

Data-Driven Modeling of Complex Business Process in Heterogeneous Environment of Healthcare Organization with Health Information Systems

Alexander KSHENIN^{a,1} and Sergey KOVALCHUK^{a,b}

^aITMO University, St. Petersburg, Russia

^bAlmazov National Medical Research Centre, St. Petersburg, Russia

Abstract. Business process modeling aims to construct digital representations of processes being executed in the company. However, models derived from the event logs of their execution tend to overcomplicate the desired representation, making them difficult to apply. The most accurate recovery of the business process model requires a comprehensive study of the various artifacts stored in the company's information system. This paper, however, aims to explore the possibility to automatically obtain the most accurate model of business process, using mutual optimization of models recovered from a set of event logs. Further, the obtained models are executed in multi-agent simulation model of company, and the resulting event logs are examined to determine patterns that are specific to distinct employees and those that generally characterize business process.

Keywords. Business process modelling, model recovery, agent-based simulation modeling, pattern identification, health information system

1. Introduction

Business Process (BP) modeling aims to construct digital representations of processes being executed in the company, that are required for the implementation of Business Process Management (BPM) in the organization and its sustainable development in constantly changing demands of the competitive environment (1).

An extensive research and optimization of processes in a medical organization make it possible to efficiently allocate and use available resources and provide treatment with no delays and at the proper volume (2). Moreover, the construction, research, and simulation of the BPs is performed when implementing the digital twin of the organization used in decision support systems (DSS) for scenario analysis (3). The health information system (HIS) of a medical organization stores large volumes of heterogeneous information about its functioning, which can be used to build and explore the BP of the healthcare system (4).

In this paper, authors consider the possibility of obtaining the most accurate model of BP in application to healthcare, using mutual optimization of models recovered from

¹ Corresponding Author, ITMO University, 197101, St. Petersburg, Russia; E-mail: Ksh.Al.Dm@gmail.com.

a set of event logs, as well as information about complex BP by examining structural and temporal patterns of BP simulation logs. The complexity of BP in this research means its diversity of execution, observed in uneven limited event logs.

Heterogeneous information about the BPs of a large medical center is used to show the proposed method. Obtainment and investigation of a set of BP implementations are performed using multi-agent simulation model developed by the authors.

2. Related Works

When the BP model is defined by an expert, the resulting representation may not reflect the actual execution of BP by employees. Models derived from the execution event logs may show excessive detail of actions, as well as they may vary from one employee executing a BP to another (5).

Different Process Mining algorithms aim to solve this problem for different types of event logs, assumptions about the investigated BPs (6). For example authors of the article (7) implement filtering of logs from rare actions that noise up the event log using their relative frequencies. Furthermore, heuristics are used to identify noise in logs. In (8) the researchers propose an approach that develops a systematic methodology for clearing event logs. Based on several datasets, 11 log imperfections patterns were identified and appropriate ways to detect and eliminate them were suggested. However, the goal of current research lies in automatic filtering of BP logs, with main assumption contained in splitting event logs on specific feature, and construction of an accurate BP model.

Recurring patterns can be found in the BP model, obtained from good quality event log without noise, they make the understanding of the processes easier or can be used to build new processes. In paper (9) the authors propose their approach for automatic pattern discovery using specified ontology from the event log generated during the execution of BPs. The sequence of events and semantics are used to determine if a pattern is present in a process. Authors of the article (10) describe an approach to identify process execution priority patterns, using the event log, to discover the impact on the overall process. Also patterns can be used to compare the performance of BP as shown in (11), where authors compare the elements of the graph built on event logs to evaluate the divergence of two logs.

3. Data

Data used in this research was provided by an Almazov National Medical Research Centre (Almazov Centre). It includes information about the movement of medical staff, derived from the Access Control System (ACS), and various actions with Electronic Health Records (EHR) from the HIS. ACS and HIS data were combined and used to extract the BP event logs for a set of employees. This study uses event logs of 24-hour nurses' shifts (log cases), generally described in table 1.

Table 1. Description of nurses' event logs

Employee	Shifts	Activities	Unique activities
№0	137	6118	239
№1	187	13644	286
№2	127	3595	239
№3	113	6427	246

Event logs belong to four nurses working in one department, with similar or close responsibilities, implementing a similar set of actions, allowing to identify and explore the model of BP. Thus, the event log of shifts contains nurse's activities – actions with HIS and movement through a set of locations in the center, with timestamps.

