who have access to a collection of EMRs and who would like to analyze CTPs and investigate typical clinical outcomes of CTPs using EMRs. Before we introduce our method, we formally define an abstract representation of EMRs, treatment patterns, and some basic notions of the proposed approach.

Definition 1 (Electronic Medical Records, EMRs) Let D be a collection of EMRs. Each EMR d in D, recording medical information of a particular patient's CTP, consists of the description on the patient-specific information, treatment interventions performed on the patient given his/her conditions and at specific time stamps during the CTP execution, and clinical outcomes resulted from the execution of the CTP.

Note that an EMR d can be considered as a mixture of static patient information that traditionally absent from claims (e.g., vital signs, comorbidities, laboratory test results, etc.), and time-stamped treatment events (e.g., medication orders, examinations, surgery, etc.) that are often packed into episodes of admissions and treatments and thus are highly irregular. Static information includes demographic variables and thus is generally moderate in dimensions. The events are, on the other side,

complex and high dimensional. These information includes all a priori plausible predictors of clinical outcomes available within the EMRs.

Definition 2 (Treatment pattern) Let $\mathcal{Z}$ be the universe of treatment patterns. Let D be a collection of EMRs. An EMR d in D is a probability distribution over a set of treatment patterns $\mathcal{Z}_d \subseteq \mathcal{Z}$. A pattern $z \in \mathcal{Z}_d$ is a probability distribution over a set of clinical words ($\mathcal{W} = \{f, v\} \cup \{a, t\}$, which consist of both a set of pairs of patient features F and their values V, and a set of treatment activities A performed on the patients and their occurring time stamps T in CTPs. Specifically, the probability distribution of a treatment pattern Z is a collection of the positive real number over W with a sum equal to 1.

Definition 3 (Patient feature) Patient features are observable clinical variables, including patient demographics, lab test results, vital signs, etc., which can have categorical or numerical values. Intuitively, patient conditions are described as a set of pairs of patient features and their values f_d, v_d, which are the most important issue physicians aim to figure out for treatments during CPs. Here $f_d = \{f_{d,i}\}_{i=1}^{N_d^f}$ represents patient features that are measured on a particular patient and recorded in an EMR d, $v_d = \{v_{d,i}\}_{i=1}^{N_d^f}$ represents their values, and N_d^f is the number of patient features recorded in d.

Definition 4 (Treatment intervention) Treatment interventions are represented as a set of treatment activities performed on a particular patient at specific time stamps during CPs, given specific patient's conditions. These treatment activities and their occurring time stamps a_d, t_d are recorded in a patient's EMR d. Here $a_d = \{a_{d,j}\}_{j=1}^{N_d^a}$ represents activity types, $t_d = \{t_{d,j}\}_{j=1}^{N_d^a}$ represents the occurring time stamps of a_d, and N_d^a is the number of treatment activities in d.

Definition 5 (Clinical outcome) A clinical outcome variable is defined as a measure of health state of a patient resulted from the execution of CTPs. Formally, we assume that there is a group of clinical outcomes, C_d, in an EMR d, including a vast range of descriptions on patient health states (e.g., mortality, transfer, or normal discharge, etc.), physiologic measures (e.g., heart attack, etc.), and patient-reported health states (e.g., Length of stay in 7 days, readmission in one month, etc.).

Method

To predict clinical outcomes for CTPs, we propose a hybrid method integrating a probabilistic topic model for latent treatment pattern discovery and multi-label classification for clinical outcome prediction. We, firstly, analyze treatments to extract essential patterns from EMRs. The probabilistic topic model reduces the high dimensional features (i.e., patient-specific information and treatment interventions) of CTPs to low dimensional features (i.e., latent treatment patterns) [2]. Then, representative multi-label classification algorithms are applied to generate a clinical outcome prediction model for CTPs. Figure 1 illustrates the procedure of our method. The steps are detailed in the following.

Initially, EMRs records both patient conditions and treatments of CTPs. Thus, EMRs can be handled by the two principal parts of our approach: a probabilistic topic model for latent treatment pattern discovery, and a multi-label classifier for multiple clinical outcome predictions. The probabilistic topic model is used to convert the data of EMRs into another representation, i.e., the original EMRs recording typical treatment behaviors of CTPs can be represented by low dimensional vectors of essential treatment patterns underlying the execution of CTPs. The multi-label classifier is then used to classify discovered treatment patterns w.r.t multiple clinical outcomes. Hence, the probabilistic topic model generates low dimension features that the multi-label classifier recognizes. In the next sections, we will describe our approach in detail.

Treatment pattern discovery

To discover latent treatment patterns from EMRs, we apply a generative model, i.e., treatment pattern model (TPM), which was proposed in our previous work [31]. TPM is an extension of Latent Dirichlet Allocation [19]. Using TPM, each patient's EMR is modeled as a multinomial distribution of treatment patterns, and each treatment pattern is modeled as a multinomial distribution of clinical events. The generative process begins by choosing a distribution over treatment patterns $\mathbf{z}=(z_{1:k})$ for a given patient record. Given a distribution of treatment topics for a specific patient record, clinical events are generated by sampling treatment patterns from this distribution. The result is a vector of N clinical words $\mathbf{w}=(w_{1:N})$ for a patient record.

Figure 2 shows the graphical model representation of TPM for mining latent treatment patterns. Z, F, and A denote variables "treatment patterns", "patient features", and "treatment activities" respectively. θ, ϕ, ξ, φ, δ, μ, and λ are distributions of treatment patterns over EMRs, treatment activities over patterns, the occurring time stamps of treatment activities over patterns, patient features over treatment patterns, the values of categorical features over treatment patterns, the mean values of numerical features over treatment patterns, and the

