Artificial Intelligence in Medicine (2005) **34**, 25–39

ELSEVIER

ARTIFICIAL INTELLIGENCE IN MEDICINE

http://www.intl.elsevierhealth.com/journals/aiim

Temporal data mining for the quality assessment of hemodialysis services

Riccardo Bellazzi[a,*], Cristiana Larizza[a], Paolo Magni[a], Roberto Bellazzi[b]

[a]*Dipartimento di Informatica e Sistemistica, Università di Pavia, via Ferrata 1, 27100 Pavia, Italy*
[b]*Unità Operativa di Nefrologia e Dialisi, S.O. Vigevano, A.O. Pavia, Corso Milano 19, 27029 Vigevano, Italy*

Received 22 March 2004; received in revised form 16 July 2004; accepted 22 July 2004

KEYWORDS
Temporal data mining;
Hemodialysis;
Rule discovery;
Temporal abstractions

Summary

Objective: This paper describes the temporal data mining aspects of a research project that deals with the definition of methods and tools for the assessment of the clinical performance of hemodialysis (HD) services, on the basis of the time series automatically collected during hemodialysis sessions.
Methods: Intelligent data analysis and temporal data mining techniques are applied to gain insight and to discover knowledge on the causes of unsatisfactory clinical results. In particular, two new methods for association rule discovery and temporal rule discovery are applied to the time series. Such methods exploit several pre-processing techniques, comprising data reduction, multi-scale filtering and temporal abstractions.
Results: We have analyzed the data of more than 5800 dialysis sessions coming from 43 different patients monitored for 19 months. The qualitative rules associating the outcome parameters and the measured variables were examined by the domain experts, which were able to distinguish between rules confirming available background knowledge and unexpected but plausible rules.
Conclusion: The new methods proposed in the paper are suitable tools for knowledge discovery in clinical time series. Their use in the context of an auditing system for dialysis management helped clinicians to improve their understanding of the patients' behavior.

1. Introduction

Health care institutions routinely collect a large quantity of clinical information about patients' status, physicians' actions (therapies, surgeries) and

* Corresponding author. Tel.: +39 0382 505720; fax: +39 0382 505373.
E-mail addresses: riccardo.bellazzi@unipv.it (R. Bellazzi), cristiana.larizza@unipv.it (C. Larizza), paolo.magni@unipv.it (P. Magni).

0933-3657/$ – see front matter
doi:10.1016/j.artmed.2004.07.010

health care processes (admissions, discharges, exams requests). Despite the abundance of this kind of data, their practical use is still limited to reimbursement and accounting procedures and, sometimes, to epidemiological studies. Moreover, the general claim of biomedical researchers is that these data, that we will refer to as *process data*, have a poor clinical relevance, since they are not collected in controlled clinical trials. However, the growing interest on knowledge management within health care institutions have highlighted the crucial role of process data for organizational learning [1,2]. One of the aspects of organizational learning is represented by the assessment of the *quality* of a hospital service, on the basis of certain performance indicators [3], either related to the *efficiency* of the hospital department or to the *efficacy* of the treatment delivered. In this paper we are interested in the development of data mining tools for assessing the *efficacy* of the treatment delivered by a hospital hemodialysis department (HHD) on the basis of the process data routinely collected during hemodialysis sessions. HHDs manage chronic patients that undergo blood depuration (hemodialysis) through an extra-corporal circuit three times a week for 4 h. The data accumulated over time for each patient contain the set of variables that are monitored during each dialysis session. In other words, the data collected are *sequences* (*inter-session data*) of *multidimensional time series* (*intra-session data*). The majority of these process data is typically neglected in clinical practice, since they are synthesized by few clinical indicators observed at the beginning and at the end of each treatment session. Such clinical indicators are usually related to the well being of patients, and do not contain detailed information about the quality of the treatment, in terms, for example, of blood depuration efficiency or nurse interventions during the dialysis sessions. The goal of an *auditing* system for quality assessment is, therefore, to fully exploit the process data that may be automatically collected in order to: (i) assess the performance of the overall HHD; (ii) assess the performance achieved for each patient; (iii) highlight problems and understand their reasons. The steps (i)—(iii) need first to define a suitable set of automatically computable performance indicators and then to analyze the dialysis temporal patterns, by studying both inter- and intra-dialysis data. In particular, the design and implementation of this system requires the use of methodological tools to perform a crucial temporal data mining task [4]: the discovery of *patient-specific* relationships between the time patterns of monitoring variables and the dialysis performance indexes.

Such relationships can be expressed in terms of association rules, in which the head of the rule is a particular performance index value, and the body is a particular temporal pattern which is frequently associated to that index. An example of these kinds of rules is: ''the dialysis has an insufficient weight loss caused by intra-dialysis hypotension when the trend of systolic pressure is decreasing and the values of heart rate are high''. A more ambitious goal is to extract temporal rules, i.e. rules in which the body is related to the head of the rule by some kind of temporal relationship. In this case, rules can be of the kind ''the dialysis has an insufficient performance due to an insufficient weight loss caused by intra-dialysis hypotension if it *is preceded by* a time period in which the trend of systolic pressure has been increasing and the values of the heart rate have been high.''

In this paper, we introduce a couple of methods for the discovery of patient-specific temporal patterns able to deal with both types of goals. The methodological approach is presented in the context of a project aimed at designing, implementing and testing a system for quality assessment of dialysis sessions. In particular, the paper first describes the application domain and the basic goals of the project; then, it details the methods applied for time series data analysis and temporal rule discovery; finally, it shows both the results of the clinical usage of the system and the outputs of the temporal data mining approaches.

2. End stage renal failure and hemodialysis

End stage renal disease (ESRD) is a severe chronic condition that corresponds to the final stage of kidney failure. In ESRD, kidneys are not anymore able to clear blood from metabolites and to eliminate water from the body. Without medical intervention, ESRD leads to death. In 1999, the ESRD incidence in Italy was of 130 cases per million; due to population ageing, the incidence of the disease is constantly increasing in all industrialized countries [5]. The elective treatment of ESRD is kidney transplant. Blood-filtering dialysis treatment is provided as a suitable alternative to transplant to people in waiting list or to people that cannot be transplanted at all, such as elderly patients. Two main classes of dialysis treatments are now-a-days available: hemodialysis (HD) and peritoneal dialysis. More than 80% of the ESRD patients are treated with HD. In HD the blood passes through an extra-corporal circuit where metabolites (e.g. urea) are eliminated, the acid-base equilibrium is re-established and the