e-Health – For Continuity of Care
C. Lovis et al. (Eds.)

doi:10.3233/978-1-61499-432-9-715

Automatic Variance Analysis of Multistage Care Pathways

Xiang LI[a,1], Haifeng LIU[a], Shilei ZHANG[a], Jing MEI[a], Guotong XIE[a], Yiqin YU[a], Jing LI[a], and Geetika T. LAKSHMANAN[b]
[a]*IBM Research, Beijing, China*
[b]*IBM T. J. Watson Research Center, Hawthorne, NY, USA*

Abstract. A care pathway (CP) is a standardized process that consists of multiple care stages, clinical activities and their relations, aimed at ensuring and enhancing the quality of care. However, actual care may deviate from the planned CP, and analysis of these deviations can help clinicians refine the CP and reduce medical errors. In this paper, we propose a CP variance analysis method to automatically identify the deviations between actual patient traces in electronic medical records (EMR) and a multistage CP. As the care stage information is usually unavailable in EMR, we first align every trace with the CP using a hidden Markov model. From the aligned traces, we report three types of deviations for every care stage: additional activities, absent activities and violated constraints, which are identified by using the techniques of temporal logic and binomial tests. The method has been applied to a CP for the management of congestive heart failure and real world EMR, providing meaningful evidence for the further improvement of care quality.

Keywords. Care Pathway, Variance Analysis, Hidden Markov Model, Temporal Logic, Care Quality Improvement

Introduction

A care/clinical pathway (CP) is a complex intervention for the mutual decision making and organization of care processes for a well-defined group of patients during a well-defined period [1]. The aim of a CP is to enhance the quality of care by improving patient outcomes and promoting patient safety. A CP may consist of multiple care stages corresponding to different disease progression conditions, where each stage contains various clinical activities such as diagnoses, medications and lab tests, as well as the temporal dependencies and numeric preconditions of the activities.

Although a CP defines a standardized care process for a specific clinical condition, actual care will inevitably deviate from the planned CP [2]. Clinicians may ignore expected activities and violate constraints during actual care. Moreover, new activities that are not defined in the CP can also be executed. Variance analysis is the process of collecting and analyzing these deviations, aimed at evaluating and revising the CP as well as reducing medical errors. However, it is difficult and error-prone for clinicians to report deviations manually [2]. Since electronic medical records (EMR) track the real care progression of patients, a potential improvement for variance analysis is to identify deviations automatically by comparing the medical records against the CP [3].

1 Corresponding Author. E-mail: lixiang@cn.ibm.com

It is a challenging problem, however, to detect the deviations for a complex CP with multiple care stages. A clinical activity can be planned in different stages (e.g., ACEI, or angiotensin-converting enzyme inhibitors, are medications targeted at every stage of CHF care) and the stage information is normally unavailable in EMR, which may result in ambiguity of mapping a care event to the activity of the correct stage. An existing solution for CP variance analysis [3] is designed for CPs that can be represented as single directed graphs, without considering CPs with multiple stages. In the neighboring field of business process management, the conformance checking method [4] that can be used to detect deviations is also limited to non-hierarchical process models and cannot distinguish duplicate activities defined in different stages.

In this paper, we address this issue by proposing a variance analysis method to identify the deviations for a multistage CP. Given a CP and a set of patient traces from EMR, we first create a hidden Markov model (HMM) from the CP, with the purpose of annotating each care event with a stage. From these aligned patient traces, we statistically examine three categories of deviations: additional activities, absent activities and violated constraints, based on the techniques of first-order linear temporal logic (FO-LTL) and binomial tests. The identified deviations can provide valuable information for the further improvement of patient safety and care quality.

