Business Process Analysis in Healthcare Environments

Álvaro Rebuge
Supervisor: Diogo R. Ferreira

IST – Technical University of Lisbon, Portugal

Abstract. Performing business process analysis in healthcare organizations is particularly difficult due to the highly dynamic, complex, ad hoc, and multi-disciplinary nature of healthcare processes. Process mining is a promising approach to obtain a better understanding about those processes by analyzing event data recorded in healthcare information systems. However, not all process mining techniques perform well in capturing the complex and ad hoc nature of clinical workflows. In this work we introduce a methodology for the application of process mining techniques that leads to the identification of regular behavior, process variants, and exceptional medical cases. The approach is demonstrated in a case study conducted at a hospital emergency service. For this purpose, we implemented the methodology in a tool that integrates the main stages of process analysis. The tool is specific to the case study, but the same methodology can be used in other healthcare environments.

Keywords. Business Process Analysis, Healthcare Processes, Process Mining, Sequence Clustering

1 Introduction

Healthcare organizations worldwide constantly struggle to control and optimize their business processes, as means to improve quality and efficiency while reducing costs [1]. The discipline of Business Process Analysis (BPA) [2] becomes extremely important to enable effective control and improvement of healthcare processes. BPA aims to provide organizations with knowledge to understand how their processes are currently being performed, so that one can detect points of improvement and take action in alignment with strategical objectives [2]. Traditional BPA approaches are mainly based on lengthy discussions with workers, extensive document analysis, careful observation of participants, etc. These efforts are time-demanding, costly, and depend on subjective descriptions made by people, which may not be aligned with reality [3, 4]. The highly dynamic, complex, multi-disciplinary, and ad-hoc nature of healthcare processes aggravates the problem [5]. As result, the value of BPA remains unrealized for most part of healthcare organizations.

However, healthcare is increasingly supported by information systems that record detailed and valuable data about the execution of underlying processes, such as which tasks where performed and for which case, who performed the task, when the task was performed, what informational resource was used, etc. These data can be organized in the so-called event logs so that one has a history of what happened during process runtime.

In this context, process mining [6] offers a promising, innovative approach to solve, or minimize, the problems of traditional BPA. Process mining offers agile means to devise, monitor, analyze, understand and improve different views of real (not assumed) processes, by automatically extracting process knowledge from event logs recorded by systems. Process analysis can be performed on different perspectives [7]: (1) the control-flow perspective focuses on extracting models describing process behavior, namely the activities in the process and their order of execution; (2) the organizational perspective focuses on the relationships between the agents who performed the activities (3) the data perspective focuses on properties and data elements associated with individual process instances or tasks; (4) the performance perspective focus the detection of bottlenecks, as well as the computation of performance indicators, such as throughput times and sojourn times. Other issues such as conformance checking [8] can be studied as well.

The extraction of process knowledge from systems, which can be made automatically to a large extent, can reduce the time and cost required for BPA. Moreover, the models acquired from process mining are based on real executions of the processes; therefore, one gains insight about what is actually happening, and ultimately the knowledge provided by process mining can be used for effective improvement of those processes and of their supporting systems. This work explores the usefulness of process mining to mitigate the problems of traditional BPA in healthcare environments. We also introduce a methodology for the application of process mining techniques that leads to the identification of regular behavior, process variants, and exceptional cases. The approach is demonstrated in a case study conducted at a hospital emergency service. Here we focus the analysis of the acute stroke process. In [5] we already published results of other analyses conducted. The remaining of this document is as follows. Section 2 discusses related work. Section 3 explains the methodology followed. In section 4 the case study is introduced and the acute stroke process is analyzed. Section 5 concludes the document with a discussion on advantages and points that still need improvement.