4. Methods

The ProFIT library (<https://github.com/Siella/ProFIT>), already applied to mining BPs in healthcare (4), was used in this research to obtain a set of BP models from the logs. Each BP model is obtained in form of a weighted graph, described by a weighted transition matrix. Graph nodes represent activities from the log, edges – sequential transitions between them. The weight of an edge is the frequency of transition occurrence in logs.

Authors propose the identification of the most accurate model of BP which includes:

- The construction of executable BP models $\{M_i\}_{i=1}^n$ for a set of n event logs.
- Mutual optimization of models – obtainment of a set of filtering parameters $\{f_i\}_{i=1}^n$ for activities and transitions (with preassigned bounds). Let $M_i(f_i)$ be a BP model, obtained from event logs i with activities' filtering f_i . $N_{i,f_i} = N(M_i(f_i))$ – is the corresponding set of activities, then

$$\{f_i\}_{i=1}^n: \frac{1}{n} \sum_i \frac{N_{i,f_i} \setminus \cup_{j \neq i} N_{j,f_j}}{N_{i,f_i}} \rightarrow \min \tag{1}$$

The set of event logs can be derived by splitting general event log on specific feature, for example, employee – as presented in this research. Thus, proposed frequency filtering automatically remove nodes and edges, which occur very rarely in event logs and turn the result into “spaghetti-like” models, and consider information about other models in every BP model, making them more accurate.

To construct an executable BP model, graph is supplemented with transition probabilities and distributions of execution times, and the process of BP occurrence. Models are executed in multi-agent BP simulation model, developed by the authors to produce simulation logs used in pattern identification. Each pattern is a sequence of activities, that appear in several event logs with specified significance threshold.

5. Results

A set of four executable BP models of nurses' shifts were selected, obtained, optimized. For each nurse, BP simulation modeling was performed, resulting in a set of event logs of BP model implementation. Further, patterns of length 2 – 5 were identified for each simulation log. The analysis of patterns showed that nurses №0 and №3 have the closest BP models, followed by nurse №1 and then №2. The filtering parameters obtained with the optimization of BP model on the subset of original event logs of two employees (№0 and №3) are shown in table 2.

Table 2. Filtering parameters for actions and transitions of two BP models optimization

Employee	f_i of activities	f_i of transitions
№0	82	5
№3	77	

The visualization of optimized BP model, that corresponds to employee №0 with filtering parameters from table 2, is shown on Figure 1.