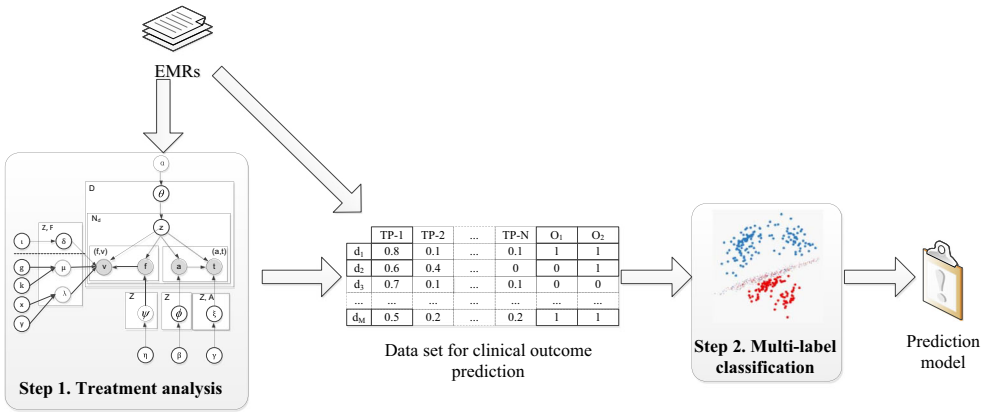

Fig. 1 The procedure of the proposed approach

precisions of numerical features over treatment patterns, respectively. w represents clinical words, which are either a pair of patient feature f and its value v or a clinical event e (a pair of treatment activity a and its occurring time stamp t). N_d represents the number of occurrences of clinical words for one EMR d. D represents the number of EMRs in the collected data set. The hyperparameters α, β, γ, η, ι, g, k, x, y control dispersions of treatment patterns, clinical activities and their occurring time stamps, patient features, values for categorical features, values' means for numerical features, and values' precisions for numerical features, respectively.

The generative process for TPM is the same as that of standard topic models. Each EMR d in D has treatment pattern proportions θ_d that are sampled from a Dirichlet distribution with prior α. θ_d stands for the probability of assigning a treatment pattern z to a clinical word generated from d. Note that clinical words are either pairs of patient features with their values, or clinical events. For clinical events, a treatment pattern z is associated with a multinomial distribution ϕ_z over the activity type a of an event e, and a multinomial distribution $\xi_{z,a}$ over the occurring time stamp of e for pattern z. For patient features, a treatment pattern z is chosen from the pattern proportions, and then a patient feature f is sampled from a treatment pattern-specific multinomial distribution ψ_z.

Note that there are two kinds of patient features, i.e., categorical features and numerical features. A categorical feature $f_{d,i}$ with its value $v_{d,i}$, is generated from the distribution $\delta_{z_{d,i}, f_{d,i}}$. There are, in total, Z×F prior distributions of patient feature-value pairs, which follow a Dirichlet distribution with prior ι. For numerical features, each treatment pattern has its own value distribution, which is assumed to be a Normal distribution, Normal $\left(\mu_{z_{d,i}, f_{d,i}}, \lambda^{-1}_{z_{d,i}, f_{d,i}}\right)$.

Fitting the proposed TPM is equivalent to finding parameters α for the Dirichlet distribution, parameters β for the treatment pattern-activity type distributions, and parameters γ for the treatment pattern-activity type-time stamp distributions, parameters η for the treatment pattern-patient feature distributions, parameters ι for the treatment pattern-categorical patient feature-value distributions, and parameters g, k, x and y for the treatment pattern-numerical patient feature-value distributions, that maximize the likelihood of the data for a collection of EMRs:

$$\text{Likelihood}\ (\alpha, \beta, \gamma, \eta, \iota, g, k, x, y) = \prod_{d\in D} \int P(\theta_d|\alpha)P(z|\theta_\sigma)\prod_{e\in d}\sum_{Z\in Z}P(e.t|z, e.a, y)P(e.a|z, \beta)\prod_{(f,v)\in d}\sum_{Z\in Z}P(v|z, f, \iota, g, k, x, y)P(f|z, \eta)d\theta_\sigma \quad (1)$$

where Z is the number of treatment patterns and each patient medical record d consists of a set of clinical events e and a set of pairs of patient features and their values (f,v). For a more detailed description of the learning process, please refer to [31].

Multi-label classification for clinical outcome prediction

Using generated multinomial distributions of treatment patterns on an EMR d (i.e., $\theta_d = \{\theta_{d,z_1}, \theta_{d,z_2}, \cdots, \theta_{d,z\ |Z|}\}$), the prediction of clinical outcomes for d can be defined as a multi-label classification task that is associated with one or more classes of clinical outcomes C.

The proposed clinical outcome prediction model implements a function $f: \Theta \times C \rightarrow R$ that returns a value for each pair $(\theta_d, c) \in \Theta \times C$, which is the evidence for the fact that a test patient instance with the distributions of treatment patterns, θ_d, can be classified under an appropriate set of clinical outcome classes C_d, $C_d \in$ Powerset(C). The function of the actual value f(·,·) can be transformed into a

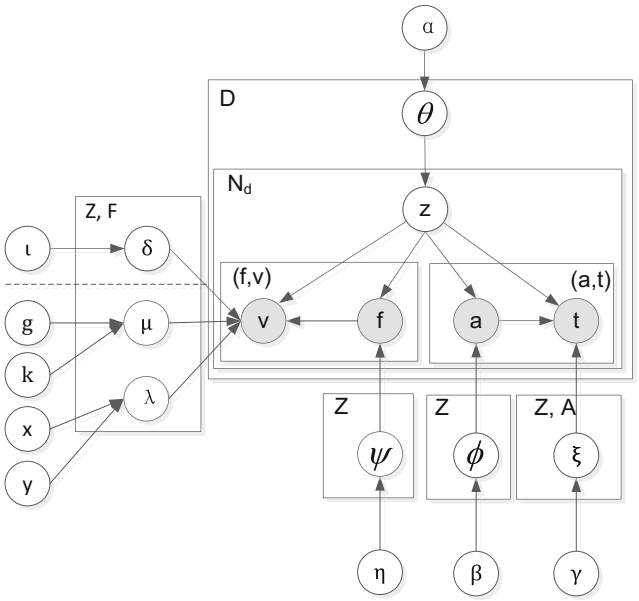

Fig. 2 The probabilistic topic model applied for treatment analysis [31].

ranking function $r(\cdot,\cdot)$, which is one-to-one mapping onto 1, 2, ⋯, |C| such that $f(\theta_d,c_1)>f(\theta_d,c_2)$, then $r(\theta_d,c_1)>r(\theta_d,c_2)$. If C_d is the set of appropriate classes for a test EMR d w.r.t its treatment pattern distributions θ_d, then a suitable classifier tends to organize the classes in C_d over those that are not in C_d.

In Table 1, for example, according to the defined problem formalization, being $\Theta=\{\theta_1,\theta_2,\theta_3\}$, the task is to associate one or more classes of clinical outcomes C={LOS in 7 days, Readmission in one month, Normal discharge} to a patient instance d_4 of which the probability distributions of treatment patterns is θ_4. For this, we can generate a multi-label classifier from both Θ and their associated clinical outcomes C, then predict the clinical outcomes for d_4 using the learned model.

