Journal of Biomedical Informatics 45 (2012) 273–291

Contents lists available at SciVerse ScienceDirect

Journal of Biomedical Informatics

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

# Querying temporal clinical databases on granular trends

Carlo Combi [a], Giuseppe Pozzi [b,*], Rosalba Rossato [a,1]

[a] Dipartimento di Informatica, Università degli Studi di Verona, Ca' Vignal 2, Strada Le Grazie, 15-37134 Verona, Italy
[b] Dipartimento di Elettronica e Informazione, Politecnico di Milano, Piazza Leonardo da Vinci, 32-20133 Milano, Italy

ARTICLE INFO

*Article history:*
Received 15 April 2011
Accepted 16 November 2011
Available online 2 December 2011

*Keywords:*
Temporal clinical databases
Temporal trends
Temporal abstractions
Queries on clinical data
Granularity
Hemodialysis
Temporal relational calculus
Temporal clinical data warehouses
Temporal clinical OLAP

ABSTRACT

This paper focuses on the identification of temporal trends involving different granularities in clinical databases, where data are temporal in nature: for example, while follow-up visit data are usually stored at the granularity of working days, queries on these data could require to consider trends either at the granularity of months ("find patients who had an increase of systolic blood pressure within a single month") or at the granularity of weeks ("find patients who had steady states of diastolic blood pressure for more than 3 weeks").

Representing and reasoning properly on temporal clinical data at different granularities are important both to guarantee the efficacy and the quality of care processes and to detect emergency situations. Temporal sequences of data acquired during a care process provide a significant source of information not only to search for a particular value or an event at a specific time, but also to detect some clinically-relevant patterns for temporal data.

We propose a general framework for the description and management of temporal trends by considering *specific temporal features* with respect to the chosen time granularity. Temporal aspects of data are considered within temporal relational databases, first formally by using a temporal extension of the relational calculus, and then by showing how to map these relational expressions to plain SQL queries. Throughout the paper we consider the clinical domain of hemodialysis, where several parameters are periodically sampled during every session.



## 1. Introduction

Health care institutions collect a huge quantity of clinical information about patients, such as that related to therapies and surgeries, and about health care processes such as admissions, discharges, and examination requests. All of these pieces of information are temporal in nature [1] and are often processed by considering different time granularities (i.e., time units). For example, follow-up visits are usually planned on working days and the recorded clinical parameters are often interpreted according to the month/season when they were observed as well as according to the therapy cycle holding when they were acquired; ICU data are acquired with timestamps up to the unit of minute/second and then may be interpreted according to the hours elapsed since the intervention. A proper representation and reasoning on temporal clinical data [2] is important both to guarantee the efficacy and the quality of care processes and to detect, as soon as possible, any emergency situation. Among the different and heterogeneous (and often domain-dependent) features of temporal clinical data, in this paper we shall focus on two general aspects of these data, namely those of (i) *temporal patterns* and of (ii) *temporal granularity*.

As for the first aspect, temporal sequences of data acquired during a care process provide a significant source of information, not only to search for a particular value or an event at a specific time, but also to analyze the frequency and the regularity of some patterns of temporal data, and to discover sets of events connected by particular temporal relationships [1]. As an example, during a hemodialysis treatment, the difference between two subsequent measurements of the patient's weight, sampled at a fixed time interval of 10 min, should not exceed a fixed value. Furthermore, sound clinical decisions require to observe and detect the increase of some given vital signs (e.g., heart rate – HR, systolic blood pressure – SBP, diastolic blood pressure – DBP) measured during the administration of a drug and to verify that, during this time period, only a finite number of exceptions (i.e., parameter values outside the specific trend) occur. Temporal patterns represent specific sequences of data values relevant to the clinical domain. Typically, temporal patterns are made of some basic *temporal trends* (e.g., increase, decrease, stationary), and state values (e.g., high, low) [3–6], which are the main components for more complex, domain- and context-dependent concepts, named knowledge-based

* Corresponding author. Fax: +39 031 332 7321.
*E-mail addresses:* carlo.combi@univr.it (C. Combi), giuseppe.pozzi@polimi.it (G. Pozzi), rossato@sci.univr.it (R. Rossato).
*URLs:* http://profs.sci.univr.it/~combi/ (C. Combi), http://home.dei.polimi.it/pozzi (G. Pozzi), http://www.scienze.univr.it (R. Rossato).
1 Present address: Mondadori Informatica, Verona, Italy.

1532-0464/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jbi.2011.11.005

temporal abstractions [7]: the Knowledge-Based Temporal Abstraction (KBTA) methodology proposed by Yuval Shahar is widely known in the medical informatics area and has been applied to several clinical domains such as cardiology, diabetology, neonatology, bone-marrow transplantation, oncology [3,8–12].

