Big Data Research ••• (••••) •••–•••

Contents lists available at ScienceDirect

Big Data Research

www.elsevier.com/locate/bdr

Big Data Compliance for Innovative Clinical Models ☆

Massimiliano Giacalone [a,*], Carlo Cusatelli [b], Vito Santarcangelo [c]

[a] Department of Economics and Statistics, University of Naples 'Federico II', Italy
[b] Ionian Department, University of Bari 'Aldo Moro', Italy
[c] Department of Mathematics and Computer Science, University of Catania, Italy

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Big Data healthcare
Clinical intelligence
Data Mining
Process Mining
Business intelligence

ABSTRACT

In the healthcare sector, information is the most important aspect, and the human body in particular is the major source of data production: as a result, the new challenge for world healthcare is to take advantage of these huge amounts of data de-structured among themselves. In order to benefit from this advantage, technology offers a solution called Big Data Analysis that allows the management of large amounts of data of a different nature and coming from different sources of a "computerized" healthcare, as there are considerable changes made by the input of digital technology in all major health areas.
Clinical intelligence consists of all the analytical methods made possible through the use of computer tools, in all the processes and disciplines of extraction and transformation of crude clinical data into significant insights, new purposes and knowledge that provide greater clinical efficacy and best health pronouncements about past performance, current operations and future events. It can therefore be stated that clinical intelligence, through patient data analysis, will become a standard operating procedure that will address all aspects of care delivery.
The purpose of this paper is to present clinical intelligence approaches through Data Mining and Process Mining, showing the differences between these two methodologies applied to perform "real process" extraction to be compared with the procedures in the corporate compliance template (the so called "Model 231") by "conformance checking".

1. Introduction

Big Data involve statistical and informative methods for analyzing very composite realities and therefore also interpreting the complexity of health data [24].

Today it is very important for business companies to implement models that consider quality and ethic aspects in all processes. It is therefore interesting the application of process mining on compliance model such as the one introduced in Italy by Legislative Decree 8 June 2001, no. 231 (the so called "Model 231"), obligatory for companies that work in healthcare.

However, sometime the audits of compliance model are conducted in a random way and can be rejected by a judge because of ineffectiveness. In this cases, the use of event logs can be very important for the tasks of auditor team that must provide objective and detailed evidences about lack of processes, that is very important for the evaluation of the effectiveness of the model implemented.

This is a Big Data scenario impacting also on the quality of clinical data, opening an important debate and research path showing important correlation with business aspects [25].

So the main contribution of the paper is to consider a process mining through a specialized algorithm applied to event logs dataset in order to determine the level of consistency with a compliance model such as the 231, mandatory for healthcare companies.

The paper is organized as follows: after a brief introduction in the first paragraph, section 2 concerns the role of Big Data in the healthcare system and the related statistical methodologies. The third paragraph lists the key points of the clinical intelligence that are presented in subsequent steps while section 4 shows the main differences between Data Mining and Process Mining.

The fifth paragraph is designed to explain the organizational medical models ex Lgs.D. 231/01 and their importance within a Big Data evolution and, for this reason, in section 6 we introduce our approach based on process mining for "Model 231".

☆ This article belongs to Big Data for Healthcare.

* Corresponding author.
E-mail addresses: massimiliano.giacalone@unina.it (M. Giacalone), carlo.cusatelli@uniba.it (C. Cusatelli), santarcangelo@dmi.unict.it (V. Santarcangelo).

https://doi.org/10.1016/j.bdr.2018.02.001
2214-5796/

Section 7 is devoted to analyze, as a "case study", the innovative model implemented in a Sicilian psycho-pedagogical medical institute, that deals with the process of assistance to patients suffering from learning disabilities. Finally in the last paragraph some final remarks are discussed.

2. Big Data in healthcare

The value proposition of Big Data Analytics in the healthcare system is inspired by the improvements that can be made and the balance between costs and outcomes, a platform capable of returning to the health world five different values [8]: right living, right care, right provider, right value, right innovation. This leads to unlimited opportunities in improving patient care and collective well-being on the one hand and reducing waste and costs on the other [5].

The variety with which health data is presented is equally remarkable [18]. For a long time, we have just logged out medical data on the paper, and unfortunately this is still happening in many small hospital settings. Following the introduction of electronic and digital formats also in the healthcare field, clinical data has assumed different formats [10]. The electronic medical records include in their data very heterogeneous information about the format: audio recordings, magnetic resonances, computerized tomographs and other diagnostic images, electrocardiograms, and the list could continue to be subtle [2,14].

