Variance analysis in task-time matrix clinical pathways

Citation for published version (APA):

Yan, H., Van Gorp, P. M. E., Kaymak, U., Ji, L., Lu, X., Chiau, C. C., Korsten, H., & Duan, H. (2017). *Variance analysis in task-time matrix clinical pathways*. Paper presented at International Conference on Biomedical and Health Informatics (BHI 2017), 16-19 February 2017, Orlando, Florida, Orlando, Florida, United States.

Document status and date:

Published: 17/02/2017

Document Version:

Accepted manuscript including changes made at the peer-review stage

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Variance Analysis in Task-Time Matrix Clinical Pathways

Hui Yan¹, Pieter Van Gorp², Uzay Kaymak², Lei Ji³, Xudong Lu¹,
Choo Chiap Chiau⁴, Hendriks H.M. Korsten⁵ and Huilong Duan¹

Abstract—Clinical pathways are popular healthcare management tools to standardise care and ensure quality. The impacts and how physicians conform to the clinical pathway give feedbacks on improvement of the execution or redesign of the clinical pathways. This paper demonstrates using Business Process Model and Notation (BPMN) language to model Task-Time matrix clinical pathways. In order to do variance analysis, an optimal alignment (between a direct BPMN model and patient traces) searching algorithm is described. This new approach shortens the time checking deviations on such clinical pathways from hours (using the existing method) to minutes. A case study on variance analysis is undertaken, where a clinical pathway from the practice and a large set of patients data from an EMR database are used. The results demonstrate that automated variance analysis between Task-Time matrix clinical pathways and EMR data is feasible and provide meaningful insights for further improvement.

I. INTRODUCTION

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]. They have emerged since 1985 [2] and used worldwide nowadays. In China, the National Health and Family Planning Commission (NHFPC) has issued more than 400 clinical pathways [3]. In 2012, the NHFPC required every tertiary- and secondary-level hospital in China to implement at least 60 clinical pathways. By 2015, 5,924 public hospitals have implemented clinical pathways systems [4]. As Electronic Medical Record (EMR) systems are widely used in many countries, it enables storage of patients data, which captures the details of the patient's progression along or around the clinical pathways. By comparing the patients medical record against the clinical pathway, we can identify where the care given has deviated from the care expected. It is termed Care Pathway Variance Analysis [5].

Researchers in [6] conducted variance analysis based on the deviations recorded manually during daily ward rounds.

*This work was sponsored by Philips Research in the scope of Brain-Bridge Project.

¹Hui Yan, Xudong Lu and Huilong Duan are with the School of Biomedical Engineering and Instrumental Science, Zhejiang University, Hangzhou, P.R. China {h_yan, lvxd, duanhl}@zju.edu.cn

²Pieter Van Gorp and Uzay Kaymak are with the School of Industrial Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands {u.kaymak, p.m.e.v.gorp}@tue.nl

³Lei Ji is with the IT Department, Chinese PLA General Hospital, Beijing, P.R. China jilei_nudt@163.com

⁴Choo Chiap Chiau is with the Philips Research, Shanghai, P.R. China choo.chiap.chiau@philips.com

⁵Hendriks H.M. Korsten is with the Anaesthesiology and Intensive Care in Catharina Ziekenhuis, Eindhoven, The Netherlands erik.korsten@cze.nl

Study in [5] developed a web-based tool COCPIT to identify and analyze deviations automatically, which represent clinical pathways in directed graphs and compare historical data in EMR against the model. However, its matching algorithm depends on high-quality data to produce meaningful results. Study in [7] aligned patient traces with the CPs using hidden Markov models (conveying multiple stages of CP) and then identify deviations using first-order linear temporal logic. Apart from the inconvenience for users to use those mathematic languages, generating exhaustive linear temporal logics individually is error-prone and arduous.

Researchers in [8], [9] have proposed using a straightforward, easy-understandable and standard process modeling language BPMN to represent CPs. Representing CP in one compact process model saves the efforts generating separated rules exhaustively. Rich semantics of BPMN support complex expressions of clinical scenario or even nested statements. Temporal relations between care activities in CPs can be represented in one process model in BPMN. BPMN further provides Time Events representing upper limit or expected transition time. Validation conditions for specific activities could also be included in Gateways or annotations.

