Process Mining Techniques: an Application to Stroke Care

Ronny MANS[a,1] , Helen SCHONENBERG[a], Giorgio LEONARDI[b], Silvia PANZARASA[b], Anna CAVALLINI[c], Silvana QUAGLINI[b], and Wil van der AALST[a]

[a] *Eindhoven University of Technology, The Netherlands*

[b] *Dept. Computer and System science, University of Pavia, Italy*

[c] *IRCCS Casimiro Mondino Foundation, Pavia, Italy*

Abstract. In a competitive health-care market, hospitals have to focus on ways to streamline their processes in order to deliver high quality care while at the same time reducing costs. To accomplish this goal, hospital managers need a thorough understanding of the actual processes. Diffusion of Information and Communication Technology tools within hospitals, such as electronic clinical charts, computerized guidelines and, more generally, decision support systems, make huge collections of data available, not only for data analysis, but also for process analysis. Process mining can be used to extract process related information (e.g., process models) from data, i.e., process mining describes a family of *a-posteriori* analysis techniques exploiting the information recorded in the event logs. This process information can be used to understand and redesign processes to become efficient high quality processes. In this paper, we apply process mining on two datasets for stroke patients and present the most interesting results. Above all, the paper demonstrates the applicability of process mining in the health-care domain.

Keywords. Data analysis-extraction tools, Process, Event-based systems

Introduction

Nowadays, Health-Care Organisations (HCOs) place strong emphasis on medical and organisational efficiency and effectiveness, to control their health-care performance and expenditures. The aim of the HCOs is to provide the highest quality services at the lowest cost [1]. Consequently, it is of the utmost importance to evaluate existing infrastructures, and the services offered by these organisations. Therefore, it is crucial to explore and process the data collected by HCO systems. These data can be process logs from a process management system, or databases from the electronic clinical chart system, or unstructured data. In fact, in modern day organisations, information and communication technologies are becoming pervasive and there is an immense growth of their use. Contemporary Information Systems (IS) have no existence of their own, but they act in the context of an organisation and its business processes [2]. Such

[1] Corresponding author: Ronny Mans, address: Paviljoen J3, 5600 MB, Eindhoven. E-mail: r.s.mans@tue.nl

systems, driven by process models to enact and manage operational business processes, are referred to as *Business Process Management Systems* (BPMSs) [3]. BPMSs record the data of the executed activities in form of event logs. An event log is like a history of what happened in the information system. This recorded data can be helpful to gain a clear picture of the underlying business process.

BPMS now also focus on Business Process Analysis (BPA), which covers functions of diagnosis and simulation of business processes. One of its emerging areas is Business Activity Monitoring (BAM), which typically focuses on performance issues without considering causal and dynamic dependencies in processes and organisations. This is where process mining techniques can be employed in order to extract process related knowledge (e.g. process models) from event logs [4]. Process mining has been applied in many domains, mainly the service industry, but also in the health-care domain [5,6,11]. In this paper, we *applied process mining to discover the procedures for treating stroke patients in different hospitals*. Additionally, *we analysed the patient related events from stroke onset till arrival in the hospital.*

The remainder of this paper is organised as follows. An overview of process mining is presented in Section 1. Section 2 presents the data used for our analyses. In Section 3 we present and discuss the obtained results. Conclusions and future work are presented in Section 4.

1. Process Mining and the PROM tool

Process mining addresses the problem that most organisations have limited information about what is actually happening. In practice, there is often a significant gap between what is prescribed or supposed to happen, and what *actually* happens. Only a concise assessment of reality, which process mining strives to deliver, can help in verifying process models, and ultimately be used in system or process redesign efforts. The goal of process mining is to extract process related information (e.g., process models) from process logs, i.e., process mining describes a family of *a-posteriori* analysis techniques exploiting the information recorded in the event logs. Typically, these approaches assume that it is possible to sequentially record events such that each event refers to an activity (i.e., a well defined step in the process) and is related to a particular case (i.e., a process instance). Furthermore, some mining techniques use additional information such as the performer or originator of the event (i.e., the person/resource executing or initiating the activity), the timestamp of the event, or data elements recorded with the event (e.g., the size of an order).

Process mining is applicable to a wide range of systems. These systems may be pure information systems (e.g., ERP systems) or systems where the hardware plays a more prominent role (e.g., embedded systems). The only requirement is that the system produces *event logs*, thus recording (parts of) the actual behaviour. Usually, hospitals have a wide variety of systems that record events. For example, in an intensive care unit, a system can record which examinations or treatments a patient undergoes and also it can record occurring complications for a patient.

