Computers in Biology and Medicine 77 (2016) 125–134

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

A guide for the application of analytics on healthcare processes: A dynamic view on patient pathways

Jasmien Lismont [a,*], Anne-Sophie Janssens [a], Irina Odnoletkova [b], Seppe vanden Broucke [a], Filip Caron [a], Jan Vanthienen [a]

[a] KU Leuven, Dept. of Decision Sciences and Information Management, Naamsestraat 69, B-3000 Leuven, Belgium
[b] KU Leuven, Dept. of Public Health, Kapucijnenvoer 35, B-3000 Leuven, Belgium

ARTICLE INFO

Article history:
Received 10 June 2016
Received in revised form
27 July 2016
Accepted 5 August 2016

Keywords:
Patient pathway
Process analytics
Healthcare processes
Electronic healthcare records
Case study

ABSTRACT

Objective: The aim of this study is to guide healthcare instances in applying process analytics on healthcare processes. Process analytics techniques can offer new insights in patient pathways, workflow processes, adherence to medical guidelines and compliance with clinical pathways, but also bring along specific challenges which will be examined and addressed in this paper.
Methods: The following methodology is proposed: log preparation, log inspection, abstraction and selection, clustering, process mining, and validation. It was applied on a case study in the type 2 diabetes mellitus domain.
Results: Several data pre-processing steps are applied and clarify the usefulness of process analytics in a healthcare setting. Healthcare utilization, such as diabetes education, is analyzed and compared with diabetes guidelines. Furthermore, we take a look at the organizational perspective and the central role of the GP. This research addresses four challenges: healthcare processes are often patient and hospital specific which leads to unique traces and unstructured processes; data is not recorded in the right format, with the right level of abstraction and time granularity; an overflow of medical activities may cloud the analysis; and analysts need to deal with data not recorded for this purpose. These challenges complicate the application of process analytics. It is explained how our methodology takes them into account.
Conclusion: Process analytics offers new insights into the medical services patients follow, how medical resources relate to each other and whether patients and healthcare processes comply with guidelines and regulations.

1. Introduction

Recently, patients and healthcare processes are increasingly more tracked in electronic healthcare records (EHR) creating a vast amount of data [1]. These data can provide us with new insights in applications such as:

- Discovering the most frequent and exceptional paths of medical services;
- Comparing pathways of different patient groups;
- Modeling medical resources, e.g. a GP, involved in treatments and analyzing potential bottlenecks in the workflow;
- Analyzing whether patients follow healthcare guidelines, e.g. the American Diabetes Association (ADA) guidelines, and whether the process complies with healthcare regulations.

More specifically, process analytics (PA) techniques can help us acquire these insights because of the dynamic perspective they offer on captured reality. This article focuses on the benefits PA can present to the healthcare environment and more specifically the diabetes domain. Type 2 diabetes mellitus (T2DM) is a growing and very present disease in our society [2]. Moreover, it raises additional challenges both from a social and from an informatics viewpoint.

This research presents a novel guide for the application of PA in healthcare. It was applied on a case study in the diabetes care sector and identifies the opportunities and challenges involved. Section 2 provides an introduction to PA. Consecutively, the case study is described in Section 3 and the specific challenges of applying PA in healthcare are discussed in Section 4. The methods are explained in Section 5 and applied to the case study in Section 6. Finally, the results are discussed in Section 7.

2. Process analytics

Process analytics: This domain refers to the advanced analysis of

* Corresponding author.
E-mail address: Jasmien.Lismont@kuleuven.be (J. Lismont).

http://dx.doi.org/10.1016/j.compbiomed.2016.08.007
0010-4825/

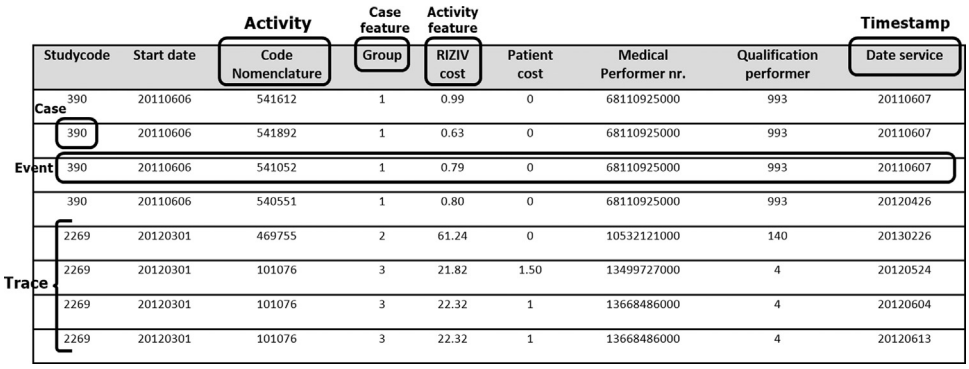

Fig. 1. Extract from the event log. (This is a 1.5-column fitting image).