In this paper, we studied the real-life CP in daily practice in Chinese hospitals. NHFPC issued CPs in documents containing a Task-Time matrix specifying main care activities and their temporal relations. Hospitals localize the CPs and embed them in their EMR systems with the same format. Such a form contains several phases, each of which specifies a period with a set of care activities. Care activities have the properties such as optional/mandatory, long-term/short-term or even conditions. We will demonstrate modelling Task-Time matrix CP in BPMN model is feasible and an automatic deviation detection method will be presented.

Though our previous work [10] studied automated deviation detection between Event Logs and complex flow charts in BPMN, it does not apply to multiple activities (more than seven) without strict temporal orders due to performance problems. The reason is that study in [10] requires a statespace generated, which takes hours or days when more than ten activities are allowed without strict temporal relations. In this paper, we reuse the basic algorithm framework of [11] but checking deviations directly between the CP model in BPMN and patient traces without statespace generation. To evaluate this method, we did a case study using real-life data on unstable angina pectoris pathway in BPMN from a Chinese hospital. The results show that our variance analysis approach is feasible and it can provide meaningful insights for further improvement.

The remainder of the paper is organized as follows. Methods are presented in Section II. Section III describes the case study. Finally, Section IV concludes and discusses possible directions for future work.

II. METHODS

A. Modeling CP in BPMN and data preparation

A CP document contains care activities from various aspects, including medication orders, nursing orders, nutrition orders, lab tests orders, and some “reminders” for physicians. Only medical orders are recorded in EMR system. On the other hand, EMR system stores extra information as well, such as when lab test results return, which is not in the scope of CP. Thus, mapping terms in EMR database to those in the CP model is the first step.

With aligned terms of activities from CP document and EMR database, the next step is to identify key activities. Those key activities work as milestones of the process whilst connecting between phases. For example, “admission”, “surgery” and “discharge” are three obvious key activities in most clinical pathways. The third step is adding the rest activities between those milestones. There are several patterns to follow: if the temporal relations among a set of activities are arbitrary, put those activities in a Parallel/Inclusive construct pair (between a Parallel/Inclusive Split and Join Gateways); If several activities are exclusive, an Exclusive construct pair serves this goal. Optional/mandatory property can be specified on the incoming Sequences of the activities. Then connecting these patterns to the milestones forms one compact CP model in BPMN. Additionally, we interpret long-term order “give aspirin” in CP to one single activity meaning “start giving aspirin”.

From historical data’s perspective, medical orders are stored in EMR systems with time stamps. It is easy and sufficient to transform the actual care events into Event Logs [12]. Event Logs consists of Event Traces, which lays Events out chronologically. For long-term medicine orders, we further extract the Events into *Start*, *Increase Dose Of*, *Decrease Dose Of* and *Stop* categories.

B. Parsing BPMN Models

Aligning an Event Trace to a CP model is indeed aligning it to one of paths in the model. In optimal alignment search algorithm, it is required that traces of activities in the model are ready. Since original .bpmn file provides graphical information of each elements (Tasks, Gateways, Subprocess and so on), we can parse a BPMN model into a simple flat model. For Subprocess, we can easily remove the hierarchy by linking entrance and exit of a Subprocess to its parent level. Since this paper deals with Task-Time matrix CP, Intermediate Events, Compensations and other advanced BPMN constructs are not necessarily in our scope. For Gateways, we specially introduce Gateway Pair Construct G_{pair} , which combine Split and Join Gateways and the elements in between into one:

Definition I A G_{pair} is a tuple $(Type, Incoming, InOutgoing, ExOutgoing)$, where:

- 1) *Type* indicates the type of the Gateways: Exclusive, Inclusive or Parallel.
- 2) *Incoming* refers to the incoming node of G_{pair} , namely the incoming node of the Split Gateway.
- 3) *InOutgoing* refers to a set of Tasks between the Split and Join Gateway constructs.
- 4) *ExOutgoing* refers to the outgoing node of the Join Gateway.