Moving to consider temporal granularity in the context of temporal clinical data, the introduced examples highlight how time units associated to clinical data need to be carefully considered both when representing and storing and when querying temporal clinical data. Indeed, different time units have to be dealt with even for the same clinical domain: for example, follow-up visits are planned on working days and are fixed by specifying hour and minute of their beginning, while hours and weeks are used as time units to temporally locate and interpret clinical treatments and related data. Furthermore, both the hemodialysis sessions and the time periods related to drug administrations can be considered as interesting observation time units, when querying and abstracting the underlying data. Thus, the considered time units can be calendric units, as days, weeks, hours, and domain-specific units, as hemodialysis sessions: each time unit represents a *time granularity*, i.e., a partition of a time domain in groups of indivisible elements called *granules* [13,14]. Temporal granularities have been deeply studied in the context of temporal databases, where several proposals deal with the definition of temporal data models and query languages, allowing one to use different time granularities both when storing and when querying data [15–19]; specific temporal query models and languages have been proposed also to support the management of temporal clinical data with multiple granularities [20–22].

With respect to this scenario, the focus and the main novelty of this paper are in filling the gap between the proposals dealing with finding temporal trends in clinical data and the proposals related to querying temporal clinical data with multiple granularities. Indeed, usually temporal trends for clinical data are preliminarily specified and derived by considering the adopted, single, granularity of underlying data; the set of derived temporal trends can be, then, queried with different granularities (e.g., "find the increasing trend for SBP starting and ending in the same month", where the increasing trend for SBP has been defined by using the granularity of minutes, i.e., the time unit used for timestamping SBP measurements). A limitation of the current proposals is that it is not usually possible to specify temporal trends based on multiple granularities, in order to query and analyze clinical data: for example, queries as "find patients having an increase of SBP all month long" or "find patients having an increase of SBP from a hemodialysis session to the next one" cannot be simply specified through the existing temporal query languages.

More specifically, we aim at defining a *general framework* for the description and the management of temporal trends by considering *specific temporal features* with respect to (possibly) multiple time granularities: we shall deal with *granular trends*, i.e., temporal trends specified according to some particular granularity and its inherent features. Temporal aspects of data are considered with respect to the context of temporal relational databases, first formally by using a temporal extension of the relational calculus, and then by showing how to map relational expressions to SQL queries.

The major novelties of this paper are:

1. the proposal of a logic-based taxonomy of granularity-dependent temporal trends for clinical data (timestamped at a single, predefined granularity);
2. the application of this taxonomy in designing SQL queries on clinical data.

The proposed framework, thus, focuses on timestamped clinical data: data can come from the measurement of some vital signs (e.g., SBP, DBP, HR), from information related to therapies (e.g., daily drug assumption, number of people involved in a psychiatric contact), from settings of medical devices (e.g., quantity of drug injected by a pump), and so on. The framework does not directly apply to biomedical signals, where sampling, noise filtering, real-time processing are the most common aspects: however, our framework could be suitably applied also to timestamped clinical parameters derived from biomedical signals (e.g., the number of ventricular ectopic and supraventricular events in a 24-h Holter monitoring device). The framework can be considered as a formal basis for the development of On-Line Analytical Processing (OLAP) tools allowing the specification and analysis of trends on time-stamped clinical data [23]: to this regard, our proposal does not deal with monitoring and real-time requirements, as we assume to have (off-line) temporal clinical data that need to be analyzed in a sophisticated way to support clinical decision making.

As a proof of concept, we designed and implemented a running prototype allowing the user to specify granular trends through a graphical interface; the prototype executes the corresponding queries on the considered clinical database and provides the user with a graphical interface displaying the found trends. We then applied the framework and the related prototype to the analysis of hemodialysis data.

The paper is organized as follows: Section 2 presents some important contributions from the literature dealing with the main topic of the paper; Section 3 provides a motivating scenario used throughout the paper, concerning hemodialysis treatments; Section 4 introduces our approach to represent temporal databases and multiple granularities, and the adopted temporal relational calculus; Section 5 focuses on the logic-based taxonomy to describe some granular trends; Section 6 describes the mapping on top of a relational data model of tables, views, and queries defined in Sections 5, 7 presents the prototype tool we developed on top of a real relational database system as a proof-of-concept of our approach and of its feasibility; finally, Section 8 reports concluding remarks and future work.

## 2. Related work

In this section, we briefly discuss some main contributions from the literature in the area of representation and reasoning on temporal clinical data abstractions, and in the area of temporal databases.

### 2.1. Representing and reasoning on temporal clinical abstractions

Several proposals focus on the representation and the analysis of temporal clinical data. In dealing with stored clinical data, we need to consider them at abstraction levels higher than the one at which they are stored [24]. A recent, comprehensive survey on temporal abstraction for clinical data analysis can be found in [25]. Temporal abstraction (TA) has been applied in many application areas after the seminal work of Shahar and co-workers [7,8]. Interesting abstraction mechanisms for decision support systems have been proposed in [3,7,9,24]; these approaches deal with data related to one patient at a time, only, and do not allow one to query the whole database of patients. Moreover, temporal abstractions cannot be specified with respect to different granularities. Our proposal provides a solution to both these limitations.

Bellazzi et al. [26] present an approach to pre-process and interpret clinical time series using TA techniques. The idea is to filter the original time series using TA and then to interpret the new and derived time series by both statistical and artificial intelligence methods. Basic abstractions extract simple trends (e.g., increase, decrease, stationary trends) or states (e.g., low, normal, high values), while trend abstractions allow one to express "fast" or "slow",