Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 119 (2017) 197–206

Procedia Computer Science

www.elsevier.com/locate/procedia

6th International Young Scientists Conference in HPC and Simulation, YSC 2017, 1-3 November 2017, Kotka, Finland

Pattern-based Mining in Electronic Health Records for Complex Clinical Process Analysis

Oleg Metsker[a]*, Ekaterina Bolgova[a], Alexey Yakovlev[a,b], Anastasia Funkner[a], Sergey Kovalchuk[a]

[a]*ITMO University, Saint Petersburg, Russia*
[b]*Almazov National Medical Research Centre, Saint Petersburg, Russia*

Abstract

This paper presents the application of text mining methods to the texts in electronic health records (EHR). It is shown in an experimental study how to raise the data possibility to reflect the real medical processes for process modeling tasks. The method is based on the patterns identified in the analysis of medical databases with the physician assistant. EHR is characterized by the gap between common semantic structure and syntactic structure what is important for complex processes modeling. This study aimed at the solution of the problem of knowledge retrieval from EHR by identifying the specifics of their semantic structure and the development of algorithms for interpretation of medical records using the text mining. The medical tests description, surgery protocols, and other medical documents contain many extremely important items for the process analysis. By automating the retrieval of significant data from EHR can be also used for knowledge bases filling. Moreover, the proposed method is developed during the study of actual Russian language medical data of Acute coronary syndrome (ACS) patients from the current specialized medical center which also valuable. The efficiency of this method is demonstrated in the course of correlation analysis of comorbidities on the treatment duration of ACS and in the case of extracted data using to develop process models with complexity metrics at the control-flow perspective of process mining techniques.

© 2018 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the scientific committee of the 6th International Young Scientist conference in HPC and Simulation

Keywords: text mining, EHR, cardiology, modeling, event data, process mining, complex process modeling

* Corresponding author. Tel.: +7-812-921-44-33.
E-mail address: olegmetsker@gmail.com

1877-0509 © 2018 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the scientific committee of the 6th International Young Scientist conference in HPC and Simulation
10.1016/j.procs.2017.11.177

1. Introduction

It is believed that process analysis is a certain means of better understanding and better evaluation of the real processes which gives good opportunities to foresee the changes. The process data must be enriching with auxiliary information and variety of attributes when we deal with clinical process analysis [1]. The EHR describe the process of medical care and compose of various elements which contained information both in structured data and natural language records such as surgery protocols, tests description, patient history, observations, clinical examination, daily status records in intensive care unit and other records. This data representing significant value for the processes analysis, process discovery, process enhancement and bottlenecks identification [2]. Moreover, data retrieval from natural language texts is a relevant research area, considerably developed and provides essential capability [3].

The healthcare records are constantly growing [4]. It becomes impossible to analyze without automation. On the one hand, medical care processes modeling is rapidly developing. As a consequence, the information processing for intelligence analysts in this area appears to be a significant problem. In this regard, the extracted information can be classified as event-driven and parametric entities: events, event attributes, terms, the terminology of the medical field and their relationship. These entities are valuable for modeling using process mining techniques. From the medical point of view valuable chronic and accompanying diseases, the presence of complications during the operation, the results of treatment, the presence of improvement or contrary deterioration, the course of the disease, subjective and objective assessment of the patient, as well as other valuable data for prognosis and decision-making in patients treating. A large variety of different medical information systems is installed in hospitals in Russia [5]. All of them are focused on the automated creation and storage of text entries in the healthcare industry and limitations in technical equipment (especially the lack of hospital networks, providing data transfer with various types of monitoring equipment) and will not allow to automatically gather structured data from patient monitoring systems. On the other hand, the medical field specificity and Russian language peculiarity greatly complicate the methods of intelligent text processing application directly. As a result, there is a gap between a significant development in the text mining field and low efficiency of these methods in medicine. Therefore, the research for adaptation of text-mining methods to the health care required this paper followed is in demand. The following parts are describing the adaptation method of text mining technique for retrieval relevant information for prediction and other important information from natural language records for aggregate semantic links with general EHR structure for treatment process analysis. The extraction of relevant information from medical records is crucially important for the analysis of medical data and for obtaining answers the questions what is the real process. The extracted information may be represented in the form of implicit events of event log or event attributes. Further in the case study presented the application of the implementing implicit events to the event log for process modeling using mining process methods which lead to complexity reduction of the patient treatment flow.

In the course of this study, the proofs that adapted text mining technique really helps to understand the processes are found. The result confirmed the research of 10061 EHR of inpatients with ACS (Acute Coronary Syndrome) treated from 2009 to 2016 years in Almazov National Medical Research Centre is also represented. Moreover, the presents result of the correlation between accompanying diseases and treatment duration obtained from a medical point of view.

2. Related work

The text mining technique is used to extract information from unstructured records [6]. Widely used for data preprocessing in government research, help desk and call center data research, market research and survey analysis [7]–[9]. Text mining methods of information extraction are classified into a pattern based and machine learning methods. Pattern based methods use templates, inference rules, and simple rule algorithms. Machine learning based methods use various texts features, the degree of their proximity and other features for training the models for information extraction with a large number of characteristics [10] with using such algorithms as conditional random fields, hidden Markov models, k-means clustering, singular value decomposition, logistic regression, decision trees, neural network, support vector machines, MARSplines, link analysis, k-nearest neighbors, word clustering.

It is in demand in many fields including medicine due to the fact that still a lot of valuable information stored in unstructured form. Traditionally, there are seven practical following directions [11]:

Download English Version:

https://daneshyari.com/en/article/6901863

Download Persian Version:

https://daneshyari.com/article/6901863

Daneshyari.com