Clinical intelligence plays a key role in the continuous improvement of healthcare and medical processes by covering all of the computer methods and disciplines of extracting and transforming raw clinical data into knowledge, insights, new discoveries affecting clinical decision-making and healthcare decisions [11]. Therefore, it requires highly sophisticated statistical methodologies and analysis, often having to deal with unstructured and complex data types. These methods include machine learning techniques, nonlinear models and multi-algorithm approaches: only through a deep understanding of the existing relationships between the data in the health databases, one can trace and identify the activities to be done and the resulting outcomes, quality, costs, complications, and security. Data is collected and stored even in real-time thanks to the introduction of electronic and also mobile medical devices which have increased the ability to record various types of events during patient hospital stay.

3. Key points in clinical intelligence

Clinical intelligence develops along eight key steps, the first two generally refer to the extraction-transfer-load (ETL) process, the remaining six relate to uploading and processing data across the various analytical platforms:

1) extraction of data;

2) data transformation: data often needs to be standardized or transformed to conform to a common standard; for example, they can be all in one format;

3) loading data: in this step the data is uploaded to the clinical intelligence platform and imported into the most appropriate architectural structure; loaded data can be structured or unstructured as are, for example, medical notes and information expressed in the form of charts;

4) term extraction: through the natural language processing programs, the uploaded data is analyzed in such a way as to extract relevant clinical information from the textual data;

5) semantic analysis: once the clinical information is extracted, semantic analysis seeks to understand the meaning and identify the relationship between the patient and the various medical terms used;

6) statistical analysis: due to the frequency and distribution of extracted data, multiple inferential analyzes can be conducted on the data itself, providing information on abnormal values, trends, data aggregations, and so on;

7) model analysis: using sophisticated techniques, borrowed from artificial intelligence, such as artificial neural networks or machine learning algorithms, it is possible to identify any causal relationships or existing correlations between the analyzed data;

8) meta-analysis: this step takes into account the results of the previous analytical steps to reach a set of conclusions and insights on the data being studied.

Big Data therefore seems to be the turning point that medicine waited, especially regarding the relationship between the patient and the doctor: Big Data, in fact, assures an immediate and complete knowledge by any clinician of the medical past of the patient. If the physician could know all sorts of pathology, surgery, as well as the health services required or benefited by the patient, he would be able to have an efficient health map and be able to do his job better. Big Data healthcare is conceived as a real digital collection of all that the patient has had, taken or required in the medical field: the Big Data, therefore, would be a true digital maxi dossier [16].

The challenge, which is also the main problem of the Big Data, is that all the information on the patient's health that is communicated through social networks to his friends or acquaintances should also be included. This would undoubtedly provide greater efficiency to local and regional health services. To understand the vast amount of potentially available data and how it has grown over time, just think of the role played by mobile technology over the last few years and, more specifically, the proliferation of smartphones and wearable devices [12]:

– the human body produces up to 150,000 billion GB of information;

– by 2020, health organizations will use 25,000 petabytes of data, 50 times the data available today;

– a standard ECG device collects approximately 1000 data points per second, while a 3D-CAT occupies more than 1 GB of space.

In addition, genomics have begun to identify the mechanisms by which genes can modify cell behavior and in particular that certain "expressions" are typical of some or part of the tumors. By acting on some genes, directly or indirectly, you can change the history of the disease [3]. It is expected that by 2025, a billion people in the world will have sequenced their genome, producing billions of gigabytes that will need to be captured, stored, distributed and analyzed to: provide the researchers with the data assets acquired by sequencing the genome of patients already affected by the tumor; define the algorithm to extract information from the collected data, with operative research techniques; develop new therapies, customized on individual cases.

4. Data Mining and Process Mining

Data Mining and Process Mining can be categorized as Business Intelligence for Big Data Analysis that refers to techniques and tools used to analyse large amounts of digital data and retrieve valuable business knowledge out of them. This purpose is as true for data mining techniques as process mining techniques, even if with different perspectives on the analysis and the results they produce. Both techniques are used to analyse large amounts of data, that it would be impossible to analyse manually and they produce information that can be used in business decisions [4]. Data mining techniques are primarily used to find patterns in a large data sets. With data mining techniques it may be possible to find that certain categories of customers demand a certain product, or to find that the customers who most frequently buy product