2 Related Work

Current research on process mining in healthcare is still in its infancy but there are already proofs of its applicability and potential. For example, Mans et al. [9] study two datasets for stroke patients, one of which encompasses the clinical course from admission to discharge, and the second focuses on pre-hospital behavior. Process mining was used with success to discover models for the whole dataset, or for only aspects of interest. By focusing on specific process paths the authors were able to discover models describing different treatment strategies between different hospitals. In a further study [10], Mans et al. also demonstrate that process mining can be used to discover and understand the careflows for gynecological oncology patients. Process mining was used to devise insightful models describing the careflows from different perspectives, including control-flow, performance, and organizational perspective. It was identified, however, the limited ability of process mining to handle the high variability in event data produced by unstructured processes often present in healthcare.Lang et al. [4] developed a data warehouse to support process mining analyses on the radiolology workflows at Erlangen University Clinic in Germany. During the study several control-flow techniques were evaluated and the authors concluded that none of the techniques alone was able to meet all the major challenges posed by healthcare processes, such as noise and incompleteness of resulting data, as well as the richness of process variants and infrequent behavior. Despite the limitations found the authors could still gather useful knowledge about the process, letting to conclude that process mining has great potential to facilitate the understanding of medical processes and their variants. The authors also agree that given the characteristics of healthcare processes it becomes important to apply clustering techniques [11–13] in order to handle the high variability in the recorded behavior. Rather than running mining techniques directly on medical event logs, which can generate very confusing models, by using clustering techniques it is possible to divide traces into clusters, such that similar types of behavior are grouped in the same cluster. Clustering techniques not only let one discover simpler process models for each cluster, but also understand regular/infrequent behavior and variants. We devised a methodology which achieves the latter purpose.

3 Methodology

The proposed methodology is an extension to the work of [14]. In [14] the authors describe a general methodology for the application of process mining techniques. This methodology (Figure 1, tasks in white) comprises (1) the preparation of an event log; (2) log inspection; (3) control-flow analysis; (4) performance analysis; (5) organizational analysis; (6) transfer of results. Log preparation builds the event log by preprocessing event data gathered from information systems. Log inspection provides a first impression about the event log; it includes the analysis of statistical information such as the number of cases, the total number of events, the distribution of number of cases per number of events, the number of different sequences, the number of originators, etc. Then follows the analysis of the control-flow, performance, and organizational perspectives of the process, using the well established process mining techniques. The final step in the methodology is the transfer of results, where the knowledge acquired during the previous analysis steps is presented to the organization for validation.

The work of [14] is extended with a new step after log inspection. This new step is a sub-methodology that includes a set of techniques to cluster the log and pre-analyze the process. The goal is not only to produce simpler models for

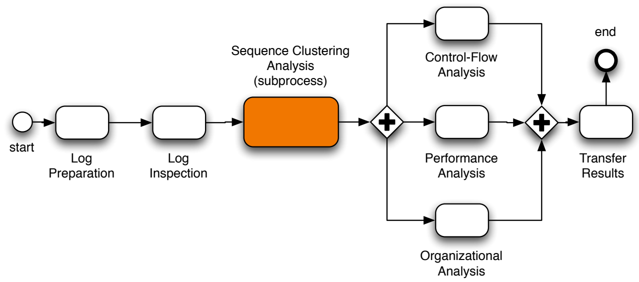

Fig. 1. The proposed methodology for BPA in healthcare is an extension of [14].

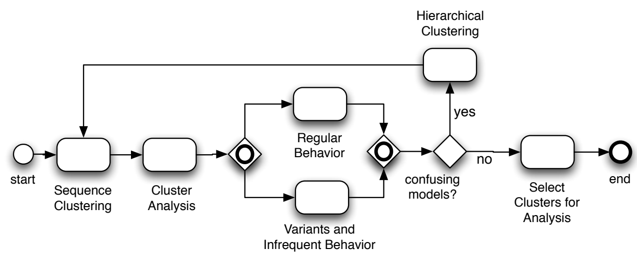

Fig. 2. The Sequence Clustering Analysis subprocess.

the next steps, but also to systematize the analysis of regular behavior, process variants and infrequent behavior. The scope is limited to a set of techniques based on sequence clustering, which we use for two main reasons: (1) it is a good approach to analyze complex, ad-hoc processes [15]; and (2) we are interested in demonstrating the usefulness of this technique for the analysis of real-life healthcare processes. A detailed view of the Sequence Clustering Analysis (SCA) is presented in Figure 2. The rationale of SCA is described at extent in [5]. Here we briefly describe each step.