To this end, we adopt three popular multi-label classification algorithms [16, 27], i.e., Binary Relevance (BR) [21], RAndom k-labELsets (RAKEL) [22], and Multi-label k-nearest neighbors (MLkNN) [23], for clinical outcome prediction. BR is a well-known one-against-all strategy. It learns m binary classifier for each different label [16, 20, 27]. When making a prediction, each binary classifier predicts whether its label is relevant for the given example or not, resulting in a set of relevant labels. RAkEL is an ensemble method for multi-label classification. It randomly breaks the initial set of labels into a number of small-sized label sets, and trains a label power-set classifier using each set of labels [16, 21, 27]. MLkNN is a theoretical approach to multi-label classification that adapts the combination of the k recovered cases of the classical k-nearest neighbors (kNN) algorithm to multiple label problems [16, 22, 27]. As these three algorithms are recognized by the community as reference algorithms, we can investigate whether a multi-label model is suitable for the prediction of clinical outcomes for CTPs.

In reality, there is limited information for a patient when he/she is in admission of the CTP. As time progresses, the information on patient status and the conditioned treatments becomes continuously updated such that the model can generate prediction updates whenever new information is received. In this sense, time plays an essential role in CTP outcome prediction, allowing calculating multiple outcomes' estimations at different clinical stages of the patient, and providing more accurate estimation since more information is available.

Formally, for a partial EMR $d=(f_1, v_1), (f_2, v_2), \cdots, (f_m, v_m), (a_1,t_1), (a_2, t_2), \cdots, (a_n,t_n)$ and a discovered treatment pattern z, we calculate the treatment pattern assignment P(z|d) as follows:

$$
\begin{aligned}
P\left(z\middle|d\right) &\propto P\left(d\middle|z\right) P\left(z\right) = P\left(z\right) \prod_{j=1}^{m} P\left(f_j\middle|z\right) P\left(v_j\middle|f_j,z\right) \prod_{i=1}^{n} P\left(a_i\middle|z\right) P\left(t_i\middle|a_i,z\right) \\
&= \frac{\sum_{d\in D_{training}} \theta_{d,z}}{\left|D_{training}\right|} \cdot \begin{cases} \prod_{j=1}^{m} \Psi_{z,f_j} \delta_{z,f_j,v_j} \prod_{i=1}^{n} \phi_{z,a_i} \xi_{z,a_i,t_i} & : f_j \text{ is a categorical feature} \\ \prod_{j=1}^{m} \Psi_{z,f_j} \; \text{Normal}\left(v_i\middle|\mu_{z,f_j},\lambda_{z,f_j}\right) \prod_{i=1}^{n} \phi_{z,a_i} \xi_{z,a_i,t_i} & : f_j \text{ is a numerical feature} \end{cases}
\end{aligned}
\tag{2}
$$

where $\theta_{d,z}$, ψ_{z,f_j}, δ_{z,f_j,v_j}, μ_{z,f_j}, λ_{z,f_j}, ϕ_{z,a_i}, and ξ_{z,a_i,t_i} are learned parameters of TPM from the training dataset. Apparently, each instance associated with a partial EMR d will have a treatment pattern assignment that will "evolve" as time advances.

Experiments and results

In this section, we set out to answer two questions using the proposed approach:

Table 1 An Example of a Multi-label Classifier for Clinical Outcome Prediction

	Clinical outcomes					
	TP1	TP2	TP3	LOS in 7 days	Readmission in 6 months	Discharge
θ_1	0.012	0.504	0.484	N	Y	Y
θ_2	0.001	0.218	0.781	N	Y	Y
θ_3	0.003	0.265	0.712	Y	N	Y
θ_4	0.023	0.413	0.564	?	?	?

1. How does the proposed approach compare to the methods directly using typical multi-label classification algorithms on EMRs without the efforts on latent treatment pattern discovery?, and
2. How does treatment analysis contributions to CTP outcome prediction? Or more specifically, how do the treatment patterns discovered characterize the set of patients and lead to specific clinical outcomes in CTPs?

We present our results motivated by the questioned above. In Section 4.1, we present the data used and describe the experiments used for CTP outcome prediction. Section 4.2 presents the results of the discovered treatment patterns from the experimental data set. In Section 4.3, we depict the performance of our method for clinical outcome prediction in the multi-label classification on both one time and sequential outcome prediction.

Experimental setup

The experimental data-set is collected from the Cardiology department of the Chinese PLA general hospital. The CTP of unstable angina is selected in the case study. Unstable angina or sometimes referred to as acute coronary syndrome is a warning sign that a heart attack may happen soon, and causes unexpected chest pain, and usually occurs while resting [23]. The population of unstable angina is huge, especially for aged people and those with associated disease such as hypertension and diabetes [24, 25]. Numerous factors impact on unstable angina patient treatment strategies and the application of interventions, such that variations could occur in a mandatory manner for unstable angina CTP. Thus, clinical outcome prediction on the unstable angina CTP will be of significant value and interest since it could provide physicians comprehensive understanding on the expected results of the CTP, and explicit suggestions for the adjustment and improvement of the CTP in concern for the patient's benefits.

In this case study, a collection of EMRs of 3492 patients following the unstable angina CTP (from 2007 to 2010) was extracted from EMRs system of the hospital to demonstrate the ability of the proposed method in CTP outcome prediction. These patients' EMRs have 84 patient features, and 201340 clinical events within 389 treatment activity types. Selected characteristics of patients are displayed in Table 2. Mean age of the patients was 63.1 years. Almost seventy percent was men, and the most five common comorbidities were hypertension (67.38%), diabetes (27.52%), hyperlipidemia (19.01%), cancer (9.02%), and renal insufficiency (5.78%).

To investigate the patient benefits from the CTP, we set out three classes (i.e., LOS, readmission time, discharge type), and 10 variables of clinical outcomes based on the suggestions of clinical experts, as shown in Table 3. Each EMR in the collection can be categorized into one or several outcome variables. All experiments were performed on a Lenevo Compatible PC with an Intel Pentium IV CPU 2.8 GHz, 4G byte main memory running on Microsoft Windows 7.