time-oriented data to derive new insights. In healthcare, one can define this process of a patient as the different steps, treatments and drugs he takes, which is sometimes also referred to as the patient's 'care journey'. In general, there are three main streams of application of PA: process discovery, conformance analysis, and process extension [3]. The former relates to descriptive techniques to evaluate patients' healthcare steps, the timing and duration of medical services and possibly the medical resources involved. Conformance analysis can involve adherence to guidelines and regulations. Extension, on the other hand, enhances an existing process model, e.g. a clinical pathway [4]. For each stream of applications, one can focus on a control-flow, organizational or case (e.g. the patient) perspective. For more information on PA, the reader is referred to van der Aalst [3] and Mans et al. [5]. The added value of process analytics in healthcare was already illustrated by Yoo et al. [6] who specifically suggest its suitability for care processes and performance analysis. Other healthcare applications can be found in Mans et al. [7] and Mans et al. [5].

Tools: Two PA tools, Disco and ProM, were applied in this research. Disco [8] is a commercial tool focusing on visualization and user-friendliness. ProM [9] is an open source tool that works with plug-ins which implement PA techniques.

3. Case description

Our study is carried out on a dataset provided by the Independent Sickness Funds of Belgium (MLOZ). In Belgium, health insurance is obligated for all its residents and is financed by means of social contributions of employers and employees. All contributions and reimbursements are controlled by the Rijksinstituut voor Ziekte- en Invaliditeitsverzekering (RIZIV) through several sickness funds. MLOZ followed patients with T2DM in the context of a nurse-led telecoaching project [10] and provided us with the medical services these patients received.

The data was received in the form of several csv files which were combined into one dataset and prepared as an event log which captures all the activities in which patients are involved. Each row in this file represents an event which is an occurrence of an activity, namely a medical service. This event is characterized by a unique case number (the patient), an activity identifier, a timestamp and optional additional case or activity features. For example, an event can be patient 1 (case) consulting his GP (activity) on the 6th of June 2011 (timestamp). We also retrieved various patient data such as age and gender, as well as data on resource utilization and the costs of care. In total there are 571 patients. Patients range between 18 and 75 years old and were diagnosed with T2DM. They were followed during one year, starting in 2011. The last registered event occurred in 2013. The granularity of the timestamp is on the level of a day. This means that we cannot distinguish a sequence of events in the same day. Arbitrarily, we assume the order of entry as the order of the events per day. Fig. 1 provides an extract of the event log as an illustration. The activities in this context are all medical services the patient received including but not limited to diabetic related services. Medical services are identified by a 6-digit standardized nomenclature code defined by the RIZIV[1] which captures the events only for the purpose of the Belgian national reimbursement scheme of healthcare insurance, e.g. code 102874 stands for an endocrinologist consultation. In this sense event codes can on the one hand be very detailed (if different costs are associated) and on the other hand too generic (e.g. not specifying the subject of a GP visit). All the services are subsequently divided into categories, identified by Odnoletkova [11]. These categories can be general healthcare services or diabetes specific healthcare services and are based on a protocol [11] and a report on diabetes by the 'Intermutualistisch Agentschap'. In an event log, one can follow a patient over time by filtering the events per case. This path of activities a patient follows, a patient's care journey, is referred to as a trace, see Fig. 1.

4. Challenges

Rebuge and Ferreira [12] identified four challenges in healthcare PA: (1) highly dynamic processes; (2) highly complex processes; (3) increasingly multi-disciplinary processes; and (4) ad-hoc processes.

Additionally, we encountered the following difficulties. Firstly, many researchers assume regular behavior, complemented by process variants, but there is no such thing in our case study. Every patient follows a unique path and is thus accompanied by a unique flow. This is common for chronic diseases, such as diabetes, as patients undergo a variety of medical services during their lifetime. In order to obtain comprehensible models, additional pre-processing of the data is required. Hripcsak et al. [13] stated with regards to this topic that patients in EHRs tend to be heterogeneous, which complicates analyses and increases variance. Secondly, there is an abundance of events since each event is characterized by a numerical RIZIV code. This is a consequence of the fact that we work with data from an administrative system with a low abstraction level [7]. Thirdly, although these nomenclature codes are very detailed in one sense, they are not in another. Most of the codes are not linked to a disease unless the event itself is linked to it. E.g. a visit to the ophthalmologist is not necessarily linked to diabetes. Fourthly, each patient was only followed during one year, which means that we cannot research the entire control flow of one patient.

[1] See Nomensoft database: https://www.riziv.fgov.be/webprd/appl/pnomen/Search.aspx?lg=N.