The first step is to run the sequence clustering algorithm as described in [16] in order to discover the behavioral patterns contained in the event log. The resulting clusters, each one associated with a first-order Markov Chain, will provide insight into the regular behavior, the infrequent behavior, and the process variants. With the behavioral patterns identified, the next step is to understand which clusters represent regular behavior, which ones contain infrequent behavior, where are the process variants, and how much do clusters differ from each other. For this purpose we make use of a cluster diagram, which is a complete, undirected graph where each node represents a cluster, annotated with its respective support (i.e. how many sequences are contained in each cluster). Each edge of the graph is weighed with a function that gives the similarity between the two cluster nodes the edge connects. In practice, the similarity measures how much the Markov chains associated with each cluster differ from each other. The next step in the analysis is to understand the regular behavior of the process. This is given by the cluster, or clusters, with highest support. Therefore, one looks at the cluster diagram and inspects the Markov chain associated with these cluster, or clusters. Once the clusters with highest support have been identified, which represent regular behavior, it is possible to consider the remaining cluster models as variants of the process. To facilitate the comparison of the different variants, one computes the minimum spanning tree of the cluster diagram as means to maximize the similarity path across cluster nodes, so that one can focus in comparing iteratively the most similar variants. Infrequent behaviors can be seen as special cases of process variants and are given by the clusters with lowest support. It may happen that some cluster models might still be hard to understand, and therefore not suitable for analysis. To mitigate this problem one can apply hierarchical sequence clustering, i.e. re-applying sequence clustering to the less understandable clusters. Finally, one selects the clusters of interest for further analysis on the different perspectives of the process.

4 Case Study

The case study takes place at Hospital of São Sebastião (HSS), a public hospital with approximately 300 beds, located in Santa Maria da Feira, Portugal. It was decided to focus the analysis on the operational activity of the emergency department (ED), because: (1) the patient's opinion about the ED has a great influence on the perceived quality and efficiency of the hospital services as a whole; (2) ED processes are one of the most complex to control. To analyze, maintain control of the large amount of complex clinical processes running at the ED is of extreme value for HSS, but it is impossible to realize with traditional methods given the prohibitive amount of time and resources it would be required. However, HSS makes use of an Electronic Patient Record system, called Medtrix, to support the daily clinical activity. The system was developed in-house and spans across the vast majority of the medical departments to provide an integrated view of all clinical information about each patient. The system database is a valuable source of data to perform process mining in order to analyze, understand the emergency service and detect points of improvement.

A process mining oriented database was created by extracting relevant data from the Medtrix database. The latter contains more than 400 non-documented tables, and a lot of effort was needed in conjunction with the system coordinator in order to find out what data could be extracted and where to find it. The resulting database contains emergency activity regarding triage of patients, treatments prescribed, diagnosis made, medical exams performed, and transfer/discharge of patients. The case study database contains several processes. Note that each diagnosis recorded (acute stroke, flu, etc) may refer to a clinical process, and the sequence of events recorded for each patient in a same diagnosis group can be used to discover and analyze the clinical process related with that diagnosis. A process mining tool, named Medtrix Process Mining Studio (MPMS), was developed which let one easily group cases with the same diagnosis, meaning one can quickly build an event log for every clinical process recorded. Furthermore, MPMS implements and integrates the main steps of the proposed methodology.

Results from the analysis of the emergency radiology process, and from an organizational analysis of the emergency service were already reported in [5], which had let us prove the usefulness of the methodology. Here we report the analysis of the acute stroke process. Results were compared with clinical guidelines [17] and validated by a HSS specialist.