Traditionally, process mining has been focusing on *discovery*, i.e., deriving process models, organisational context, and execution properties from enactment logs. It is important to mention that there is no a-priori model, but, based on an event log, some model, e.g. a Petri net, is constructed. However, process mining is not limited to process models (i.e., control flow) and recent process mining techniques are more and more focusing on other perspectives, e.g., the organisational perspective, performance perspective or the data perspective. For example, there are approaches to extract social networks from event logs and analyze them using social network analysis [7]. This allows organisations to monitor how people, groups, or software/system components are working together. Also, there are approaches to visualize performance related information, e.g. approaches that graphically show the bottlenecks and all kinds of performance indicators, e.g., average/variance of the total flow time or the time spent between two activities.

To be able to understand whether the HCOs under study achieve their goals of providing timely and high quality medical services, we conducted several experiments using the Process Mining tool called ProM, extensively described in [8]. ProM is a platform independent open source framework which supports a wide variety of process mining and data mining techniques, and can be extended by adding new functionalities in the form of plug-ins. The ProM framework can read event logs that are in the MXML format. With the ProM *import tool,* data from various systems, such as MS Access, can be converted into the MXML format.

2. The data used for process mining

Two data sets are available. One refers to the clinical course of stroke patients from their hospital admission to discharge (clinical data set), and the other one refers to the pre-hospital phase (pre-hospital behaviour data set).

2.1. Data set 1: Clinical data set

Data on 368 consecutive subjects with a confirmed diagnosis of first-ever ischemic stroke have been collected using an electronic clinical chart, developed with the relational database management system MS-Access and shared by the neurological departments or the stroke units of four Italian hospitals. Data, that comprehend diagnostic and treatment procedures, complications, etc., are labelled according to the time elapsed from the symptoms onset: the first 6 hours are the *acute stroke phase (*110 patients arrived in the hospital in this phase), and the subsequent period, up to 7 days after stroke, was considered as the *sub-acute phase*. In the past, this data set has been analysed to investigate the relationships between compliance to clinical practice guidelines and stroke outcomes, in terms of both survival and cost [9,10]. However, those analyses have been carried out through classical statistical procedures, while *this paper shows that kind of new insights can be achieved using process mining techniques*. Although it is mostly a "classical" medical database, some process information can be derived, due to the availability of timestamps for diagnostic, therapeutic and clinical events, e.g. hospital admission and discharge. More specifically, these timestamps represent the actual dates of when the associated activities occurred.

2.2. Data set 2: Pre-hospital behaviour data set

In the pre-hospital behaviour data set, collected through direct interviews with 234 patients, we find all events which occur from stroke onset till the patient's arrival in the hospital. The data set contains temporal information about the actions taken by patients, their relatives and general practitioners (GP). More specifically, as detailed temporal data exists for each event, this offers the opportunity for discovering a model showing several performance indicators, like bottlenecks or time spent in between events.

3. Results

Due to the lack of space, we only show the most significant results of our analysis that highlight the potential of process mining.

Clinical data set results - Process mining can be used to construct process models for a whole data set, or parts that are of particular interest. One result that we would like to highlight in this paper is obtained by partitioning the data set on different hospitals for patients that arrived in the acute phase. In Figure 1, we only show the treatment process for hospital 1 and 2. The models are obtained by using the Heuristics miner and only show the main flow (relationships between events) of the process and only for the frequently occurring events. Events are depicted by boxes. The numbers in the activity boxes indicate the occurrence frequency of the activity, e.g. *admission* occurs 31 times in the log of hospital 2. The upper number on the arcs indicates the reliability of the relation between the activities, e.g. for hospital 1 the reliability of *admission* followed by *neuroprotection* is 0.917. The lower number on the arcs represents the number of times this activity pattern occurred in the log. The reliability of a relationship (e.g. event *i* followed by event *j*) is not only influenced by the number of occurrences of this pattern in the log, but is also (negatively) determined by the number of occurrences of the opposite pattern (*j* followed by *i*).