Tasks, Start Event, End Event are connected using Sequence Flows in BPMN. A Gateway Pair Construct is connected through its *Incoming* and *ExOutgoing* to its outside and *InOutgoing* to its inside Tasks, which makes it flexible to link in or out of the G_{pair} when needed in optimal alignments search algorithm.

C. Aligning Event Logs to CP Models in BPMN

Optimal alignments can be created heuristically using an A* based algorithm [11]. It works as follows: starting from an initial (empty) alignment, take the first Event x from an Event Trace and take the first Task y in the CP model. If x matches with y , a new alignment is generated $\gamma_1 = \{(x, y)\}$; if x does not match y , two new alignments are generated $\gamma_2 = \{(x, \perp)\}$ and $\gamma_3 = \{(\perp, y)\}$ (“ $\perp$ ” is used to indicate a placeholder). $(x, \perp)$ means Event x is not matched, so it is a deviation as added. Similarly, $(\perp, y)$ means Task y is not matched, so y is a missing Task in the model. The one with the best potential being the optimal is selected to extend again.

In the next iteration of generating alignments, successors from both sides are needed. In an Event Trace, successor of Event x can be easily obtained chronologically, i.e. x' . In model’s perspective, successor of a Task y can be obtained through connections in the parsed flat model. In case that the successive node of Task y is a Gateway Pair Construct g_{pair} and it is of *Type* Parallel, all the Tasks in $g_{pair}.InOutgoing$ are compared to x' . If one of those Tasks y'_1 is matched to x' , the newly generated alignment is $\gamma_4 = \{(x, y), (x', y'_1)\}$; If none of the Tasks in $g_{pair}.InOutgoing$ is matched to x' , two alignments are generated: $\gamma_5 = \{(x, y), (x', \perp)\}$ and $\gamma_6 = \{(x, y), (\perp, y'_1), (\perp, y'_2), (\perp, y'_3) \dots\}$. The *Type* of g_{pair} regulates exclusive or coexisting logics between $(\perp, y'_1)$, $(\perp, y'_2)$ and $(\perp, y'_3)$. The algorithm stops when it reaches both the end of the Event Trace and the model.

The key to the success of creating and finding the optimal alignment is the evaluation function. The evaluation function $f(s)$ was defined as $f(s) = g(s) + \alpha * (h_L(s) + h_M(s))$ in [11], where $g(s)$ is the number of deviations. $h_L(s)$ is number of remaining Events in the Trace. $h_M(s)$ is the estimated remaining effort (layer in statespace) from the model’s perspective. α is the weight factor to adjust weights of deviations $g(s)$ and heuristic information for estimation.

Though the basic algorithm framework keeps the same with [11], replacing statespace into a BPMN flat model requires adjustment of $h_M(s)$ and the corresponding α . In this paper, we use the number of remaining Tasks in the model to indicate $h_M(s)$. The weight factor α is usually of a positive value but smaller than 1. It is a tradeoff

Fig. 1. Unstable Angina Clinical Pathway Model in BPMN 2.0

between efficiency and accuracy. The smaller α is, the more accurate the optimal alignment is. However, a smaller α takes longer time to complete search. We found the value of α being 0.05 useful in our case study through trial and error. Via this algorithm, optimal alignments can be discovered automatically. Thus where exactly logs deviate from models, and where they are consistent again are shown.

III. EXPERIMENT

A. Modeling Unstable Angina Pectoris Clinical Pathway in BPMN and Data Preprocessing

Our cooperation hospital has adopted clinical pathway of interventional treatment of unstable angina pectoris (version of 2009)[13], which was published by the NHFPC of China. The adopted CP is embedded in its EMR system in the form of Task-Time matrix. The localized CP has six sequential Phases. Care activities in other words, Tasks in the model are listed in each Phase. In those Tasks, we identify three main key Tasks: Admit Patient, Perform a Surgery, Discharge Patient. After “Admit Patient”, an ECG is performed to see how severe the patient is. Then a set of routine medicine and lab tests are given to the patient. They could be long-term or short-term orders. Before surgery, a set of Tasks are required to be done. After surgery, anti-coagulants should be given to prevent bleeding. Besides, ECG monitoring should be performed. The modeled Pathway is shown in Fig. 1. Note that after “Perform ECG”, two Subprocesses are followed, each of which has seven Tasks in a Parallel Gateway Pair. In the flat model, this means that those 14 Tasks are in arbitrary order.