4.1 Analysis of the Acute Stroke Process

Acute stroke is a medical emergency related with the loss of brain functions caused by interruption of blood supply to any part of the brain. It can be classified according two major and very distinct categories: (1) ischemic, when due blood blockage; or (2) hemorrhagic, when due blood leakage. Stroke remains one of the leading causes of mortality, hospitalization, and chronic disability worldwide [17].

Log Preparation and Inspection: An event log was created with MPMS by grouping the 83 cases of patients diagnosed with acute stroke. It was found 55 distinct sequences of tasks from the total 83, with 46 occurring only once. It corresponds to a percentage of unique behavior of ca. 55%, which indicates high degree of ad-hoc behavior. It was possible to identify a sub-sequence occurring in the vast majority of traces, meaning there is a behavioral pattern that confers some structure to the process despite the diversity of traces found in the event log.

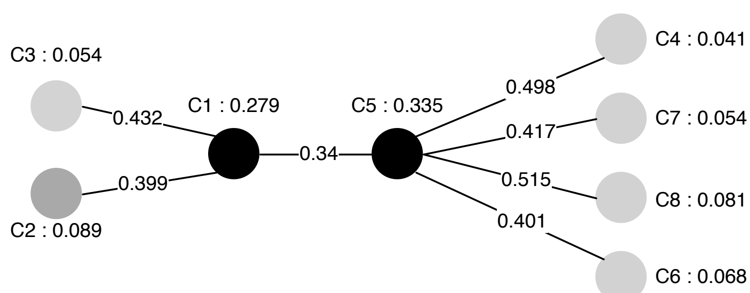

Fig. 3. Minimum spanning tree of the cluster diagram.

Sequence Clustering Analysis: Figure 3 depicts the minimum spanning tree of the cluster diagram obtained when running the sequence clustering algorithm over the event log with 8 clusters as input parameter. Cluster 5 reveals regular behavior. The respective model is shown in Figure 5 a), and one sees that for each patient the following pattern occurs `Request CT Scan > Request Chest XRay > Request Hematology > Request Biochemestry > Diagnosis` . This is the pattern conferring structure to the process, as previously referred. The request of chest x-rays as a common practice is of particular concern because guidelines tell that chest x-rays were found of little use in absence of an appropriate clinical indication (such as those presenting pulmonary or cardiac pathology); therefore, only recommended for a selected set of patients, and not as common practice. The HSS specialist confirmed that requesting chest x-rays is indeed a common internal practice. It was further explained that given the fast paced environment of the emergency it is a way to assure that the patient is not suffering from pneumonia or heart failure, which can be a cause of stroke. Practicing defensive medicine in fast-paced environments such as emergency departments is usual. It serves as means to decrease the risk of miss diagnosis; but at same time it increases the risk of resource waste.

Cluster 1 illustrated cases in which the triage is skipped (referring to patients that give entrance in life threatening conditions), as well as prescription of intravenous normal saline solution for fluid replacement, which some patients need. Cluster 4 revealed evidence of patients presenting delirium or similar complications onset, which has been associated with worse functional outcomes, higher mortality and longer hospital stay. The cases associated with this cluster were not hospitalized but rather transfered to an external hospital. Similarly, the remaining clusters revealed other variants in treatment due the different patient complications onset. An interesting pattern was discovered after applying hierarchical sequence clustering to cluster 1. Figure 4 depicts the cluster with lowest support, referring to infrequent behavior. It is observed that an anticoagulant drug (enoxaparine) is prescribed and the diagnosis of the patient is made without previous brain image study. A brain imaging study is mandatory to distinguish ischemic stroke from hemorrhagic since both conditions have incompatible treatment (a proper anticoagulant drug to treat ischemic stroke would be dangerous for a patient with hemorrhagic stroke because it would increase the bleeding).