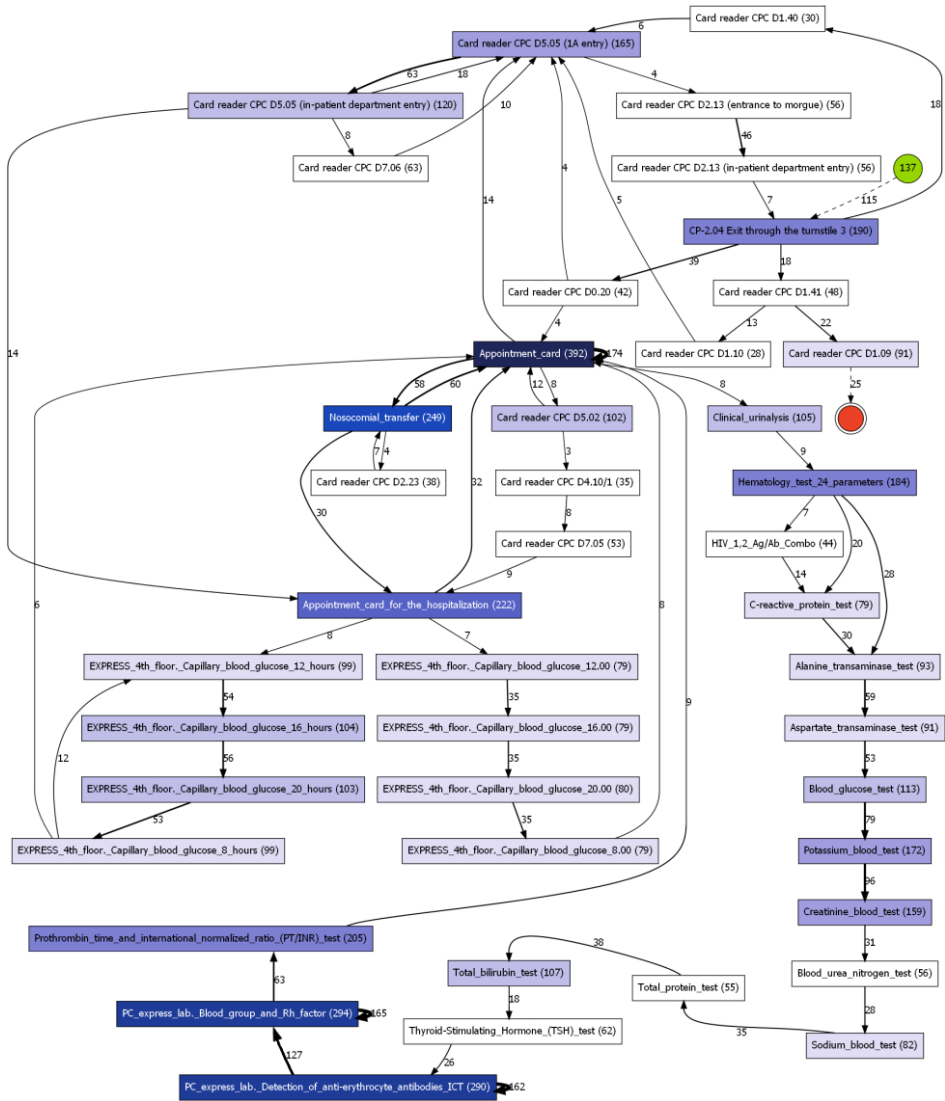


Figure 1. Visualization of optimized graph BP model corresponding to nurse №0.

Figure 2 provides the box plots of four movement patterns, that are mutual for all BP models, and consequently, employees. These patterns are situated at the very beginning of BP and may show small time loss in employee’s №1 execution (left), and higher dispersion of longer patterns’ execution times (right) that is due to untracked sub-pathways and different arriving time. The example of movement patterns, №7, is the sequence: 'CP-2.04 Exit through the turnstile 3', 'Card reader CPC D0.20', 'Card reader CPC D5.05 (1A entry)', 'Card reader CPC D5.05 (in-patient department entry)'.

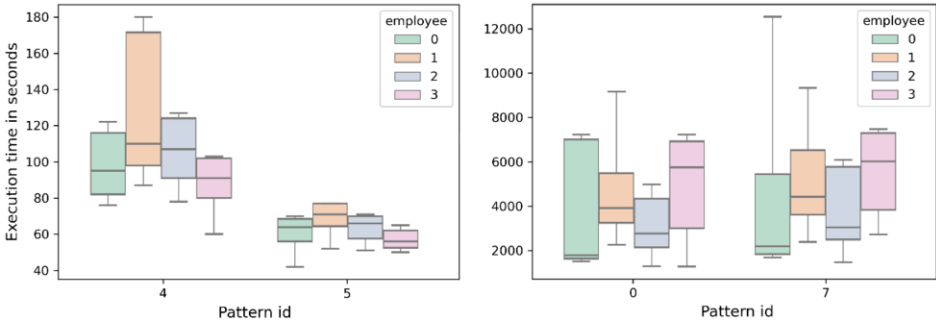


Figure 2. (left) length of two movement patterns; (right) length of four movement patterns.

Figure 3 provides the box plots of patterns related to various blood analysis tests. Though all three employees possess such competencies, employee №1 performs them in a different scenario not during BP closest to other employees’ models. Employees №0 and №3 show similar results in executing these patterns, as they are subsequences of clinical blood analysis that is highly automated. The example of tests patterns, №13, is the sequence: 'Thyroid-Stimulating Hormone (TSH) test', 'PC express lab. Detection of anti-erythrocyte antibodies ICT', 'PC express lab. Blood group and Rh factor'.

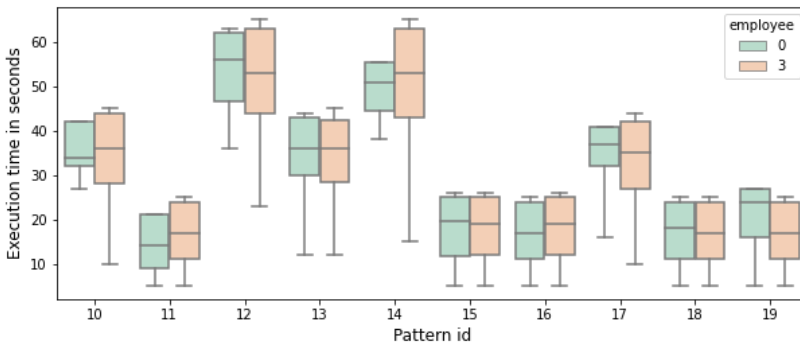


Figure 3. Blood analysis patterns.