We extract anonymous patients from historical database with the final diagnosis “Unstable Angina Pectoris” and interventional Treatment into data tables. EMR data contains medical orders, examination results, lab test results, vital signs. Medical orders are transformed into Event Logs. Long term medical orders are parsed into several Events representing start, end, dose increased or dose decreased separately. Some medical orders were missing, for example, Give Anesthetics was missing in some cases but surgery was performed. Some test orders appeared as a set of test items, which had the exact same timestamp. We use one main test order to substitute the whole set. Since this CP is a standard treatment of unstable angina pectoris patients with one interventional surgery, patients with length of stay longer than 12 days are filtered out. In total, 1051 patients out of 1542 are included in the case study.

B. Results

Fig. 2. Deviation Ratio vs. Patients Number

Running the algorithm with α of 0.05 takes 557 seconds. When α is of 0.03, it takes 812 seconds with the same results produced. However, using the method in [10] takes more than 10 hours due to generating statespace for 14 Tasks after a Parallel Split Gateway. Fig. 2 shows distribution of patients by deviation ratio. Deviation ratio is defined as number of deviations divided by length of the optimal alignment. What is worth mentioning is that the optimal alignments with minimum deviations are several. They differ at combinations of deviations (but with same number). For simplicity, we select the first one generated by the algorithm into a deviations data set.

Here we discuss two most frequent deviations.

Perform ECG in the CP is missing 767 times. We further find two-element frequent pattern (using FP-growth algorithm) containing *Perform ECG* are: *Perform ECG* is missing whilst *Perform ECG* is added between two *Give Routine Medications* for 133 times; *Perform ECG* is missing whilst *Perform ECG Monitoring* is added between two *Give Routine Medications* for 63 times; *Perform ECG* is missing whilst *Perform ECG Monitoring* is added before *Perform a Surgery* for 76 times. This phenomena shows that though the CP regulates that *Perform ECG* is before *Give Routine Medication*, 767 patients out of 1051 did not follow it. However, it occurs between *Give Routine Medications* and also before *Perform a Surgery*. This raises a natural question to physicians, whether the CP should be redesigned if those deviations are reasonable. The whole process of variance analysis is automated.

Give Drugs Relating Myocardial Ischemia Protection is added between *Give Routine Medications* for 541 times. *Drugs Relating Myocardial Ischemia Protection* is a collection of several drugs, including cinepazide maleate injection, alprostadil injection and others. These drugs are not specified in the CP, so they are categorized into a collection based on its effect in data preprocessing. Again, from the relating frequent patterns, we find that *Give ACEI/ARB inhibitors* is missing for 279 times, *Give Anticoagulants* is missing for 188 times, *Give Beta Blocker* is missing for 188 times, *Give High-intensity Statin* is missing for 72 times. It seems that these *Drugs Relating Myocardial Ischemia Protection* are substituting the traditional drugs used in cardiovascular disease. This should be confirmed and discussed with the experts. Moreover, counting detailed number of each drug in drugs relating myocardial ischemia protection helps in variance analysis.

IV. DISCUSSION AND CONCLUSION

This paper provides a method checking deviations between CPs in Task-Time matrix format and patients' Event Logs from EMR database. Identified deviations help understand existing situations, find hidden problems of the existing CP models and suggests possible redesign ideas. One limitation of this research is that EMR data does not represent the real execution or performing of the care. Physicians may perform some tasks and then register them in the EMR system, especially in emergencies. Though this paper uses

BPMN language to model CPs, it does not cover a complete set of BPMN constructs. Advanced constructs such as loops, Intermediate Events, Compensations in BPMN are not supported in this paper. This method does support multiple Tasks in parallel without compromising efficiency.