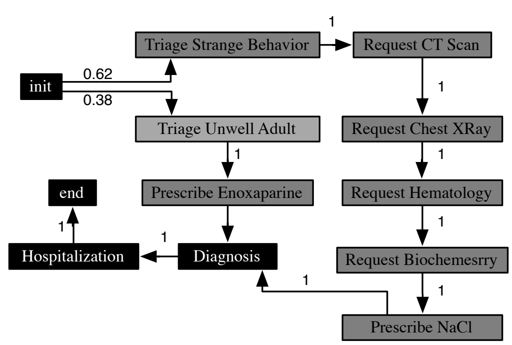

Fig. 4. Process model depicting infrequent behavior discovered in the stroke process and revealing malpractice.

Performance Analysis: The observed process throughput ranges from 37.7 to 552.2 minutes, with an average value of 282 minutes. The dotted chart analysis showed larger timespans from the exams requests to the patient diagnosis, indicating a potential bottleneck. Figure 5 b) depicts the Markov Chain of cluster 5 with each transitions colored according the mean transition time observed. Transitions marked as red, yellow and green are showing high, medium and low values, respectively. Transitions in black refer to values equaling zero. Without compromising the analysis, let us abstract from the different triages. We can also abstract from the different exam requests, in a sense that in this cluster exams are requested in parallel (note the transition times between the sequence of exams equaling zero). Having said that, looking at Figure 5 a), one identifies two points of concern: (1) the transition time from triaging a patient to the exams requests, let us say T1; and (2) the transition time from exam requests to diagnosis, let us say T2. Figure 5 b) plots the histogram and density function of T1. It also shows an equivalent density function for global cases. We assume reasonable to describe both distributions according a log-normal law. In average, stroke patients receive medical attention after triage in 24.73 (-5.11, +6.44) minutes. Guidelines recommend emergency triage in less than 10 minutes. HSS presents an average deviation from guidelines of 14.73 (-5.11, +6.44) minutes, but the long tail of the density curves shows the occurrence of preoccupying transition times. Figure 5 c) plots the histogram and the density function considered for T2. The latter is expressed according a gamma distribution with $shape \approx 3.62\ (-0.921, +1.23)$ and $scale \approx 60.28\ (-16.26, +22.27)$. By definition, the mean transition time $\bar{x} = shape * scale \approx 218.21\ (-14.98, +22.39)$ minutes, i.e. it takes an average of 218.21 (-14.98, +22.39) minutes from exam requests to diagnosis. This transition time is the main responsible for the high values of throughput time observed, and it indicates the problem may be due the performance of laboratory or radiology processes.

The root cause of the performance problems was found at the emergency radiology process; specifically, it was found a bottleneck at the computed tomography (CT) workflow, which is required for all stroke victims. Figure 5 d) shows a discovered variant of the emergency radiology process regarding the specific path followed for handling CT exams. Transitions are colored in order to show where bottlenecks are. From Figure 5 d) one sees the bottleneck at `To Report ITM > Reported by ITM` , which refer to the time that ITM (an external entity) spends to deliver the report of a CT exam. HSS outsources the reporting of CT exams to ITM and expects report delivery within one hour. Figure 5 e) depicts the empirical and theoretical cumulative density function (CDF) on the transition times observed for `To Report ITM > Reported by ITM` . The transition exceeds the expected hour in ca. 40% of cases.