6. Discussion

BP, explored in this research does not possess parallel execution of several activities, and moreover, is entirely executed by a single employee. The opposite case of the second feature, nonetheless, can be considered within the proposed method. In this case, the resulting activity patterns will correspond to all possible subsets of employees, that are involved in executing them. One can analyze the time and structure diversity in such patterns and identify activities and/or employees where the loss of efficiency occurs.

Execution patterns in simulation logs are rather simple, however, they reflect the process of execution – as a sequence of activities, so they are easy to interpret and analyze. Box plots comparison can show both the personalized loss of execution time and the efficiency of automation. When changing subsets of models one can identify structural differences between BP executed by different employees.

The results can be improved with a more representative variable for graph model optimization. The procedure, that can handle possible changes in models' subset, will allow to dynamically explore more diverse models and save time, not starting from the very beginning. The most resource-demanding part of the research method is simulation modelling; however, all BP models can be executed in parallel, vastly reducing time.

7. Conclusion

In this research, authors propose method to obtain the most accurate model of BP, using mutual optimization of models, derived from a set of event logs. The identification of BP model is followed by the examination of structural and temporal patterns in BP simulation logs. The study includes the identification and research of nurses' shifts BP, using the data of Almazov Centre, that show an interpretable result.

Acknowledgments

This research is financially supported by The Russian Scientific Foundation, Agreement #17-15-01177.

References

- [1] Shaw DR, Holland CP, Kawalek P, Snowdon B, Warboys B. Elements of a business process management system: Theory and practice. *Bus Process Manag J.* 2007;13(1).
- [2] Rojas E, Munoz-Gama J, Sepúlveda M, Capurro D. Process mining in healthcare: A literature review. *J Biomed Inform [Internet].* 2016 Jun;61:224–36. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1532046416300296>
- [3] Sulis E, Tema P, Di Leva A, Boella G, Boccuzzi A. Agent-oriented Decision Support System for Business Processes Management with Genetic Algorithm Optimization: an Application in Healthcare. *J Med Syst.* 2020 Aug 2;44(9):157. doi: 10.1007/s10916-020-01608-4.
- [4] Elkhovskaya L, Kabyshev M, Funkner A, Balakhontceva M, Fonin V, Kovalchuk S. Personalized assistance for patients with chronic diseases through multi-level distributed healthcare process assessment. *Stud Health Technol Inform.* 2019;261:309-312.
- [5] van der Aalst W. Process mining: discovering and improving Spaghetti and Lasagna processes. 2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM) 2012. DOI: 10.1109/CIDM.2011.6129461
- [6] Van Dongen BF, Alves De Medeiros AK, Wen L. Process mining: Overview and outlook of Petri net discovery algorithms. *Lecture Notes in Computer Science* 2009; 5460: 225-242.
- [7] Conforti R, La Rosa M, Ter Hofstede AHM. Filtering Out Infrequent Behavior from Business Process Event Logs. *IEEE Trans Knowl Data Eng.* 2017;29(2).
- [8] Suriadi S, Andrews R, ter Hofstede AHM, Wynn MT. Event log imperfection patterns for process mining: Towards a systematic approach to cleaning event logs. *Inf Syst.* 2017;64:132-150.
- [9] Ferreira DR, Thom LH. A semantic approach to the discovery of workflow activity patterns in event logs. In: *International Journal of Business Process Integration and Management.* 2012; 6(1). <https://doi.org/10.1504/IJBPI.2012.047909>
- [10] Suriadi S, Wynn MT, Xu J, van der Aalst WMP, ter Hofstede AHM. Discovering work prioritisation patterns from event logs. *Decis Support Syst.* 2017;100.
- [11] van Beest NRTP, Dumas M, García-Bañuelos L, La Rosa M. Log delta analysis: Interpretable differencing of business process event logs. *Lecture Notes in Computer Science* 2015;9253:386-405. DOI: 10.1007/978-3-319-23063-4_26