Treatment Pattern discovery

As we mentioned above, we employ a probabilistic topic model, i.e., treatment pattern model, which is presented in our previous work [3], to analyze treatment behaviors of CTPs. Regarding parameters of TPM, we set Dirichlet priors α, β, γ, η, and ι as 0.1, 0.01, 0.01, 0.01, and 0.01, respectively, which are common settings in literature. In addition, we set $g_f = \widehat{\mu}_f$, k_f=1, x_f=1, and $y_f = 1 + \widehat{\lambda}_f^2$, where $\widehat{\mu}_f$ and $\widehat{\lambda}_f$ are the empirical mean and variance of the value of patient feature f. The number of iterations of the Markov chain for Gibbs sampling is set to 1000. Unusually, Gibbs sampling converges before 1000 iterations.

Table 2 Selected characteristics of the study patients

Full cohort (n = 3492)	
Demographics	
Age (years) (mean, SD)	63.1, 12.24
Male gender (n(%))	69.9%
Comorbidities (n(%))	
Hypertension	67.38%
Diabetes	27.52%
Hyperlipidemia	19.01%
Cancer	9.02%
Renal insufficiency	5.78%
Vital signs	
SBP	132.2, 17.85
DBP	77.4, 10.4

Table 3 Clinical Outcome Variables Used in This Study

C_1	LOS ≤ 7 days	C_6	Readmission in six months
C_2	7 days < LOS ≤ 14 days	C_7	Readmission larger than six months or without readmission
C_3	14 days < LOS ≤ 28 days	C_8	Normal discharge
C_4	LOS > 28 days	C_9	Transfer
C_5	Readmission in one month	C_{10}	Death

Regarding the number of treatment patterns Z, we set Z to 3. In this case, the proposed approach can achieve the best performance on CTP outcome prediction, as illustrated in the next subsection. Table 4 shows the derived treatment patterns with the number of patterns to be 3. For each pattern, we include the top-ranked 20 patient features and their values, ranked by the probability of a feature and its value given the pattern, i.e. $P(f, v|z)=P(f|z)P(v|f, z)$. In addition, top-ranked 50 treatment activities are listed in Table 4, ranked by the probability of an activity and its occurring time stamps given the pattern, i.e., $\sum_{t\in T}P(a, t|z)$. As shown in Table 4 different patterns might have different patient features and values. For example, the average value of age in the treatment pattern-1 is 60.13, and that in the treatment pattern-2 is 68.21. We can distinguish this from the value (range) of patient features. As well, discovered treatment patterns show latent correlations between patient features, from which we can find some interesting phenomena. For example, all three patterns show that there exist latent correlations between unstable angina and Hypertension, which indicate that there is a large probability for patients with unstable angina to have the complication *Hypertension* at the same time.

Furthermore, there obviously exist different treatment behaviors given different patient conditions. For example, patients who follow the discovered pattern-1 are probably performed stent placement, PTCA and PCI surgeries, which are representative treatment interventions for the unstable angina CTP. Patients who follow pattern-2 might undergo conservative treatments at first, and then are performed stent placement and PCI surgeries. Additionally, we notice that patients who follow pattern-2 have a significant probability to have coronary stenosis, and are generally with more complications than patients who follow pattern-1, which may result in the conservative treatments, and the delay of surgery. Patients who follow pattern-3 are with more complications in comparison with the other patterns, e.g., hypertension, diabetes, insufficient kidney, etc. In particular, several specific treatment interventions, such as "Coronary Artery Bypass Grafting" which was note generated for the other patterns, are observed in this pattern-3.

Clinical outcome prediction

Using TPM, underlying treatment patterns and their distributions in the collection of EMRs are derived. Based on the generated results of treatment pattern analysis, we used a well-known multi-label classification tool, i.e., MEKA [26], for the further task of clinical outcome prediction. Specifically, we selected three typical multi-label classification algorithms, BR, RAkEL, and MLkNN of MEKA, and compare their performances on the collected EMRs with and without the step of treatment pattern analysis.

We employ Accuracy and Hamming-Loss to evaluate performance on the outcome prediction [27]. Accuracy is measured by the Hamming-Loss which symmetrically measures how close the true set of clinical outcome labels C_j is to the predicted set of clinical outcome labels P_j. It is not as 'easy' as Hamming-Loss. The value of Accuracy is calculated by [27]:

$$\text{Accuracy} = \frac{1}{|D|}\sum_{j=1}^{|D|}\frac{\left|P_j \bigcap C_j\right|}{\left|P_j \cup C_j\right|} \tag{3}$$

Hamming-Loss evaluates how often a test instance is categorized in a class. The higher the value of the Hamming-Loss is, the more accurate class recommendations of the classifier. The value of the Hamming-Loss is calculated as follows [27]:

$$\text{Hamming}-\text{Loss} = \frac{1}{|D|}\sum_{j=1}^{|D|}\frac{1}{|C|}\left|P_j \Delta C_i\right| \tag{4}$$

where D is the collection of EMRs, C is a set of clinical outcome labels, P_j is the predicted set of clinical outcome labels of the test EMR d_j, C_j is the true set of clinical outcome labels of d_j, and Δ stands for the symmetric difference of two sets and corresponds to the XOR operation in Boolean logic.

To calculate Accuracy and Hamming-Loss, we split the dataset into a training set and a testing set, and evaluate the performance by 10-fold cross-validation in all experiments.

One time prediction

The one time prediction assumes that all information is available when the CTP's execution of a patient is complete. This assumption was followed implicitly by some of the earlier studies, mostly to understand the feasibility of using EMRs to support CTP analysis. We will follow the same reasoning and leave the more realistic scenario of sequential prediction for the next section.

Table 4 Discovered Treatment Patterns from the Unstable Angina EMRs.