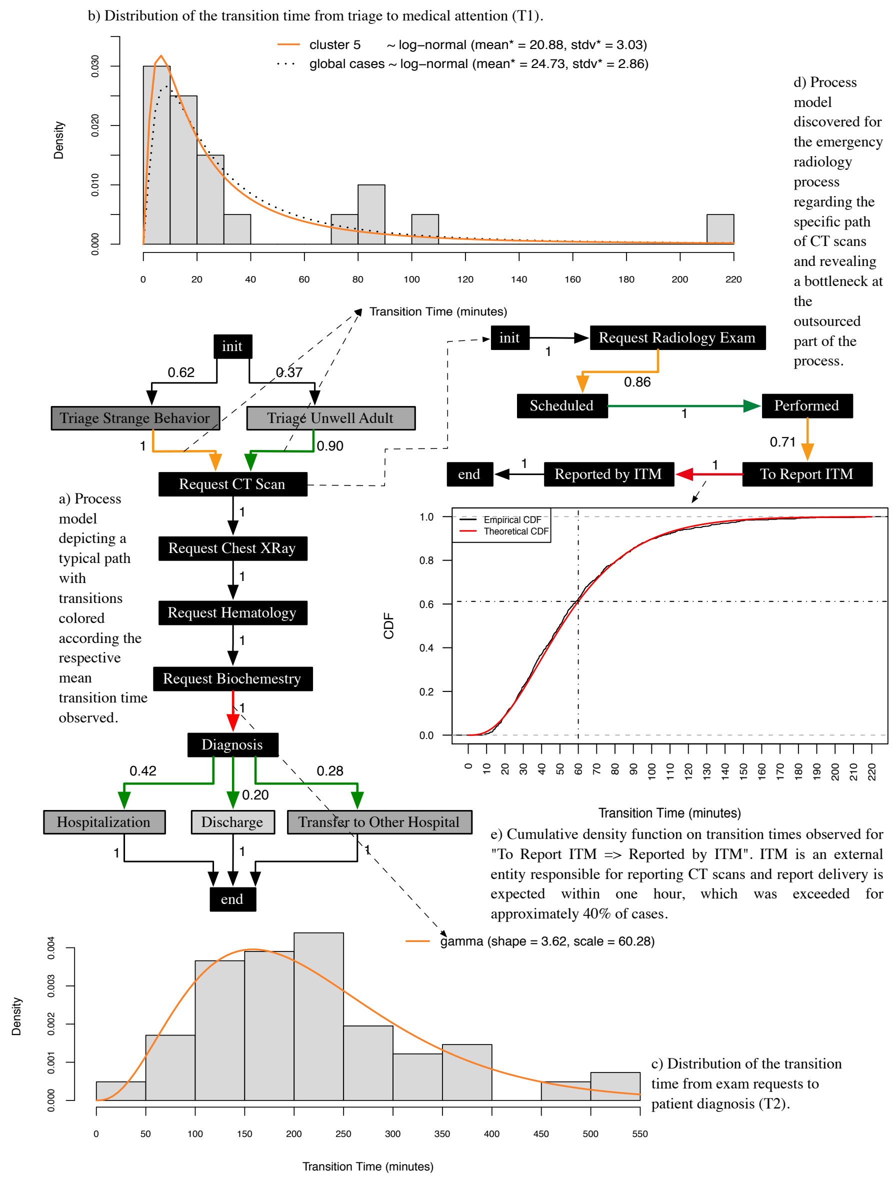

Fig. 5. Control-flow and performance models discovered for cluster 5 of the acute stroke process and their interconnection with the workflow of CT scans observed at the emergency radiology process.

5 Conclusion

Process mining can provide insight into the flow of healthcare processes, their performance, and their adherence to institutional or clinical guidelines. Sequence clustering analysis proved particularly useful to handle the large diversity of log traces, as well as to discern and analyze different behavioral patterns, including regular/infrequent behavior and their variants. It was possible to discover clinical malpractice, potential waste of resources, performance bottlenecks, violation of service level agreements, and potential violations of internal practices. How professionals perceive reality differed in several occasions from the objective results provided by process mining.

The process mining setting developed proved of extreme value for the hospital. Despite the limited resources and effort HSS allocates to perform process analysis, the hospital has now means to be in control of the emergency processes by relying on objective data. Most importantly, the cost of analysis was reduced to an extent that it is now feasible for the organization to perform it whenever is needed. The scalability of the solution is also worth to note because any emergency process can be analyzed attending the patient diagnosis. This would be unfeasible to achieve or maintain with traditional BPA. Moreover, to scale the solution to other departments is now just a matter of acquiring additional data from the hospital system.

