BISE manuscript No.
(will be inserted by the editor)

Applying process mining techniques in a real healthcare case study

M. Mecella · **F. Covino** · **A. Marrella** ·
S. Agostinelli

Received: date / Accepted: date

Abstract The healthcare organizations are under increasing pressure to improve productivity, gain competitive advantage and reduce costs. For this reason, healthcare organizations, such as hospitals try to streamline their processes. In this paper we demonstrate the applicability of process mining in the healthcare domain, using a real case study of '*San Carlo di Nancy*' hospital in Rome (GVM Group). We apply process mining techniques to obtain meaningful knowledge about the patient careflows from so-called event logs, obtained from raw data of hospital information systems. We analyzed these logs using the ProM framework from three different perspectives: the control flow perspective, the organizational perspective and the timing perspective. The results show that process mining can be used to provide new insights that facilitate the improvement of existing careflows.

Keywords Process mining · Healthcare · ProM

M. Mecella
Professor in Engineering in Computer Science at Sapienza University of Rome
E-mail: mecella@dis.uniroma1.it

A. Marrella
Postdoctoral research fellow in Computer Science and Engineering at Sapienza University of Rome
E-mail: a.marrella@gmail.com

F. Covino
CMMI Business Analyst at ENAV
E-mail: federico.covino@gmail.com

S. Agostinelli
Graduated in Computer Engineering at Sapienza University of Rome
E-mail: simone.agostinelli.sa@gmail.com

1 Introduction

Nowdays, the hospitals try to streamline their processes in order to deliver high quality care while at the same time improving revenues and reducing costs. More and more pressure is put on hospitals to work in the most efficient way as possible, whereas in the future, an increase in the demand for care is expected. A complicating factor is that healthcare is characterized by highly complex and extremely flexible patient care processes, also referred to as '*careflows*'. In healthcare organisations, a wide range of processes with different characteristics and requirements are daily managed and executed. The delivery of complex care may involve several departments and organisations, and requires an active collaboration between different professionals and practitioners having heterogeneus skills. Healthcare is thus widely recognised as one of the most promising, yet challenging, domains for the adoption of process-oriented solutions. We demonstrates the applicability of process mining in the healthcare domain, using a real case study of '*San Carlo di Nancy*' hospital in Rome (GVM Group). We apply process mining techniques to obtain meaningful knowledge about the patient careflows from so-called '*event logs*' obtained from raw data of hospital information systems. Process mining aims at extracting process knowledge from that logs in order to discover, for example, both typical paths followed by particular groups of patients and strong collaboration between different hospitalization wards. We analyzed the different careflows both under the control flow perspective (emphasizing the differences between the process models obtained from different cluster of patients), the organizational perspective (looking at the social networks we were able to discover the relationship between the resources of the patient careflow) and the performance perspective (looking at the timing perspective of different activities performed by the patients we were able to discover bottlenecks in the patient careflow). In order to do so, we extracted the event logs from the raw datasets of '*San Carlo di Nancy*' hospital and we analyzed them using *ProM*: the process mining framework. The datasets in question are the following:

- *Ambulatori* (outpatient clinic): each row stores the information about a single healthcare service.
- *Pronto soccorso* (emergency room): each row represents a single emergency room activity.
- *Ricoveri* (hospitalizations): each row represents a single hospitalization taken by a patient.

These three datasets contain raw data about patients treated in both year 2016 and May 2017 for which all the treatment activities have been recorded. We did not use any a-priori knowledge about the careflows of the patients of '*San Carlo di Nancy*' hospital and did not have any process model at hand. The data analyzed are the standard ones of the National Health Service (Servizio Sanitario Nazionale) that the hospitals interchanged with the Regional Authorities (Enti Regione). Therefore the presented analysis can be replicated nationwide.

2 Analysis

In this section, we discuss the results obtained by applying process mining techniques (process discovery, social network analysis and performance analysis) on the datasets of '*San Carlo di Nancy*' hospital already '*converted*' to event logs: '*Ambulatori*', '*Pronto soccorso*' and '*Ricoveri*'. The process mining techniques we used for discovering the process model of the totality of patients (blueprint) are: the *Alpha Miner* algorithm, the *Heuristic Miner* algorithm and the *Inductive Miner* algorithm. In order to establish which process model out of three is representing better the behaviour observed in the event log, and therefore which mining algorithm is the best, we computed the four quality metrics replay fitness, precision, generalization and simplicity. ProM 6.7 allows to compute replay fitness, precision and generalization only. Therefore simplicity is calculated on the basis of this metric: $\#activities + \#splits + \#joins$. The grater is the value, the less is the simplicity of the process model. At the end, the best algorithm will be chosen on the basis of the just mentioned quality metrics. Since the inductive visual miner is high parametrisable (i.e., we can set a-priori the replay fitness we would like to reach in order to obtain a process model in function of this fitness value) therefore it's the algorithm with the best compromise between the just mentioned quality metrics. For this reason, we picked the inductive visual miner as process discovery technique. After discovered the blueprint of all the three datasets, we have performed ad-hoc analysis for these set of clusters:

– foreign patients [1] vs. italian patients [2]
– patients with an age lower or equal than 25 years (resident in Latium)[3] vs. patients with an age greater than 25 years (resident in Latium) [4]

Only for '*Ricoveri*' dataset we have performed a further in depth analysis for these specialist branches: *chirurgia generale*, *medicina generale* and *ortopedia and traumatologia* looking at both pre-operative hospitalization time (the time a patient is waiting for the surgical intervention) and post-operative hospitalization time (the time a patient is waiting for being discharged). After did this phase of analysis and presented results, we performed a second ad-hoc analysis, with the goal of refining the overall procedure, on the following set of clusters:

– patients with an age lower or equal than 40 years (resident in Latium)[5] vs. patients with an age greater than 40 years (resident in Latium) [6]
– patients performing health services in *laboratorio analisi*, with medical prescription compiled by doctors of other hospitals, belonging to any specialist branches [7]
– patients coming from emergency room with hospitalization outcome [8] vs. patients hospitalized not coming from emergency room [9], partitioned in specialist branches
– patients hospitalized with reservation [10], partitioned in different priority classes
– patients hospitalized in *chirurgia endocrina* with main diagnosis 24200/2410/2411 /2419/193 and surgical intervention 064/062 [11].

The third and last ad-hoc analysis involved the following set of clusters:

- patients performing health services in *laboratorio radiologia*, with medical prescription compiled by doctors of other hospitals, belonging to any specialist branches [11]
- Integration of economic values regarding both the patients hospitalized coming from emergency room [12] and patients hospitalized not coming from emergency room [13].
- patients hospitalized resident in Latium [14] vs. patients hospitalized resident in a region different from Latium [15]

In the last analysis phase we have also removed the duplicate data from both the '*Pronto soccorso*' and '*Ricoveri*' datasets.

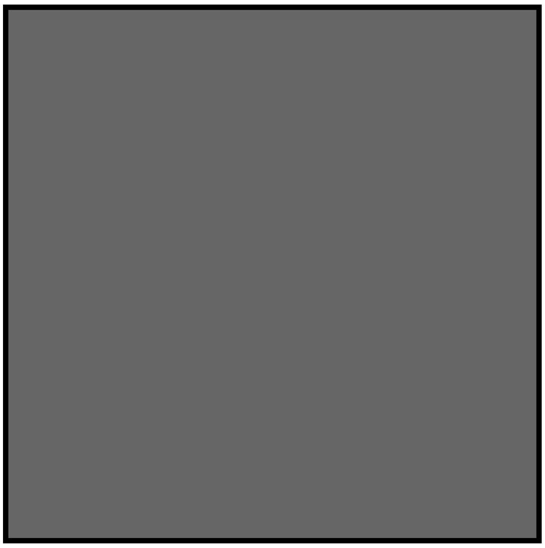

Fig. 1 Please write your figure caption here

Table 1 Please write your table caption here

first	second	third
number	number	number
number	number	number

2.1 Ambulatori dataset

We first showed the blueprint of '*Ambulatori*' in terms of process model, social network and transition system. In this way, we were able to see how patients within the dataset behave, the resources involved in the healthcare flow and the most time consuming activities. Then, we performed the ad-hoc analysis for the clusters already defined in the previous page.

3 Conclusions

References

1. Author, Article title, Journal, Volume, page numbers (year)
2. Author, Book title, page numbers. Publisher, place (year)

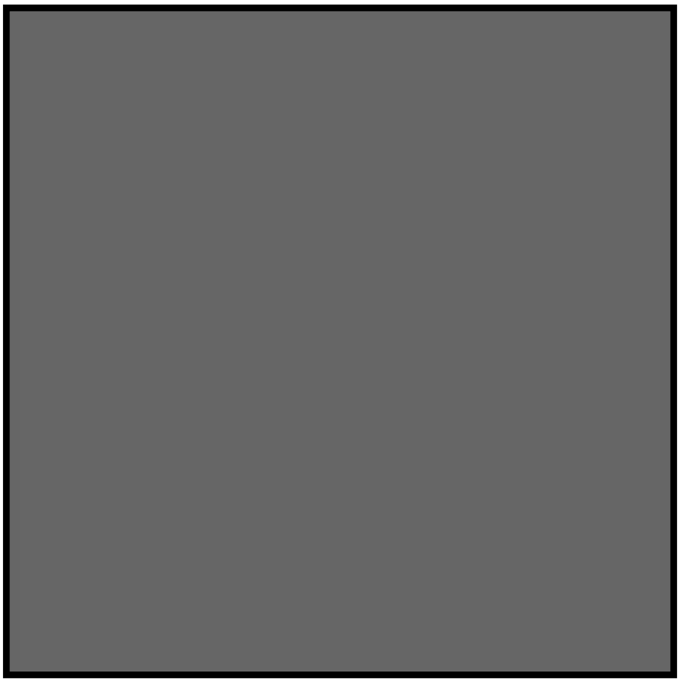

Fig. 2 Please write your figure caption here