Treatment pattern-1		
Patient features	Attack of angina	True
	Creatinine	Normal
	Sodium	Normal
	Hemoglobin measurement	Normal
	Creatine kinase	Normal
	Mean corpuscular hemoglobin amount	Normal
	Lactate dehydrogenase	Normal
	Platelet counts	Normal
	Mean corpuscular hemoglobin concentration	Normal
	Potassium	Normal
	Mean platelet volume measurement	Normal
	Low-density lipoprotein cholesterol	Normal
	Gender	Male
	Qualitative urine glucose test	907.052, 246.692
	Creatine kinase isoenzyme	Normal
	Total cholesterol	Normal
	Glucose	Normal
	Hypertension	True
	Triglycerides	Normal
	Age	60.131, 11.115
Treatment intervention	Biochemical tests items, Antianginal drugs, Antiplatelet drugs, Drurgs for cardiovascular system disease, Ultrasonography, Drugs for regulating blood lipids, Routine care, Routine blood test, Anticoagulants, Admission, Second-grade care, Coagulation test, Drug replacement, Routine urine test, Routine faeces test, Serum test, Drugs for lowering blood pressure, Discharge, Identification of blood type, Occult blood test, Draw off blood, Coronary angiography, Local anesthetics, Drugs of calcium channel blockers, Anesthesia, Radiation, Analysis of blood and ions, Puncture, Biochemical test, Check items of coagulation test, Sedative hypnotic and anti-anxiety drugs, Durgs of angiotensin-converting enzyme inhibitiors, Drugs for expanding peripheral vascular, Injection ways, ECG, Stent placement, PCI, Drugs of proton pump inhibitors, Troponin T, First-grade care, Oxygen inhalation, Parenteral nutrition drugs, Drugs for curing angina, Monitoring of ECG and blood oxygen, Other kinds of antidiabetic drugs	
Treatment pattern-2		
Patient features	Sodium	True
	Attack of angina	Normal
	Lactate dehydrogenase	Normal
	Mean corpuscular hemoglobin amount	Normal
	Creatine kinase	Normal
	Platelet counts	Normal
	Potassium	Normal
	Low-density lipoprotein cholesterol	Normal
	Creatine kinase isoenzyme	Normal
	Age	68.211, 13.7578
	Mean corpuscular hemoglobin concentration	Normal
	Creatinnine	Normal
	Total cholesterol	Normal
	Mean platelet volume measurement	Normal
	Hemoglobin measurement	Normal
	Hypertension	True
	Triglyceride	Normal
	Gender	Male
	High-density lipoprotein cholesterol	Normal
	Glucose	Normal

Table 4 (continued)

Treatment pattern-1		
Treatment intervention	Biochemical tests items, Draw off blood, Antianginal drugs, Diuretics, Drug replacement, Sedative hypnotic and anti-anxiety drugs, Routine blood test, Electrolyte regulating drugs, Analysis of blood and ions, Anticoagulants, Discharge, Expert consultation, Antiplatelet drugs, Oxygen inhalation, Ultrasonography, Drugs for lowering blood pressure, Drugs for regulating blood glucose, Drugs for cardiovascular system disease, Drugs of calcium channel blockers, Drugs used for non-cardiovascular system disease, Coronary angiography, Troponin T, Puncture, Anesthesia, Local anesthetics, Biochemical test, Drugs of proton pump inhibitors, First-grade care, Coagulation test, Stent placement, Second-grade care, PCI, Antiarrhythmic drugs, Parenteral nutrition drugs, Radiation, ECG, Routine care, Cephalosporins, Other kinds of antidiabetic drugs, Monitoring of ECG and blood oxygen, Drugs for regulating blood lipids, CT, Drugs of angiotensin-converting enzyme inhibitors, Transferred, Drugs of digitalis glycosidic, Test for TB, Occult blood test, Expectorants, Check items of coagulation test, Routine faeces test	
Treatment pattern-3		
Patient features	Qualitative urine glucose test	511.854, 359.980
	Platelet volume measurement	0.197, 0.060
	Measurement of Troponin T	High
	Quantitative determination of creatine kinase isoenzyme	High
	Diabetes	True
	Hypertension	True
	Brain natriuretic peptide precursor	High
	Cardiac insufficiency	True
	Insufficiency of kidney function	True
	Lactate dehydrogenase	High
	Hyperlipidemia	True
	Attack of angina	True
	Platelet aggregation	Normal
	History of CHD	True
	Tumor	True
	Artery stenosis	True
	Creatine kinase	High
	Post-PCI	True
	Pulmonary disease	True
	Mean platelet volume measurement	High
Treatment intervention	Diuretics, Biochemical tests items, Draw off blood, Routine blood test, Electrolyte regulating drugs, Antianginal drugs, Drugs for regulating blood glucose, Analysis of blood and ions, Drugs for lowering blood pressure, Sedative hypnotic and anti-anxiety drugs, Oxygen inhalation, Transferred, Radiation, Blood pressure, Plasma and plasma substitutes, Troponin T, Drugs of digitalis glycosidic, Biochemical test, Drugs for cardiovascular system disease, Antiplatelet drugs, Drugs of calcium channel blockers, Drugs of angiotensin-converting enzyme inhibitors, Coagulation test, Cephaloporins, Test for TB, Penicillin, Parenteral nutrition drugs, Expectorants, Anticoagulants, Second-grade care, Discharge, Antiarrhythmic drugs, Narcotic analgesics, Other kinds of antidiabetic drugs, X-ray, Anti-shock and vasoactive drugs, Puncture, Drugs of proton pump inhibitors, Calcium-regulating drugs, Antiepileptic and anticonvulsants drugs, Drugs used for non-cardiovascular system disease, First-grade care, Antispasmodic drugs gastrointestinal, Echocardiography, Monitoring of ECG and blood oxygen, Antipyretic and analgesic anti-inflammatory drugs, Ultrasonography, ECG, CABG, Intravenous anesthetics	

As mentioned above, constructing of a TPM is to fit latent treatment patterns to the collection of EMRs such that high-dimensional features of CTPs can be reduced to low-dimensional features. In Section 3.2, we conducted treatment pattern analysis using TPM with different number of treatment patterns (Z=1, 2, 3, …, 10) from the experimental EMRs. Based on the results of treatment analysis, we performed MLkNN, RAkEL, and BR for clinical outcome prediction. In comparison, we conducted MLkNN, RAkEL, and BR on the collection of EMRs directly without efforts on treatment analysis. For this, all 69 patient features and 382 treatment activity types are taken as input features for clinical outcome prediction.