1. Methods

In this section, we introduce our CP variance analysis method in detail. The inputs of our method are a CP and patient traces of a cohort in EMR, while the output is a variance analysis report including the deviations. Since there is no agreed standard for modeling a CP [3], we use a general CP representation. Other formats that have the ability to define activities and their relations can be converted to our representation easily. Let A be the universal set of clinical activities, where each activity is defined as a unique string (e.g., a code from a standard terminology, or a use-defined value such as "ACEI"). Then a CP is represented as $M = (A_M, S, \Delta, C)$, where $A_M \subset A$ is a set of CP activities, S is a set of care stages, Δ: $A_M \rightarrow S$ is a many-to-many mapping between activities and stages, and C is a set of constraints (see Section 1.2). In EMR, a care event is represented as $e = (p, a, g, v, t)$, where p is the patient ID, $a \in A$ is the care activity (e.g., "ACEI"), g is the activity type (e.g., "medication"), v is the numeric value (e.g., the drug dose), and t is the occurring time. For a specific patient p, his/her care events can be sorted by t to generate a patient trace $\sigma = <e_1, e_2, \ldots, e_l>$. As illustrated in Figure 1, our method consists of two phases: alignment and variance identification.

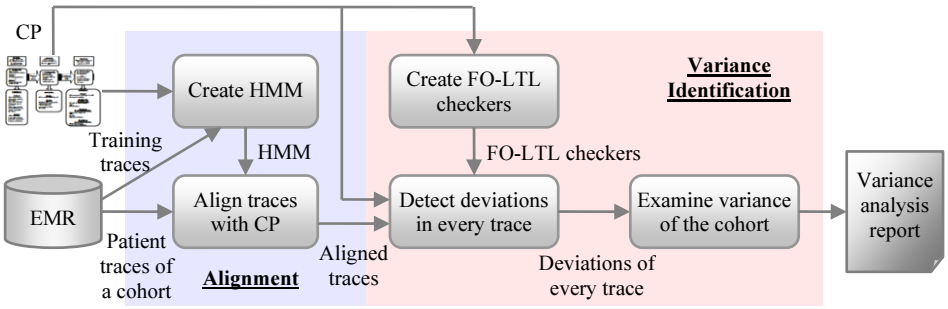

Figure 1. Method flowchart of CP variance analysis

1.1. Alignment between Patient Traces and CP

Notice that the care stage of an event is not available in EMR, and there is a many-to-many mapping between activities and stages in a multistage CP. In order to identify the deviations, we must first decide when an event matches one of alternative stages. Furthermore, in real patient data, many events that are not defined in the CP may occur frequently. The care stages of these additional events should also be determined. This ambiguity cannot be resolved correctly by independently annotating each event. To address the problem, we use a HMM [5] to model the stage transitions of a CP, by which a globally optimal stage sequence can be found for a given patient trace.

The structure of HMM for variance analysis is derived from the CP. We first create a hidden state for each care stage $s \in S$, and an observation for each activity $a \in A_M$. As new activities that are not defined in the CP can be observed in patient traces, we also create an "additional" observation for each activity type (e.g., "Additional medication"). The probabilities of the HMM are initially assigned according to the CP, where the start probabilities $P(s)$ and the transition probabilities $P(s'|s)$ satisfy the temporal constraints defined in C, and the emission probabilities $P(x|s)$ conform to the activity-stage mapping Δ. Figure 2 shows an example of HMM, where HTN (hypertension) is defined in both stages and MI (myocardial infarctions) is defined in Stage B.

Based on the HMM structure, each patient trace $\sigma = \langle e_1, e_2, \ldots, e_l \rangle$ is converted to an observation sequence $o = \langle x_1, x_2, \ldots, x_l \rangle$. For every event $e_i = (p_i, a_i, g_i, v_i, t_i)$ in σ, if $a_i \in A_M$ then $x_i = a_i$, otherwise the additional observation for g_i is assigned to x_i. Then we train the HMM using a set of training patient traces, by applying the Baum-Welch algorithm [5]. After training, we can align a given patient trace σ with the HMM by using the Viterbi algorithm [5], which finds the most likely sequence of care stages $\tau = \langle s_1, s_2, \ldots, s_l \rangle$ by maximizing $P(s_1) \cdot P(s_2|s_1) \cdot \ldots \cdot P(s_l|s_{l-1}) \cdot P(x_1|s_1) \cdot \ldots \cdot P(x_l|s_l)$.

