Decision Support Systems 50 (2011) 439–448

Contents lists available at ScienceDirect

Decision Support Systems

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

Using data mining techniques to predict hospitalization of hemodialysis patients

Jinn-Yi Yeh [a,*], Tai-Hsi Wu [b], Chuan-Wei Tsao [a]

[a] Department of Management Information Systems, National Chiayi University, Taiwan
[b] Department of Business Administration, National Taipei University, Taiwan

ARTICLE INFO

Article history:
Received 20 February 2009
Received in revised form 22 October 2010
Accepted 1 November 2010
Available online 6 November 2010

Keywords:
Hemodialysis
Temporal abstract
Data mining
Healthcare quality

ABSTRACT

Hemodialysis patients might suffer from unhealthy care behaviors or long-term dialysis treatments and need to be hospitalized. If the hospitalization rate of a hemodialysis center is high, its service quality will be low. Therefore, decreasing hospitalization rate is a crucial problem for health care centers. This study combines temporal abstraction with data mining techniques for analyzing dialysis patients' biochemical data to develop a decision support system. The mined temporal patterns are helpful for clinicians to predict hospitalization of hemodialysis patients and to suggest immediate treatments to avoid hospitalization.

1. Introduction

End stage renal disease (ESRD), commonly known as uremia, is a severe chronic state corresponding to the final stage of kidney failure. In ESRD, kidneys are not able to purify blood from metabolites or to exclude water from the body. Without medical intervention, ESRD patients may die or remain in intensive care unit (ICU) for a long time. These patients require either a kidney transplant or blood-filtering dialysis treatment. The former treatment is difficult to obtain because of a long waiting list and certain patients, such as the elderly, cannot undergo a transplant. The latter includes two main categories, hemodialysis (HD) and peritoneal dialysis (PD). In HD, the blood passes through an extra-corporal circuit where metabolites (e.g. urea) are eliminated. The acid-based equilibrium is re-established and excess water is removed [3]. PD works on the same principles of solute diffusion and fluid ultra filtration as HD, but the blood is cleaned inside the body rather than through a machine [16]. More than 80% of ESRD patients are currently treated with HD [3]. HD patients typically undergo a dialysis session for 4 h, three times a week. During the long-term dialysis treatment, patients will likely receive hospitalization due to caregiver carelessness or other infections. This has been the main reason for HD patient hospitalization in previous years.

High hospitalization rate for a hospital hemodialysis department (HHD) means low service quality in health care. Therefore, the HHD focuses on reducing hospitalization rate. Preventing hospitalization of HD patients from the perspective of preventive medicine is also very important. This paper develops a decision support system to predict hospitalization of HD patients based on a real dataset collected from a hemodialysis center in Taiwan. The HHD examines HD patients receiving long-term treatment to obtain biochemical data during hemodiaysis sessions, such as hematocrit (Hct), albumin, alkaline-p, cholesterol, triglyceride, blood urea nitrogen (BUN), creatinine, uric acid, Na, etc. [25]. The accumulated data over time contains a set of patient variables that are monitored during each dialysis session. The collected data are sequences of multidimensional time series [3].

For time series data, the temporal abstraction (TA) method proposed by Shahar [22] can be integrated with data mining techniques to support data analysis. For example, Bellazzi et al. [3] successfully applied temporal data mining techniques for assessing the clinical performance of HD services such as preprocessing, data reduction, multi-scale filtering, association rule discovery, etc. They found their approach to be suitable for knowledge discovery in clinical time series. Using an auditing system context for dialysis management helped clinicians improve their understanding of patients' behavior. Adlassnig et al. [1] proposed and discussed promising research directions in the field of TA and temporal reasoning in medicine. They identified and focused on fuzzy logic, temporal reasoning and data mining, health information systems, and temporal clinical databases and recommended developing decision support systems to properly manage the multifaceted temporal aspects of information and knowledge encountered by physicians in their clinical work. Stacey and McGregor [23] surveyed previous research in developing intelligent clinical data analysis systems that incorporate TA mechanisms and present research trends. They suggested the necessity of fusing data mining and TA processes to fully exploit new knowledge from stored clinical data through data mining and apply it to data abstraction.

* Corresponding author. Tel.: +886 5 2732899; fax: +886 5 2732893.
E-mail addresses: jyeh@mail.ncyu.edu.tw (J.-Y. Yeh), taiwu@mail.ntpu.edu.tw (T.-H. Wu).

0167-9236/$ – see front matter
doi:10.1016/j.dss.2010.11.001

For TA rule mining, Sacchi et al. [21] proposed a new kind of TA rule and related algorithms for the extraction of temporal relationships between complex patterns defined over time series. Their approaches could be used in a variety of application domains, and they were already tested on two different biomedical problems. Concaro et al. [5] developed a general methodology for the mining of TA rules on sequences of hybrid events for Diabetes Mellitus. The method was capable to characterize subgroups of subjects, highlighting interesting frequent temporal associations between diagnostic or therapeutic patterns and patterns related to the patients' clinical condition. They concluded that the approach could find a practice for the evaluation of the pertinence of the care delivery flow for specific pathologies.