Figure 3 shows detailed experimental results of TPM-Z+MLkNN, TPM-Z+RAkEL, TPM-Z+BR (Z=1, 2, 3, … , 10), and MLkNN, RAkEL, and BR on the performance of one time prediction. In comparison with the typical multi-label classification algorithms, i.e., MLkNN, RAkEL, and BR, the proposed three hybrid methods achieve the better performance on CTP outcome prediction even when Z has a very small

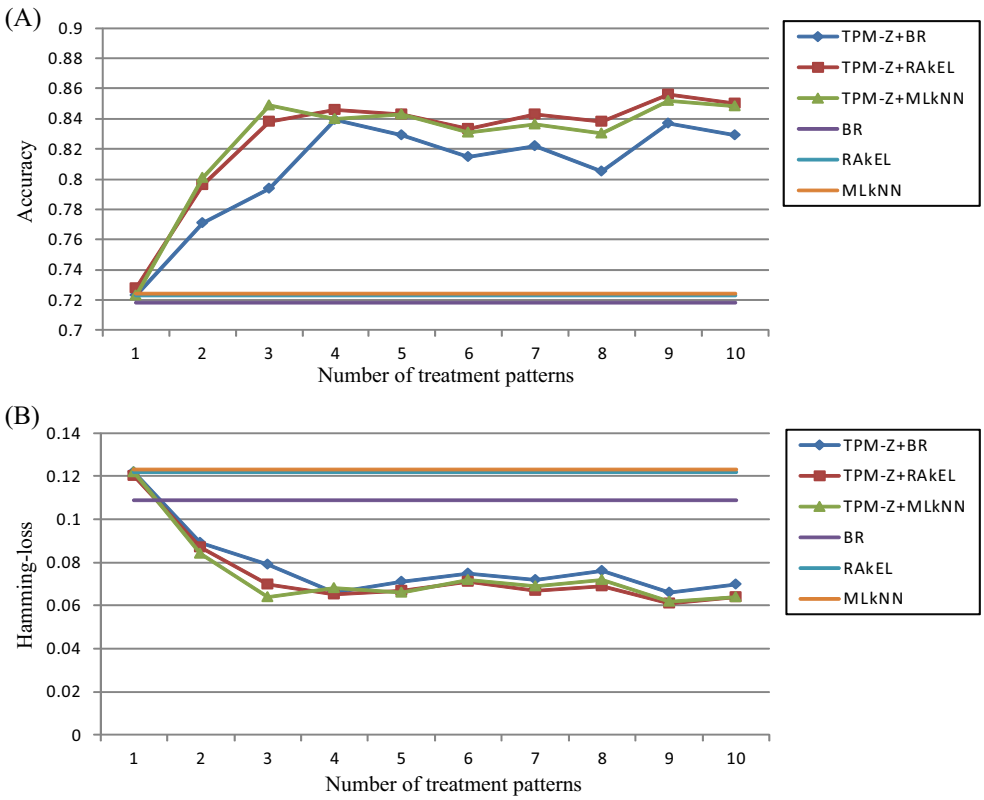

Fig. 3 The performance of the proposed approach with different number of treatment patterns.

value (e.g., Z=2). For example, the Accuracy achieved by TPM-3+MLkNN and MLkNN are 0.849 and 0.724, respectively, i.e., roughly 17% improvements in the quality of Accuracy with the treatment analysis, which is quite remarkable for clinical outcome prediction.

In terms of the number of treatment patterns Z, The experimental results show that Z has weak impacts on the prediction performance for the proposed hybrid methods when Z is larger than a threshold value. As depicted in Fig. 3a and b, the curves of Accuracy (or Hamming-loss) are close with each other for TPM-Z+MLkNN, TPM-Z+RAkEL, and TPM-Z+BR. In addition, as Z increases, the Accuracy increases (Hamming-Loss decreases) quickly at first, and then remains stable with the further increases of Z for the proposed hybrid methods (i.e., TPM-Z+MLkNN, TPM-Z+RAkEL, and TPM-Z+BR). We observe that the proposed hybrid methods achieve the stable performance when Z is larger than 3 in terms of Accuracy and Hamming-loss. To this end, TPM-3+MLkNN (Z=3) might be the suitable method for CTP outcome prediction, w.r.t the experimental dataset.

In terms of run-times of the outcome prediction, the proposed TPM-Z+MLkNN, TPM-Z+RAkEL, TPM-Z+BR outperform algorithms MLkNN, RAkEL, and BR significantly, as shown in Table 5. Note that an important factor influencing run-times of multi-label classifications is the variable dimensions of data. As an efficient step of our approach, the treatment analysis can not only discover the underlying treatment patterns from EMRs, but also reduce the dimensions of CTP features of EMRs significantly (from hundreds to be several), which is

Table 5 Run-times of the CTP Outcome Prediction for the Experimental Dataset (*sec.*)

	Number of treatment patterns *Z*									
	1	2	3	4	5	6	7	8	9	10
TPM-*Z*+BR	0.249	0.291	0.702	1.289	1.86	1.712	2.266	2.288	2.75	3.165
TPM-*Z*+RAkEL	13.061	11.692	12.969	16.036	19.609	19.566	21.409	21.952	22.379	24.905
TPM-*Z*+MLkNN	6.535	1.595	1.691	1.849	1.596	3.371	3.905	3.753	3.473	4.727
BR	906.217									
RAkEL	1975.377									
MLkNN	357.678									

benefitted to multi-label classification. The experimental results validate our assumption explicitly that the treatment analysis can improve the performance on CTP outcome prediction.

Sequential prediction

To validate the sequential prediction on clinical outcomes in the execution of CTPs, we split the dataset into two parts: a training set comparing 66% EMRs of the overall dataset, and a test set comparing 33% EMRs of the dataset. We learned a TPM model for the treatment analysis and a clinical outcome prediction model, all from the training set, and then test the prediction model on the test set taking into account the elapsed time τ.

Intuitively, the quality of predictions should gradually improve as new information about treatments become available, particularly when also considering the elapsed time. Figure 4 shows the performance on the Accuracy and Hamming-loss over a specific time period from admission to a particular time instance τ using TPM-3+BR, TPM-3+RAkEL, and TPM-3+MLkNN (i.e., the number of treatment patterns Z=3). For comparison purpose, at each of these moments, we also obtain prediction updates from the benchmark algorithms BR, RAkEl, and MLkNN. Since the general LOS of the unstable angina TP is between 2 and 3 weeks, we set the whole prediction period as 21 days. As expected from Fig. 4, the Accuracy of the proposed hybrid methods increases gradually (Hamming-loss decreases) with the increases of τ. On the contrary, the Accuracy of the benchmark algorithms does not increase but decrease slightly (Hamming-loss increases) with the increases of τ, which lead to worse performance than the proposed approach. The main cause of this phenomenon is that there are many variant treatment behaviors emerging during the execution of CTPs. If there is a high number of variations, the individual treatment behaviors will be very different such that the predicted value will be unreliable than in the situation where most treatment behaviors are similar. On the other side, our approach is robust with these variant treatments even that the prediction quality can be improved when more medical information is obtained from the time progress.