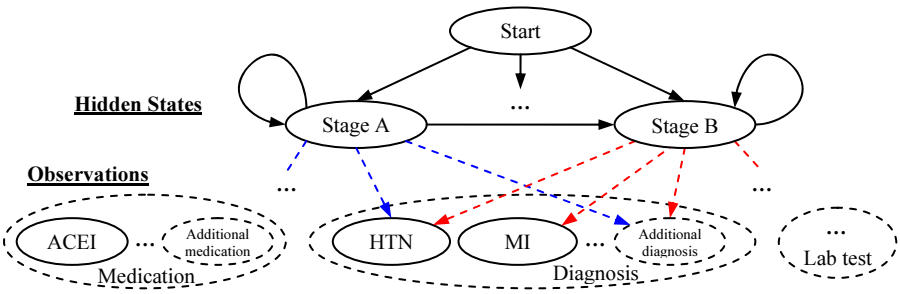

Figure 2. HMM structure derived from a CP

1.2. Variance Identification

Given an aligned patient trace, we can detect three categories of deviations. The first category is additional activities, which are executed in the patient trace but *not* defined in the CP. The second is absent activities, which are defined in the CP but *not* executed in the patient trace. To identify these activity deviations, we first build a null set $A_+(s)$ for every stage s, as well as a set $A_-(s)$ that contains all pre-defined activities in s. Then for each event e_i in the patient trace σ and its annotated stage s_i in τ, if $a_i \notin A_M$ then a_i is added to $A_+(s_i)$, otherwise a_i is removed from $A_-(s_i)$. After the iteration, $A_+(s)$ contains the additional activities for the stage s, while the absent activities remain in $A_-(s)$.

The third category is violated constraints, which are defined in the CP but *not* satisfied in the patient trace. Notice that in a CP, both temporal dependencies and numeric conditions of the activities should be formally represented. For this purpose, we use FO-LTL to define the constraints, which is constructed by combining first-order formulas by temporal and Boolean operators [6]. For detecting the violated constraints, we translate each FO-LTL constraint $c \in C$ into a Büchi automaton B_c [6], which is a non-deterministic finite automaton with an acceptance condition for input sequences. If the patient trace σ cannot be accepted by B_c, then c is a violated constraint for σ. Figure 3 shows an example automaton derived from a FO-LTL constraint, which means that "stop using ACEI if Creatinine (CR) increases to >3.0 (mg/dL)".

Based on the deviations identified in every trace of a patient cohort, we can examine the deviations of the whole cohort. Because deviations occurring by chance alone are less meaningful, we should check whether each deviation is statistically significant for the cohort. Given a deviation δ in the stage s, let n be the number of patients who experienced s, and n_δ the number of patients for whom δ is identified, the support degree of δ is defined as $d_\delta = n_\delta / n$. Then we perform a one-tailed binomial test to check its statistical significance. Since different categories of deviations may have different variabilities, we define a local variance threshold for each category. Let θ be the threshold for the category of δ, and H_0: $d_\delta \leq \theta$ be the null hypothesis. If $d_\delta > \theta$ with p-value < 0.05, then δ is significant for the cohort and will be reported to clinicians.

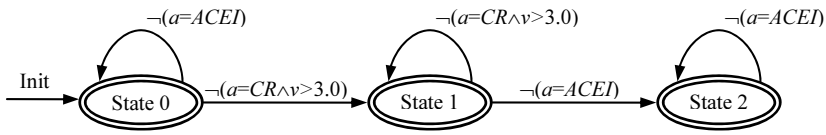

Figure 3. Büchi automaton for "$\forall a, v.\ \Upsilon((a=ACEI)\rightarrow\Box((a=CR\wedge v>3.0)\rightarrow\Box(\neg(a=ACEI))))$"