Based on the literature review, this study integrates TA with data mining techniques for analyzing biochemical data of HD patients to discover temporal patterns resulting in hospitalization. This work develops a decision support system to provide clinicians with association rules and the probability of HD patients' hospitalization for implementing preventive medicine to decrease hospitalization incidence. This system will hopefully help to understand patients' changing biomedical data that leads to hospitalization and to improve service quality of the hemodialysis center. The remainder of this paper is arranged as follows: The Materials and methods Section describes hemodialysis and temporal abstraction, the Development of decision support system Section demonstrates the development of the decision support system used in this paper, the Computational results Section illustrates the experimental results using the combined approach for hemodialysis patients' data analysis, and the last section give the conclusions.

2. Materials and methods

2.1. Hemodialysis

HD for ESRD patients is typically performed in a clinic setting. The diffusion process exchanges solutes and metabolites across a semi-permeable membrane, separating the blood and dialysate. Water is removed from the body using a negative pressure gradient in a process called ultra-filtration. After transit through the dialyzer, the clean, filtered blood is returned to the body. Typically, HD is performed three times a week for about four hours each session. The cost of providing care for HD patients is high. Finding ways to improve patient outcomes and reduce dialysis cost is important. Kusiak et al. [16] demonstrated that data mining, data transformation, data partitioning, and decision-making algorithms are useful for predicting dialysis patient survival. They applied a rough set theory and decision-tree algorithms to analyze biochemical data of HD patients. Sixteen classifiers were produced by these two methods to make predictions. A simple voting scheme was used, with each classifier having one vote. The decision outcome with the maximum number of votes resulted in the predicted outcome. These rules were used to predict the survival of new unseen patients. The results provide a base for analyzing HD patients' data in Taiwan.

The required HD service for ESRD patients has dramatically increased each year in Taiwan. Due to the special characteristics of these patients, both hospitalization times and hospitalization costs are higher than for other patients. Preventive medicine could decrease these costs by analyzing HD lab data to find possible factors. The Bureau of National Health Insurance in Taiwan has also put forward a professional health care quality index according to various overall quality schemes for long-term monitoring. The main items for assessing hemodialysis center quality include: hospitalization rate, serum albumin, clearance rate of urea nitrogen (Kt/V), hematocrit, death rate, sinus reconstruction rate, and weaning rate. Hong [11] used multiple minimum support association rules to discover hospitalization cause and effect factors by analyzing HD lab data. His approach increased the accuracy of health care results and shortened the analyzing period. However, he did not consider time series data recorded during HD sessions.

2.2. Temporal abstraction

Temporal abstraction (TA) is an artificial intelligence technique, which integrates domain knowledge into the data analysis process. TA outlines the evolutionary process of temporal data through a qualitative presentation mode, such as level shifts, periods of stability and trends. Shahar [22] defined TA as a program given a set of time series data including variables, external events, and abstract. The generated abstract description represents previous and current states and data trends. The TA program converts patients' data from a low-level quantitative format to a high-level qualitative description. This presentation format is close to the clinician's specialty vocabulary [23]. Clinicians and domain experts typically work together to discuss rules and knowledge based on TA. These rules and knowledge are very important for generating significant and data-dependent TA and to determine whether these abstractions can be explained correctly to work out correct diagnoses.

Generally, TA can be obtained from both basic TA and complex TA, described respectively as follows. Basic TA is typically indicated by combining state and trend from time series data of existing episodes. The state can be classified as low, normal, and high values. The trend can be classified as increased, decreased, and stable patterns [22]. An episode refers to the data in a time interval [3]. Complex TA describes the temporal relation between basic TAs and complex TAs. Typical complex TAs typically use the temporal operator proposed by [2] to concatenate basic TA (see Fig. 1). The most used operator is Meet, referring to the successively presented precedence order between basic TAs [23]. For example, if a rule states that two basic TAs, A-high and B-high, correspond to result C, and A occurs earlier than B, it can be indicated by A-high Meet B-high Then C.

Initially, TA is applied to data monitoring of patients in intensive care units (ICU) based on the intelligent data analysis (IDA) system which detects abnormal phenomena of patients' temporal data. This provides clinicians with relevant temporal information of patients for subsequent treatments [17]. Fig. 2 shows the IDA system schema, containing several different missions: data validation, data representation, data interpretation, and control tasks. TA is applicable to data representation and the interpretation phase. When TA is applied to the data analysis system during the reasoning process, interpretation or reasoning can be carried out by comparing predefined patterns defined by clinicians or derived from machine learning techniques. Finally, the control tasks present proper treatments according to abnormal phenomena of the data.

Relation	Examples
A operator(PRECEDE) B	Overlap : aaaaa bbbbb
	Meet: aaaaa bbbbb
	Before : aaaaa bbbbb
	Equal : aaaaa bbbbb

Fig. 1. Temporal operator [3].