By comparing the obtained process models we observed a difference in treatment strategies between different hospitals. Most notably, hospital 2 performs hypertension therapy earlier and much more than the other hospitals. It is known that antihypertensive treatment is a common practice, although not always justified by

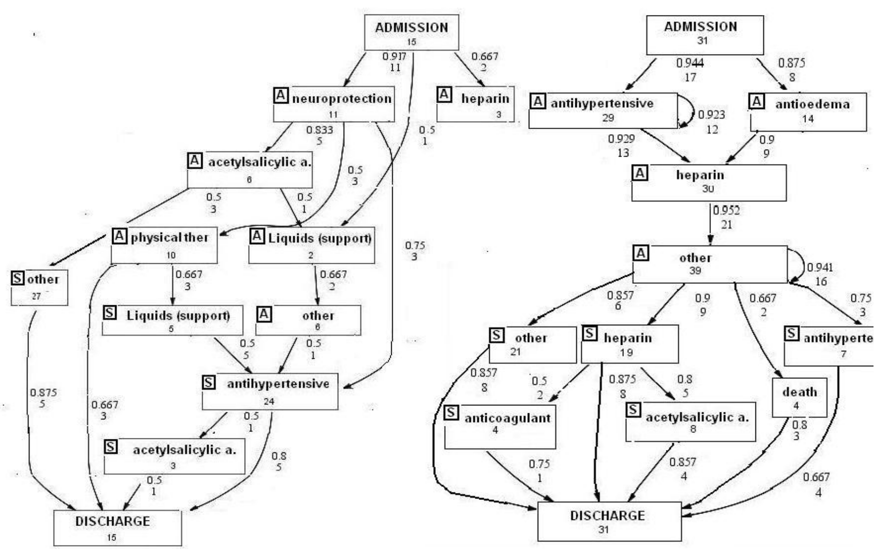

Figure 1. The treatment processes mined for the two hospitals H1(left) and H2(right). With respect to the original output, to facilitate the reader, labels "A" and "S" have been added indicating events belonging to the acute and subacute phase respectively.

scientific evidence. Hospital 1 seems to be more "research-addressed", since it adopts therapeutic protocols such as neuroprotection, and also is more compliant with the more recent guidelines, that recommend early physical therapy. Physicians can benefit from these results and look for motivations behind these differences.

Pre-hospital behaviour data set results - Pre-hospital data are useful to discover the population attitude versus stroke, motivations for possible delays and to indicate efficient pathways to reach the hospital. The question is interesting because stroke is, as a medical emergency, similar to a heart attack. But, opposite to heart attack, stroke is not well-known within the population, and signs and symptoms are often under-estimated by patients, their relatives and even GPs. As a result, often patients arrive in the hospital when their temporal window for effective treatments is over. In the pre-hospital behaviour data set we find detailed temporal data for each event, which offers the opportunity for discovering a model showing several performance indicators, like bottlenecks or time spent in between events.

Figure 2 shows a performance analysis plug-in of ProM which projects timing information on places and transitions in a Petri net. It graphically shows, for a part of

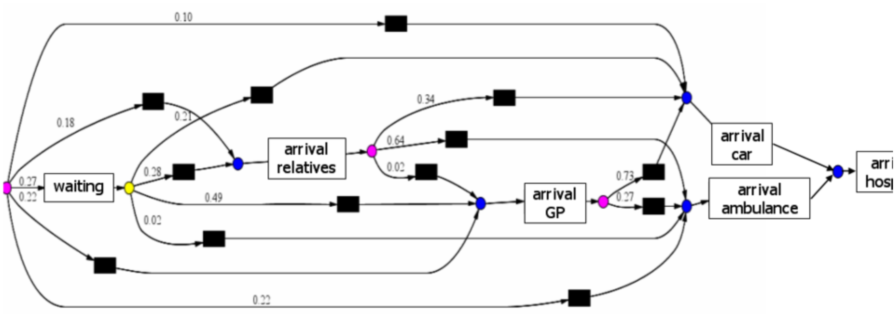

Figure 2. The Petri net representing the pre-hospital process. White transitions represent an event while black transitions are not linked to any event and are only added because of routing purposes.

the discovered Petri net of the pre-hospital process, the bottlenecks and all kinds of performance indicators, e.g. average/variance of the total flow time or the time spent between activities. In particular, bottlenecks can be identified by searching for places which indicate a high waiting time to the next non-black transition. In Figure 2, places coloured blue, yellow or purple represent respectively a low (<6 hours), medium (<12 hours and >=6 hours) and high average waiting time (>=12 hours) in that place. Amongst others, what can be derived from the picture is that after occurrence of the events 'waiting' (patient stopped waiting), 'arrival relatives' and 'arrival GP' on average it still takes considerable time before the next action occurs. Note that the time in between stroke onset and the arrival in the hospital is on average 28 hours and has a standard deviation of 45 hours. Possibly, this is due to underestimation of the stroke symptoms by patients/relatives/GPs. At the same time, a great variability among people is shown. This can be derived from Figure 2 which shows the most frequent paths followed after the occurrence of an event. For example, after stroke onset, 27% of the patients decided to wait instead of calling relatives (18%) or calling a GP (22%).