A critical point to improve concerns the usability of process mining and the understandability of results to non-experts. IT and non-IT professionals showed interest in using process mining, but they find the concept very technical. Usability has to be improved to a minimum point at which professionals can be trained without needing deep technical understanding, otherwise the hospital is forced to invest in specialists, which is unlikely to happen given the economical panorama and organizational culture.

References

1. Lenz, R., Kuhn, K.A.: Towards a continuous evolution and adaptation of information systems in healthcare. International Journal of Medical Informatics **73**(1) (2004) 75 – 89
2. Darnton, G., Darton, M.: Business Process Analysis. International Thompson Business Press (1997)
3. Van Der Aalst, W., Ter Hofstede, A., Kiepuszewski, B., Barros, A.: Workflow patterns. Distributed and parallel databases **14**(1) (2003) 5–51
4. LANG3bl, M., BURKLE, T., LAUMANN, S., PROKOSCH, H.: Process Mining for Clinical Workflows: Challenges and Current Limitations. EHealth Beyond the Horizon: Get It There (2008) 229
5. Rebuge, Á., Ferreira, D.: Business process analysis in healthcare environments: A methodology based on process mining. Information Systems (2012)
6. Van der Aalst, W., Van Dongen, B., Herbst, J., Maruster, L., Schimm, G., Weijters, A.: Workflow mining: A survey of issues and approaches. Data & Knowledge Engineering **47**(2) (2003) 237–267

7. Van Der Aalst, W., Reijers, H., Weijters, A., Van Dongen, B., Alves de Medeiros, A., Song, M., Verbeek, H.: Business process mining: An industrial application. Information Systems **32**(5) (2007) 713–732
8. Rozinat, A., van der Aalst, W.: Conformance checking of processes based on monitoring real behavior. Information Systems **33**(1) (2008) 64–95
9. Mans, R., Schonenberg, H., Leonardi, G., Panzarasa, S., Cavallini, A., Quaglini, S., van der AALST, W.: Process mining techniques: an application to stroke care. Stud Health Technol Inform **136** (2008) 573–8
10. Mans, R., Schonenberg, M., Song, M., van der Aalst, W., Bakker, P.: Application of Process Mining in Healthcare–A Case Study in a Dutch Hospital. In: Biomedical Engineering Systems and Technologies, Communications in Computer and Information Science, Volume 25. ISBN 978-3-540-92218-6. Springer Berlin Heidelberg, 2009, p. 425, Springer (2009) 425
11. Ferreira, D.R.: Applied sequence clustering techniques for process mining. In Cardoso, J., van der Aalst, W., eds.: Handbook of Research on Business Process Modeling. Information Science Reference. IGI Global (2009) 492–513
12. De Medeiros, A., Guzzo, A., Greco, G., van der Aalst, W., Weijters, A., van Dongen, B., Sacca, D.: Process mining based on clustering: A quest for precision. Lecture Notes in Computer Science **4928** (2008) 17
13. Song, M., Gunther, C., van der Aalst, W.: Trace clustering in process mining. In: Proceedings of the 4th Workshop on Business Process Intelligence (BPI 08) in conjunction with BPM 2008, Springer (2008)
14. Bozkaya, M., Gabriels, J., van der Werf, J.: Process Diagnostics: a Method Based on Process Mining. In: Information, Process, and Knowledge Management, 2009. eKNOW'09. International Conference on. (2009) 22–27
15. Veiga, G.M., Ferreira, D.R.: Understanding spaghetti models with sequence clustering for ProM. In: Business Process Intelligence (BPI 2009): Workshop Proceedings, Ulm, Germany (September 2009)
16. Cadez, I., Heckerman, D., Meek, C., Smyth, P., White, S.: Visualization of navigation patterns on a web site using model-based clustering. In: Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, ACM New York, NY, USA (2000) 280–284
17. Swain, S., Turner, C., Tyrrell, P., Rudd, A.: Diagnosis and initial management of acute stroke and transient ischaemic attack: summary of nice guidance. BMJ **337** (2008)