In addition, the proposed TPM-3+BR, TPM-3+RAkEL, and TPM-3+MLkNN outperform the benchmark algorithms on both metrics when $\tau > 8$, i.e., after 8 days of admission, although they are close with each other during the first stage of the CTP. Since the whole prediction period is 21 days, and the average LOS for the collected EMRs is 18 days. We consider it is an acceptable result. It indicates that the treatment analysis is able to improve the prediction quality.

Note that MLkNN, RAkEL, and BR are discriminative algorithms. The main purpose of these benchmark algorithms is to classify EMRs regarding specific clinical outcomes. On the other hand, the proposed hybrid methods, by combining treatment topic analysis and multi-label classification, are typically generative. Thus, they cannot only classify EMRs regarding to specific clinical outcomes, but also reveal details of underlying treatment behaviors for CTPs. Thus, they provide useful insight into CTPs, and can hence be straightforwardly included explicitly as background knowledge for further analytical objectives. Since CTPs may not perform as desired according to obtained clinical outcomes, and have to be readjusted/redesigned consequently, actionable knowledge discovered from our models can be used as a feedback tool that helps in auditing and analyzing already enacted CTPs, and can also provide a valuable reference for medical staff to redesign and continuously optimize CTPs.

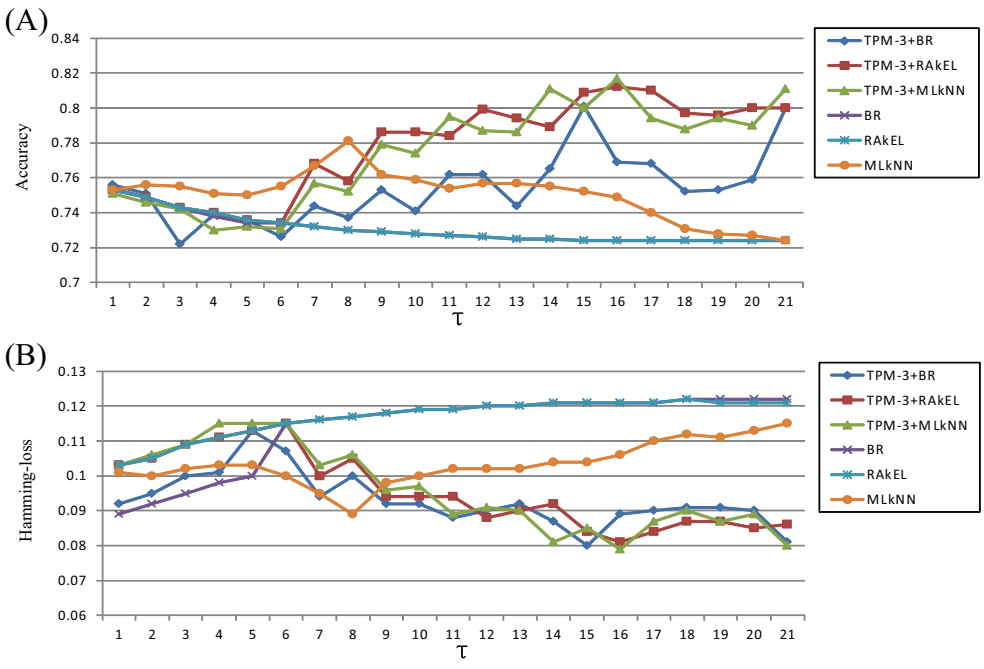

Fig. 4 Sequential prediction using the collected unstable angina EMRs.

Conclusion

Inspired by machine learning methods from the topic modeling community, we convert a collection of EMRs into a series of documents consisting of sets of discrete clinical words, i.e. patient features and their values, and treatment interventions and their occurring time stamps. These sets are then mined for treatment analysis, using a probabilistic model presented in our previous work [3]. The generated results of treatment analysis can be potentially exploited for further applications. In particular, we present a clinical outcome prediction model by applying typical multi-label classification algorithms on the generated results of treatment analysis. Unlike current clinical outcome prediction methods that commonly consider a prior knowledge, such as demographic data or patient's physiological conditions in terms of a single outcome class, e.g., LOS, Readmission time, etc., the proposed approach bases its estimation on the temporal evolution of the patient, and thus continuously generates predictions in terms of multiple clinical outcomes, as new information about treatments becomes available in form of EMR updates during the execution of CTPs.

In an evaluation using 3492 pieces of EMRs of the unstable angina CTP, we showed that the discovered treatment patterns, as effective reflections of the real executing scenarios of CTPs, can improve the quality of clinical outcome prediction significantly. In terms of the accuracy of prediction, the proposed approach outperforms the benchmark algorithms. It indicates that the generated results of treatment analysis, therefore, can help understand the patient conditions, treatment inventions, and clinical outcomes resulted from the execution of CTPs, in an integrated view.

In future studies, the proposed approach will be extended to predict a vast range of CTP outcomes, including physiologic measures (e.g., blood pressure), laboratory test results (e.g., serum cholesterol) and patient-reported health states (e.g., functional status and symptoms) [30]. In this way, it may provide insights into the CTP, gain a deeper understanding of the situations in which the proposed prediction technique performs well, and improves the utilization of health-care resources while maintaining and even enhancing the quality of health service in CTPs [18].

The issue of meaningful or secondary use of EMRs represents a promising and interesting research direction in health informatics [3]. Our study indicates the feasibility of exploring EMRs to support treatment analysis and CTP outcome prediction. Note that the proposed treatment analysis model offers many avenues for future expansions and applications, e.g., personalized treatment recommendations, treatment grouping and identification within the same therapy and treatment intention, and anomaly detection from normal treatment behaviors, etc. As our future work, we will address these tasks by exploiting the potential of our approach on a larger scale of EMR collections.

Acknowledgement This work was supported by the National Nature Science Foundation of China under Grant No 81101126 and 61562088, and the Fundamental Research Funds for the Central Universities under Grant No 2014QNA5014. The author would like to give special thanks to all experts who cooperate in the evaluation of the proposed method.