Clearly, different kind of performance indicators can be obtained for the discovered Petri net. Moreover, once such a Petri net is available, simulations with different parameters can be run to see what the consequences are after removal of a bottleneck, e.g. change in throughput time.

4. Conclusion and future work

This work showed that process mining techniques can be applied successfully to clinical data to gain a better understanding of different clinical pathways adopted by different hospitals and for different groups of patients. It is interesting to analyse the differences, to establish whether they concern only the scheduling of the various tasks or also the tasks themselves. In this way, not only different practices may be discovered that are used to treat similar patients, but also unexpected behaviour may be highlighted.

Also, we have visualized the pre-hospitalisation pathways and identified bottlenecks. Even more interesting results could be found if additional data would be available from other health-care units involved in stroke management, like emergency rooms and rehabilitation clinics. But this requires high integration of the information systems involved, which is not (yet) the case. We believe that making health-care administrators aware about the potential of process mining can foster them to promote this kind of integration.

In this paper, we have applied process mining from a discovery point of view. An interesting future development would be to apply conformance testing. This would enable the comparison between (formal models of) medical guidelines and the execution in practice, i.e. the analysis of non-compliances.

References

[1] Anyanwu K, Sheth A, Cardoso J, Miller J, Kochut K; Health-care Enterprise Process Development and Integration. Journal of Research and Practice in Information Technology, 35(2):83–98, 2003.

[2] Hammori M, Herbst J, Kleiner N; Interactive Workflow Mining – Requirements, Concepts and Implementation. Data and Knowledge Engineering, 56(1):41-63, 2006.

[3] van der Aalst WMP, ter Hofstede AHM, Weske M; Business Process Management: a Survey. In: Proceedings of the 2003 International Conference on Business Process Management (BPM2003). Lecture Notes of Computer Science 2678, 1-12, Berlin, Springer-Verlag, 2003.

[4] van der Aalst WMP, van Dongen BF, Herbst J, Maruster L, Schimm G, Weijters AJMM; Workflow Mining: A Survey of Issues and Approaches. Data and Knowledge Engineering, 47:2, 237-267, 2003.

[5] Yanga WS, Hwang SY; A Process-Mining Framework for the Detection of Health-care Fraud and Abuse. Expert Systems with Applications, 31(1):56-68, 2006.

[6] Lin F, Chou SC, Pan SM, Chen YM; Mining Time Dependency Patterns in Clinical Pathways. International Journal of Medical Informatics, 62(1):11-25, 2001.

[7] van der Aalst WMP, Reijers HA, Song, M; Discovering Social Networks from Event Logs. Computer Supported Cooperative Work 14(6):549-593, 2005.

[8] van der Aalst WMP, van Dongen BF, Günther CW, Mans RS., et al. ProM 4.0: Comprehensive Support for Real Process Analysis. In J. Kleijn and A. Yakovlev, editors, Application and Theory of Petri Nets, pages 484-494. Springer-Verlag, Berlin, 2007.

[9] Micieli G, Cavallini A, Quaglini S; Guideline Application for Decision Making in Ischemic Stroke (GLADIS) Study Group. Guideline compliance improves stroke outcome: a preliminary study in 4 districts in the Italian region of Lombardia. Stroke, 33(5):1341-7, 2002.

[10] Quaglini S, Cavallini A, Gerzeli S, Micieli G, GLADIS Study Group; Economic Benefit from Clinical Practice Guideline Compliance in Stroke Patient Management. Health Policy, 69(3):305-315, 2004.

[11] Mans RS, Schonenberg MH, Song M, Aalst WMP van der, Process Mining in Healthcare: A Case Study. In: Proceedings of Healthinf 2008, International Conference on Health Informatics, INSTICC, 1(1):118-125, 2008.