The foundation of optimal alignment search technique is a global optimization problem. Due to this, adding one fake deviation (due to wrong registration time) could make the original deviation invisible. Also, when the deviation ratio rises to a high value, the identified deviations are not reliable. Distinguishing different deviations by assigning different value in the evaluation function is one possible solution. Even though deviations do not necessarily lead to negative effects and some deviations may even save the patient's life, some negative deviations can be avoided if the clinical practitioners are aware of the corresponding outcomes and avoid the deviation timely. Future research direction applies using data mining techniques to understand deviations and their effects, so that clinical pathways can be improved.

REFERENCES

- [1] K. Vanhaecht, K. De Witte, and W. Sermeus, "The impact of clinical pathways on the organisation of care processes," 2007.
- [2] L. J. Carpenito-Moyet, *Nursing care plans & documentation: nursing diagnoses and collaborative problems*. Lippincott Williams & Wilkins, 2009.
- [3] X. Y. He, M. K. Bundorf, J. J. Gu, P. Zhou, and D. Xue, "Compliance with clinical pathways for inpatient care in chinese public hospitals," *BMC health services research*, vol. 15, no. 1, p. 1, 2015.
- [4] J. A. He and W. Yang, "Clinical pathways in china—an evaluation," *International journal of health care quality assurance*, vol. 28, no. 4, pp. 394–411, 2015.
- [5] J. Ainsworth and I. Buchan, "Cocpit: a tool for integrated care pathway variance analysis," *Studies in health technology and informatics*, vol. 180, pp. 995–999, 2012.
- [6] P. Dalton, D. Macintosh, and B. Pearson, "Variance analysis in clinical pathways for total hip and knee joint arthroplasty," *Journal of quality in clinical practice*, vol. 20, no. 4, pp. 145–149, 2000.
- [7] X. Li, H. Liu, S. Zhang, J. Mei, G. Xie, Y. Yu, J. Li, and G. T. Lakshmanan, "Automatic variance analysis of multistage care pathways," *Studies in health technology and informatics*, vol. 205, pp. 715–719, 2013.
- [8] R. Muller and A. Rogge-Solti, "BPMN for healthcare processes," *Services und ihre Komposition Proceedings of the 3rd Central-European workshop on Services and their Composition*, vol. 705, pp. 65–72, 2011.
- [9] H. Scheuerlein, F. Rauchfuss, Y. Dittmar, R. Molle, T. Lehmann, N. Pienkos, and U. Settmacher, "New methods for clinical pathways-Business Process Modeling Notation (BPMN) and Tangible Business Process Modeling (t.BPM)," *Langenbeck's archives of surgery / Deutsche Gesellschaft für Chirurgie*, vol. 397, no. 5, pp. 755–61, June 2012.
- [10] H. Yan, P. Van Gorp, U. Kaymak, X. Lu, R. Vdovjak, H. H. Korsten, and H. Duan, "Analyzing conformance to clinical protocols involving advanced synchronizations," in *2013 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2013, pp. 61–68.
- [11] H. Yan, P. van Gorp, X. Lu, C. Chiau, U. Kaymak, and H. Duan, "Deviations discovery using aligning event logs to business processes using a*," Technical Report, 2016. [Online]. Available: <http://onderzoekschool-beta.nl/working-papers/deviations-discovery-using-aligning-event-logs-business-processes-using/>
- [12] W. M. P. van der Aalst, *Process Mining - Discovery, Conformance and Enhancement of Business Processes*. Springer, 2011.
- [13] Y. Bi, R. Gao, A. Patel, S. Su, W. Gao, D. Hu, D. Huang, L. Kong, W. Qi, Y. Wu, et al., "Evidence-based medication use among chinese patients with acute coronary syndromes at the time of hospital discharge and 1 year after hospitalization: results from the clinical pathways for acute coronary syndromes in china (cpacs) study," *American heart journal*, vol. 157, no. 3, pp. 509–516, 2009.