



Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

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



A two-step approach for mining patient treatment pathways in administrative healthcare databases

Ahmed Najjar^{a,*}, Daniel Reinharz^b, Catherine Girouard^c, Christian Gagné^a

^a Laboratoire de vision et systèmes numériques, Département de génie électrique et de génie informatique, Université Laval, Québec, QC G1V 0A6, Canada

^b Laboratoire de simulation du dépistage, Département de médecine sociale et préventive, Université Laval, Québec, QC G1V 0A6, Canada

^c CISSS Chaudière-Appalaches, Secteur Alphonse-Desjardins, Lévis, QC G6V 3Z1, Canada

ARTICLE INFO

Article history:

Received 13 November 2016
Received in revised form 8 March 2018
Accepted 22 March 2018

Keywords:

Process clustering
Process mining
Mixed variables
HMM
k-Prototypes
Healthcare databases
Medical treatment process

ABSTRACT

Clustering electronic medical records allows the discovery of information on healthcare practices. Entries in such medical records are usually composed of a succession of diagnostics or therapeutic steps. The corresponding processes are complex and heterogeneous since they depend on medical knowledge integrating clinical guidelines, the physician's individual experience, and patient data and conditions. To analyze such data, we are first proposing to cluster medical visits, consultations, and hospital stays into homogeneous groups, and then to construct higher-level patient treatment pathways over these different groups. These pathways are then also clustered to distill typical pathways, enabling interpretation of clusters by experts. This approach is evaluated on a real-world administrative database of elderly people in Québec suffering from heart failures.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

Health Informatics is a rapidly growing field that is concerned with applying computer science and information technology to medical and health data [3]. The huge and increasing amount of data in the healthcare databases is a source of information of considerable value for those concerned by how to improve health outcomes and how to better control the rising cost of healthcare. Administrative healthcare databases can unveil the constraints of reality, capturing elements from a great variety of real medical care situations. However, the existing databases contain a huge amount of data. For example, an average of 80–86 million medical services are provided annually to the Quebec population. For 2005–2006, there were nearly 714,000 acute care and more than 465,000 one-day surgeries, for which rather details information has been captured [15]. Yet, there is a lack of effective analysis tools to explore and extract the potential rather detailed information that can be brought by analyzing these databases. Exploring the databases to their full potential represents a challenge as it requires complex preprocessing steps and appropriate developed methods.

Process mining [21,22] is an emerging field in the use of all this information for the benefit of the population. This approach is based on the assumption that each event is an instance of a given process. It consists in developing techniques that allow the production of meaningful clusters of similar types of behavior that are grouped together. In healthcare, it is judicious to distinguish between organizational processes and the medical treatment process. As mentioned by Lenz and Reichert [12], the organizational processes help to coordinate inter-operating healthcare professionals and organizational units. The medical treatment process is linked to the patient. The patient specific medical treatment process depends on case-specific decisions. Decisions are made by interpreting patient specific data according to medical knowledge. This decision process is very complex because it includes medical knowledge, medical guidelines and the individual experience of physicians. Yet, over the last decade, some researchers have proposed techniques and methodologies to cluster such healthcare processes.

Although the medical processes are composed of complex events, previous works found in the literature relied on predetermined events. Moreover, many of the proposals made are not suitable for large-scale datasets. Our contribution is to propose a scalable approach based on a two-step clustering method to handle such data. As the first step, we cluster the medical events that are composing the processes. As the second step, we construct higher-level medical treatment pathways and cluster them to iden-

* Corresponding author.

E-mail addresses: ahmed.najjar.1@ulaval.ca (A. Najjar), daniel.reinharz@fmed.ulaval.ca (D. Reinharz), catherine.girouard83@gmail.com (C. Girouard), christian.gagne@gel.ulaval.ca (C. Gagné)

<https://doi.org/10.1016/j.artmed.2018.03.004>

0933-3657/© 2018 Elsevier B.V. All rights reserved.

tify typical ones. Furthermore, we use a real-world administrative database of elderly people in Québec suffering from heart failures to demonstrate the applicability and the scalability of the proposed methodology.

This paper is organized as follows. We first define the problem in Section 2, followed by an overview of relevant medical process clustering approaches in Section 3. We then present our methodology and the proposed algorithm in Section 4. We propose our methodology to analyze results in Section 5. A case study on the clustering of an administrative healthcare database validating our methodology and algorithm follows in Section 6. Finally, we conclude the paper in Section 7 with some consideration on the relevance and importance of this work.

2. Problem definition

2.1. Case study

The health system and social services in the Province of Quebec are mainly public. The RAMQ (Régie de l'assurance-maladie du Québec) acts as the health insurer for Quebec residents who are covered by a universal public health insurance program (virtually 100% of the people living in the province). The MSSS (Ministère de la Santé et des Services sociaux du Québec) is responsible for the administration of health and social services in the province. Almost all physicians (98%) are participants of the public health system and are exclusively paid by the RAMQ. The health system in Quebec is composed of institutions, community organizations, clinics and private cabinets, all publicly funded. The RAMQ therefore records information on the vast majority of medical, and social services provided to the population.

