

Article

Abstract:

This master thesis, aim to discover the impact of process mining techniques on healthcare domain and by comparing these techniques in order to evaluate which one cope better with healthcare insights. Healthcare domain is characterized by its cross-functional and multidisciplinary processes. Nowadays healthcare organizations are fighting to provide high performing medical services. To achieve this goal we need first of all to analyze the Real event data. The process mining research area aims at extracting useful information from hospitals information systems. Then, we work on process mining techniques that automatically discover the process model underlying the existing processes. The performance of process mining techniques on less structured healthcare processes is investigated by using data from foreign healthcare organization. The evaluation will be based on five main criteria which consist on; precision, generalization, fitness, and simplicity and last but not least soundness. To realize this work we will use some plug-ins implemented in the open source Process Mining Framework (ProM). The focus of this work is to use all these tools and techniques to gain information about the Less structured and flexible processes in Health-care and demonstrate which technique is more suitable for unstructured data.

Introduction

This paper aims at exploring how we can exploit process mining [6] techniques to get insights into healthcare processes. Healthcare domain, as seen in the previous section, suffers from many problems such as cost and time consuming. To be able to understand whether a healthcare organization achieves its goals of providing timely, cost effective and quality medical services, we need to analyze its healthcare processes using process discovery techniques and conformance checking[8] [4]. Therefore we used recorded histories of process events that occurred over time in a healthcare process to experiment with some of the process mining algorithms available in the Process Mining (ProM) framework. The focus was on the ability of these algorithms to capture the underlying process. To do so we will use three main algorithms, Inductive Miner (IM), fuzzy Miner (FM) and Heuristic Miner (HM)[1][9],[7].

Processing of the log

The research goal is defined as follows: Study and evaluate the performance of the process mining algorithms in healthcare processes. As the focus of these algorithms is to study the control-flow i.e. the execution order of tasks in a healthcare process, we chose three of the most robust algorithms implemented in ProM: (HM), (FM) and (IM)[1][9]. These algorithms generate a process model underlying an event log based on causal dependencies between tasks and are robust to noise and imbalance. This leads to our second research goal: Analyze the process models derived from the Heuristics Miner algorithm and the inductive miner algorithm and Fuzzy Miner[1],[5],[6] in order to evaluate them depending on a set of criteria such as simplicity, generalization, precision, and fitness. HM, FM and IM are available as the Heuristics Mining, fuzzy Miner and the inductive miner plug-ins in the ProM framework. To be able to do this work, we will use a real event log extracted from a foreign hospital. This event log contains 11143 cases, 150291 events and 624 classes. Data for this study directly comes from the hospital information systems, where all patients' activities are recorded, from visits, lab tests, and administrative tariff and so on. The time span is from Mondays (January 2005) till Thursday (March 2008).

The goal of this master thesis is to compare process mining techniques to show which one cope well with healthcare event log. To achieve this purpose we used different plug-ins implemented in ProM tool which is an open source and free tool. It contains a lot of plug-ins than can we use to discover event logs. The comparison will be based on five criteria[1]:

- **Soundness** or correctness in process mining science is a very important part. A sound model means that this model can reach the end of the executions without any bugs, no tokens left behind, or dead transitions. Sound model is the one which doesn't contain any anomalies (livelocks or deadlocks. . .)
- **Precision** it means that the model mustn't allow for "too much" behavior. A model that does not precise is "under fitting".
- **Generalization** the model should generalize and not restrict behavior. A model that does not generalize is "over fitting" readability.
- **Simplicity** in the context of process discovery, this means that the simplest model that can explain the behavior seen in the log is the best model.

- **Fitness** A model has a perfect fitness if all traces in the log can be replayed by the model from beginning to end.

During the execution of the event log with our three algorithms we have note variance results. The table below concludes the results of the experiments done with three process mining discovery; Inductive miner, Heuristic Miner and fuzzy miner. This comparison is based on the five criteria mentioned in the beginning of this report which are soundness, simplicity, generalization, fitness and precision.

The experiments achieved by applying process mining techniques ‘ heuristic, fuzzy and inductive miner’ algorithms on real data obtained from the information system of the foreign hospital shows that these process mining techniques help a lot to analyze the healthcare event data. Process mining techniques help to discover healthcare domain anomalies and detect problems that we can face every day in our hospitals. The result of these experiments are concluded in table1.

Criteria	Fuzzy miner	Inductive Miner + +	Heuristic Miner
generalization	+	+	+
Precision	-	+	+/-
Soundness	∅	+	+/-
Simplicity	-	+	+/-
fitness	∅	+	+

Table 1 contains the main results of the execution of the event log with our algorithms. We can note here that inductive miner algorithm is typically the most convenient one. It can deal better with healthcare data i.e. spaghetti like models comparing to other algorithms.

Conclusion

In this paper, we have focused on the applicability of process mining on healthcare data using the Inductive Miner; Heuristic Miner and fuzzy Miner algorithms. The goal of this research is to degage the most convenient algorithm that can cope with unstructured information such as those extracted from hospitals information systems. To be able to evaluate which algorithm is the most convenient for healthcare analyses we have discovered the event log with three process mining techniques; IM, HM and FM The models yielded after the application of plug-ins implemented in ProM shows some weakness and strength points in each model. The interpretations of these models has proved that the inductive miner is the most convenient one and can deal with unstructured data well. The analyses achieved on the event log using the three

process discovery techniques FM, IM and HM demonstrate that unstructured information can be discovered successfully but there is still no algorithm which can discover healthcare data perfectly; each one suffer from some deficiencies that hinder process discovery especially with missing information in the event log because generally mistakes can happen while the registration of the data or while the transformation to a CVS or XES files [1]. As a next step, we will try to apply the process mining techniques on real event data extracted from Tunisian hospitals and try to analyze them in order to find out the main problems that our hospitals suffer from and ameliorate their performances.

BIBLIOGRAPHY

- [1] Will van der Aalst, *process mining*, coursera, 2016.
- [2] C.Günther, *Process Mining in Flexible Environments PhD Thesis*, Springer, 2009.
- [3] Coursera, *Coursera Application of Process Mining in Healthcare – A Case Study in a Dutch Hospital*, 2017.
- [4] Maikel L van Eck et al., “PM²: A Process Mining Project Methodology”, in: *International Conference on Advanced Information Systems Engineering*, Springer, 2015, pp. 297–313.
- [5] Process Mining Group, <http://www.processmining.org/online/fuzzyminer>,» coursera, 2016.
- [6] Process Mining Group, www.futurelearn.com, coursera, 2017.
- [7] Felix Kossak et al., “A Rigorous Semantics for BPMN 2.0 Process Diagrams”, in: *A Rigorous Semantics for BPMN 2.0 Process Diagrams*, Springer, 2014, pp. 29–152.
- [8] Process Mining, “Discovery, Conformance and Enhancement of Business Processes”, in: *Springer-Verlag 8* (2011), p. 18.
- [9] Wil MP Van der Aalst, *Process mining: data science in action*, Springer, 2016.