

Chapter 51

Extraction of Medical Pathways from Electronic Patient Records

Dario Antonelli
Politecnico di Torino, Italy

Elena Baralis
Politecnico di Torino, Italy

Giulia Bruno
Politecnico di Torino, Italy

Silvia Chiusano
Politecnico di Torino, Italy

Naeem A. Mahoto
Politecnico di Torino, Italy

Caterina Petrigli
Politecnico di Torino, Italy

ABSTRACT

With the introduction of electronic medical records, a large amount of patients' medical data has been available. An actual problem in this domain is to perform reverse engineering of the medical treatment process to highlight medical pathways typically adopted for specific health conditions. This chapter addresses the ability of sequential data mining techniques to reconstruct the actual medical pathways followed by patients. Detected medical pathways are in the form of sets of exams frequently done together, sequences of exam sets frequently followed by patients and frequent correlations between exam sets. The analysis shows that the majority of the extracted pathways are consistent with the medical guidelines, but also reveals some unexpected results, which can be useful both to enrich existing guidelines and to improve the public sanitary service.

INTRODUCTION

A huge amount of medical data storing the medical history of patients has made available in recent years by the introduction of electronic medical records. Data mining techniques can be profitably exploited to support healthcare decision making, by extracting a variety of information from these

large medical data collections. For example, it is possible to analyze the relationships between medical treatment and final patient condition, or the medical pathways frequently followed by patients with a given disease.

An actual problem in this domain is to perform reverse engineering of the medical treatment process to highlight medical pathways typically adopted for specific health conditions, as well as discovering deviations with respect to predefined

DOI: 10.4018/978-1-4666-2455-9.ch051

care guidelines. This information can support healthcare organizations in improving the current treatment process or assessing new guidelines.

Care guidelines represent standard medical pathways that have been defined for a variety of clinical conditions. They specify the sequence and timing of actions necessary to provide treatments to patients with optimal effectiveness and efficiency. Sometimes care guidelines include suggestions to treat a particular health state but the real application depend on specific cases and there is not a unique right guideline. Also in this scenario it is important to extract the frequent medical pathways to analyze the most frequent patients' behavior and thus measure the patients' accessibility to the sanitary system.

The approach proposed in this chapter relies on sequential pattern mining to analyze a collection of patients' medical data and extract from it the medical pathways. The analysis is performed on the patients' medical treatments of a group of pregnant women provided by a Healthcare Territorial Agency in the Piedmont region of Italy. Sequential pattern mining is used to reconstruct the actual diagnostic services accessed with high frequency by patients and even their temporal relationships. To this aim, raw medical logs are collected, cleaned, and integrated into a sequence data structure on which the mining process takes place.

Detected medical pathways include the sets of exams frequently done together, the sequences of exam sets frequently followed by patients and the frequent correlations between exam sets. Our analysis show that the majority of the extracted pathways are consistent with the medical guidelines, but also revealed some unexpected results. For example, some important exams occur with a lower frequency than expected, while some exams not included in the guidelines frequently appear in the considered medical data. These results are useful both to enrich guidelines by considering the exams actually done by patients and to improve the public sanitary service by investigating causes that lead patients to do exams privately.

BACKGROUND

The use of data mining techniques in healthcare institutes has taken great attention due to the large amount of generated data (Hardin, 2007). A lot of research has been carried out for enhancing and improving medical practices, disease management and resource utilization.

The medical treatment relationships and condition of a patient for a given disease can be extracted by means of data mining techniques (Cerrito, 2007). The decision support tools for clinical healthcare based on data mining techniques are addressed in Siddiq (2009), Kazemzadeh (2006) and Palniappan (2008). However, these works do not exploit real datasets for experiments. In Stolba (2007), data warehousing and data mining techniques are emphasized essential to provide evidence-based guidelines for clinicians. Health care resources optimal utilization is focused in Dart (2003) and Rossille (2008).

Chen (2007) presents possible side effects of using multiple drugs during pregnancy period with the use of association rule mining approach. The SmartRule technique is used to mine association rules from a saved tabular pregnancy data and finds Maximum Frequent Itemsets (MFI) based on user specified minimum support threshold. The subset of MFIs is selected with targeted attributes by users to derive association rules for a given support and confidence level. The author tries to highlight and warn the drugs that may cause harm to unborn babies.

Gosain (2009) addresses decision tree and association rule mining techniques for finding human immune-deficiency virus (HIV) in order to get insight impacts in management strategies against HIV. This work analyses a real world dataset and investigates its association with data stored through the antiretroviral therapy system, which is a software developed for medical organizations. Decision tree technique identified patterns of higher support and confidence based on several attributes such as age, gender and

13 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/extraction-medical-pathways-electronic-patient/73481

Related Content

Determination of Unithood and Termhood for Term Recognition

Wilson Wong (2009). *Handbook of Research on Text and Web Mining Technologies* (pp. 500-529).

www.irma-international.org/chapter/determination-unithood-termhood-term-recognition/21743

Adding Electric Vehicle Modeling Capability to an Agent-Based Transport Simulation

Rashid A. Waraich, Gil Georges, Matthias D. Galusand Kay W. Axhausen (2014). *Data Science and Simulation in Transportation Research* (pp. 282-318).

www.irma-international.org/chapter/adding-electric-vehicle-modeling-capability-to-an-agent-based-transport-simulation/90076

A Presentation Model & Non-Traditional Visualization for OLAP

Andreas Maniatis, Panos Vassiliadis, Spiros Skiadopoulos, Yannis Vassiliou, George Mavrogonatos and Ilias Michalarias (2005). *International Journal of Data Warehousing and Mining* (pp. 1-36).

www.irma-international.org/article/presentation-model-non-traditional-visualization/1746

Analyzing AI-Generated Packaging's Impact on Consumer Satisfaction With Three Types of Datasets

Tao Chen, Ding Bang Luhand Jin Guang Wang (2023). *International Journal of Data Warehousing and Mining* (pp. 1-17).

www.irma-international.org/article/analyzing-ai-generated-packagings-impact-on-consumer-satisfaction-with-three-types-of-datasets/334024

A Parameterized Framework for Clustering Streams

Vasudha Bhatnagar, Sharanjit Kaur and Laurent Mignet (2009). *International Journal of Data Warehousing and Mining* (pp. 36-56).

www.irma-international.org/article/parameterized-framework-clustering-streams/1822