Heart failure disease is a significant cause of the use of health care resources [13,14]. This disease is a clinical syndrome that normally requires health care to be provided by both specialist and family physicians. Practitioners have guidelines to help them to diagnose and treat the disease. Yet, guidelines that most often apply to an average patient, might not always be relevant for individual cases. This is not surprising given that every patient differs from the “average patient” and has his own peculiar needs. Every health care provider functions with his own biases. The management of heart failure is therefore complex and reflects the integration of decisions made by many actors from different disciplines, but also by the patients [14]. Indeed, in real-life, decisions made on care to be provided often differ from evidence-based recommendations [10].

2.2. Aim

The main goal of our work is to propose a methodology allowing the construction of patient treatment pathways from administrative relational databases, to cluster them in homogenous clusters and to analyze and describe these clusters. Thereby, our approach allows the extraction of latent patterns from patient treatment pathways. As a proof of concept, we are studying medical services given to patients over 65 years old, who live in the province of Quebec (Canada) and who suffer from a heart failure disease.

2.3. Datasets

For this purpose, we have been granted access to administrative health care databases of the RAMQ and MSSS. These databases record all medical acts from health care professionals that are covered by the RAMQ and all hospital stays taking place in the Province of Quebec. Our intent is to exploit these data to reconstruct and cluster patient treatment pathways for elderly people suffering from this disease. We have two databases. The first one contains

data for all hospital stays that occurred in Quebec. These hospitals provide general and specialized care. These data, compiled by the hospitals, relate to acute care (physical and mental) and one-day surgery. The data are organized into 5 tables: hospital stays, diagnostics, services, intensive care, and interventions. The second database contains information on physician fees for medical services according to the health insurance plan administered by the RAMQ. It contains the table of medical services and the table of patient information. For our experiments, we selected individuals in these databases with at least one diagnosis of heart failure (i.e., ICD-10 diagnosis codes 428.0, 428.1, or 428.9) made between January 1, 2000 and December 31, 2005. We rejected individuals who were not 65 years or older at the earliest consultation date or earliest departure date from hospital stays. We obtained 180,027 individuals. We were interested in building and mining the treatment pathways between January 1, 2000 and December 31, 2009 for these patients.

3. Related work

In recent years, researchers have become interested in discovering process models from unlabeled event logs. Some proposals have been made for the clustering of medical processes. Ferreira et al. [6] proposed sequence clustering as an approach to deal with processes. They considered that this approach is a good candidate for process clustering. Indeed, sequence clustering is a collection of techniques with the goal of partitioning a number of sequences into meaningful groups. To reach this aim, they used a clustering algorithm based on first-order Markov chains. Afterwards, many studies based on the use of this approach have been proposed to cluster medical processes. For instance, Rebuge et al. [19] proposed a methodology based on a first-order Markov chain to cluster processes composed of the care events occurring in the emergency department of a hospital. They were interested in the radiology workflow of emergency patients, which is an organizational healthcare process. The events used in their studies include 12 different tasks: the exam request and the 11 possible states of the exam. Furthermore, Elghazel et al. [5] assumed that the clinical pathway is a sequence of hospital stays and each hospital stay is characterized by two qualitative items. Therefore, they used a similarity based approach to compute the clinical pathway dissimilarities and a method based on graph coloring to cluster pathways. The behavior of each cluster is then governed by a finite-state Markov chain model.

Huang et al. [8,9] have applied latent Dirichlet allocation (LDA) to discover latent patterns as a probabilistic combination of clinical activities. They assumed that a patient clinical pathway is represented by a mixture of treatment patterns. They applied LDA to two specific care flow logs concerning intracranial hemorrhage and cerebral infarction, extracted from a hospital information system. The model gives the clinical activity density estimation for each pattern, from which the probabilistic association between an activity and a pattern can be obtained.

All of these studies rely on processes composed of relatively simple and well-defined events. Moreover, each of them was concerned with a specific aspect of a patient's pathway but not with providing an overall view of the care provided. Rebuge et al. [19] worked on organizational processes but not on the patient treatment pathways. Huang et al. [8,9] and Elghazel et al. [5] worked on clustering patient treatment pathways. Huang et al. were interested in the patient treatment pathway in hospital stays which gives a local and micro view in a specific site. Elghazel et al. were interested in a succession of hospital stays. This work provides a more macro and global view but does not involve the complete patient treatment pathways since there are services other than hospital

Download English Version:

<https://daneshyari.com/en/article/6853303>

Download Persian Version:

<https://daneshyari.com/article/6853303>

[Daneshyari.com](https://daneshyari.com)