References

1. Huang, Z., Juarze, J. M., Duan, H., and Li, H., Length of stay prediction for clinical treatment process using temporal similarity. *Expert Sys Appli* 40(16):6330–6339, 2013.
2. Huang, Z., Lu, X., Duan, H., and Fan, W., Summarizing clinical pathways from event logs. *J Biomed Inform* 46(1):111–127, 2013.
3. Huang, Z., Dong, W., Ji, L., Gan, C., Lu, X., and Duan, H., Discovery of clinical pathway patterns from event logs using probabilistic topic models. *J Biomed Inform* 47:39–57, 2014.
4. Gang, D., Zhibin, J., Xiaodi, D., and Yang, Y., Intelligent ensemble T–S fuzzy neural networks with RCDPSO_DM optimization for effective handling of complex clinical pathway variances. *Comput Biol Med* 43(6):613–634, 2013.
5. Zhengxing Huang, Xudong Lu, Chenxi Gan, and Huilong Duan. Variation prediction in clinical processes. In M. Peleg, N. Lavrac, & C. Combi (Eds.), Artificial Intelligence in Medicine. Lecture notes in computer science (vol. 6747, pp. 286–295). Berlin/Heidelberg: Springer
6. Yang, C.-S., Wei, C.-P., Yuan, C.-C., and Schoung, J.-Y., Predicting the length of hospital stay of burn patients: comparisons of prediction accuracy among different clinical stages. *Decis Support Syst* 50(1):325–335, 2010.
7. Ng, S.-K., McLachlan, G. J., and Lee, A. H., An incremental EM-based learning approach for on-line prediction of hospital resource utilization. *Artif Intell Med* 36(3):257–267, 2006.
8. Tu, J. V., and Michael, R. J. G., Use of a neural network as a predictive instrument for length of stay in the intensive care unit following cardiac surgery. *Comput Biomed Res* 26(3):220–229, 1993.
9. Kim, D., Shin, H., Young, S. S., and Ju Han, K., Synergistic effect of different levels of genomic data for cancer clinical outcome prediction. *J Biomed Inform* 45(6):1191–1198, 2012.
10. Adeyemi, S., Demir, E., and Chaussalet, T., Towards an evidence-based decision making healthcare system management: modelling patient pathways to improve clinical outcomes. *Decis Support Syst* 55(1):117–125, 2013.
11. Charlene, R., Staggers, W. N., and Tamara, L., Reviewing the impact of computerized provider order entry on clinical outcomes: the quality of systematic reviews. *International Journal of Medical Informatics* 81(4):219–231, 2012.
12. Zhengxing Huang, Xudong Lu, Huilong Duan. Anomaly detection for clinical processes, AIMA2012
13. Stefania Montani, Giorgio Leonardi: Retrieval and clustering for supporting business process adjustment and analysis. Inf. Syst. 40: 128–141, 2014.
14. Lu, X., Zhengxing, H., and Huilong, D., Supporting adaptive clinical treatment processes through recommendations. *Comput Methods Prog Biomed* 107(3):413–424, 2012.
15. Huang, Z., Lu, X., and Duan, H., Latent treatment pattern discovery for clinical processes. *J Med Syst* 37(2):1–10, 2013.
16. Tsoumakas, G., and Katakis, I., Multi-label classification: an overview. *Int J Data Ware Min* 3(3):1–13, 2007.
17. Lehman, L., Adams, R., Mayaud, L., Moody, G., Malhotra, A., Mark, R., Nemati, S. A Physiological Time Series Dynamics-Based Approach to Patient Monitoring and Outcome Prediction,

IEEE Journal of Biomedical and Health Informatics, doi: 10.1109/JBHI.2014.2330827.
18. Gang, D., Jiang, Z., Yang, Y., and Diao, X., Clinical pathways scheduling using hybrid genetic algorithm. *J Med Syst* 37:9945, 2013.
19. Blei, D. M., Ng, A. Y., and Jordan, M. I., Latent Dirichlet allocation. *J Mach Learn Res* 3(4–5):993–1022, 2003.
20. Jesse Read, Bernhard Pfahringer, Geoff Holmes, and Eibe Frank. 2009. Classifier Chains for Multi-label Classification. In Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases: Part II (ECML PKDD '09), pp:254–269, Springer-Verlag, Berlin, Heidelberg, 2009.
21. Elisseeff, A., and Weston, J., A kernel method for multi-labelled classification. *In Advances in Neural Inform Proc Syst* 14:681–687, 2001.
22. Zhang, M.-L., and Zhou, Z.-H., *A k-nearest neighbor based algorithm for multi-label classification, 2005 I.E. international conference on granular computing, 2:718–721*, 2005.
23. Catherwood, E., and O'Rourke, D. J., Critical pathway management of unstable angina. *Prog Cardiovasc Dis* 3(3):121–148, 1994.
24. 2012ACCF/AHA focused update of the guideline for the management of patients with Unstable Angina/Non-ST-Elevation myocardial infarction (updating the 2007 guideline and replacing the 2011 focused update). Circulation, 126(7):875–910, 2012.
25. Wei Dong, Zhengxing Huang, Lei Ji, Huilong Duan. A genetic fuzzy system for unstable angina risk assessment, BMC Medical Informatics and Decision Making, 14:12, 2014
26. MEKA, http://meka.sourceforge.net/. Last access on 2014-4-30.
27. Madjarov, G., Kocev, D., Gjorgjevikj, D., and Džeroski, S., An extensive experimental comparison of methods for multi-label learning. *Pattern Recogn* 45(9):3084–3104, 2012.
28. Gultepe, E., Green, J. P., Hien, N., Jason, A., Timothy, A., and Ilias, T., From vital signs to clinical outcomes for patients with sepsis: a machine learning basis for a clinical decision support system. *J Am Med Inform Assoc* 21(2):315–325, 2014.
29. Traber Davis Giardina, Shailaja Menon, Danielle E Parrish, Dean F Sittig, Hardeep Singh. Patient access to medical records and healthcare outcomes: a systematic review, J Am Med Inform Assoc, Published Online First: 23 October 2013, doi:10.1136/amiajnl-2013-002239.
30. Zifang Huang, Mei-Ling Shyu, Tien, J.M., Vigoda, M.M., Birnbach, D.J. Knowledge-Assisted Sequential Pattern Analysis With Heuristic Parameter Tuning for Labor Contraction Prediction, IEEE Journal of Biomedical and Health Informatics, 18(2):492–499, 2014.
31. Huang, Z., Dong, W., Ji, L., Bath, P., and Duan, H., On mining latent treatment patterns from electronic medical records. *Data Min Knowl Disc* 29(4):914–949, 2015.