2. Results

In this section we present the details of our experiments on a multistage CP and real world EMR data. The CP is derived from a clinical guideline for the management of CHF [7], including 3 stages, 64 activities (24 diagnoses, 22 medications and 18 lab tests) and 86 constraints among the activities (e.g., the constraint shown in Figure 3). The EMR data contains 134,902 care events (79,966 diagnoses, 35,384 medications and 19,552 lab tests) that have occurred over the course of 4 years, from a cohort of 430 CHF patients with COPD (chronic obstructive pulmonary disease) condition. In the experiment, we set the variance thresholds for additional activities, absent activities and violated constraints to 0.2, 0.9 and 0.2 respectively. Then the three categories of deviations are identified accordingly. The statistics of the deviations is shown in Table 1. Notice that in Stage C, much more additional activities are identified, while pre-defined activities are seldom absent. That is probably because the disease condition of Stage C is more severe and complex, and more interventions were given.

Table 1. Statistical results of the variance analysis

Stage	#Activities	#Constraints	#Additional Activities	#Absent Activities	#Violated Constraints
Stage A	18	28	6	8	15
Stage B	22	28	4	6	14
Stage C	24	30	26	1	16

The analysis of the deviations reveals that, most of the additional activities are relevant to the target condition. For example, the top 10 additional activities (except COPD) in Stage C include: Glucocorticoids (GCs), Bronchodilators, Quinolones and Statins, which are useful medications for treating CHF with COPD; Anemias and Joint disorders, which are possible complications of COPD; and HbA1c, which is usually monitored when using GCs. These deviations are helpful in developing the CP specific to the management of CHF with COPD. Besides, a majority of the absent activities and violated constraints are due to the non-compliance to the pre-defined lab tests. Many indicators planned in the CP were not monitored on schedule. These non-compliance situations could be reported to clinicians to promote the safety and quality of care.

3. Discussion

Variance analysis forms a critical part of the quality improvement cycle for CPs [2]. Unfortunately, healthcare organizations rarely collect and analyze variance data when using CPs due to the difficulty in identifying deviations [2]. This paper proposed a novel approach to automate the process of CP variance analysis, particularly for CPs with multiple stages. Experimental results show that our method has the potential to provide meaningful information for refining CPs and improving care practices.

One limitation of our current approach is that we do not distinguish positive deviations from negative ones in terms of patient outcomes, which may provide additional evidence for further improvement. To address this problem, we are planning to develop an approach to rank the identified deviations according to their correlation with a particular outcome. Another limitation is that the time intervals between adjacent events are not modeled in HMM. The intervals, however, may vary between 0 day and several years, which should be taken into account during trace alignment. A potential improvement is continuous-time HMM, where the transition probabilities are defined as functions of time. Other future work includes run-time CP recommendation, which aims to provide decision support during CP execution, by suggesting new activities or alerting violations according to the results of variance analysis.

References

[1] Vanhaecht K, De Witte K, Sermeus W. The impact of clinical pathways on the organisation of care processes [dissertation]. Belgium: Katholieke Universiteit Leuven; 2007.
[2] Hyett KL, Podosky M, Santamaria N, Ham JC. Valuing variance: the importance of variance analysis in clinical pathways utilisation. Aust Health Rev. 2007 Nov; 31(4): 565-70.
[3] Ainsworth J, Buchan I. COCPIT: a tool for integrated care pathway variance analysis. Stud Health Technol Inform. 2012; 180: 995-9.
[4] de Leoni M, Maggi FM, van der Aalst WMP. Aligning event logs and declarative process models for conformance checking. In: Proceedings of the 10th international conference on Business Process Management; 2012 Sep 3-6; Tallinn, Estonia. Berlin: Springer; 2012. p. 82-97.
[5] Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE. 1989 Feb; 77(2): 257-86.
[6] De Masellis R, Su J. Runtime enforcement of first-order ltl properties on data-aware business processes. In: Proceedings of the 11th International Conference on Service Oriented Computing; 2013 Dec 2-5; Berlin, Germany. Berlin: Springer; 2013. p. 54-68.
[7] Hunt SA, Abraham WT, Chin MH, Feldman AM, Francis GS, Ganiats TG, et al. ACC/AHA 2005 guideline update for the diagnosis and management of chronic heart failure in the adult. Circulation. 2005 Sep 20; 112(12): 154-235.