A Probabilistic Method for Multiple-Patient Temporal Abstraction

Michael Ramati and Yuval Shahar
Medical Informatics Research Center
Department of Information Engineering
Ben-Gurion University
P.O.B. 653, 84105 Beer-Sheva, Israel
{ramatim, yshahar}@bgu.ac.il

Abstract

Several systems have been designed to reason about longitudinal patient data in terms of abstract, clinically meaningful concepts derived from raw time-stamped clinical data. All approaches had to some degree severe limitations in their treatment of incompleteness and uncertainty that typically underlie the raw data, on which the temporal reasoning is performed, and have generally narrowed their interest to a single subject. We have designed a new probability-oriented methodology to overcome these conceptual and computational limitations. The new method includes also a practical parallel computational model that is geared specifically for implementing our probabilistic approach in the case of abstraction of a large number of electronic medical records.

1 Introduction

The commonly occurring task of Temporal Abstraction (TA) was originally defined as the problem of converting a series of time-oriented raw data (e.g., a time-stamped series of chemotherapy-administration events and various hematological laboratory tests) into interval-based higher-level concepts (e.g., a pattern of bone-marrow toxicity grades specific to a particular chemotherapy-related context) [Shahar, 1997]. Several of the main objectives involved in solving this task include the need for a formal representation that facilitates acquisition, maintenance, sharing, and reuse of the required temporal abstraction knowledge. Most of these aspects were catered for by the Knowledge-Based Temporal Abstraction (KBTA) method [Shahar, 1997] and its extensions [O'Connor et al., 2001; Spokoiny and Shahar 2001; Balaban et al., 2004]. Nevertheless, these solutions, although being evaluated as fruitful, maintained several unsolved subproblems. These subproblems seem common to some of other methods suggested for solving the TA task as well as closely related systems applied in the clinical domain [De Zegher-Geets, 1987; Kohane, 1987; Russ, 1989; Kahn, 1991; Haimowitz and Kohane, 1993; Miksch et al., 1997; Salatian and Hunter, 1999]. Thus, Considering these challenging subproblems suggests an additional method.

At least three subproblems in the former methods can be pointed out, which we propose to solve through the method discussed in this paper. First, raw clinical data, to which the temporal reasoning is being applied, are assumed as certain – that is, typically no mechanism is suggested for handling the inherent impreciseness of the laboratory tests taken to obtain the clinical data. Second, current mechanisms used for completing missing data in an electronic medical record are typically not sound and are incomplete. For example, in the case of the KBTA method, a knowledge-based interpolation mechanism is used [Shahar, 1999]. However, completion of missing values is supported only for bridging gaps between two intervals, in which the proposition (e.g., anemia level) had the same value (e.g., moderate anemia). Furthermore, the value concluded by inference is too crisp, and a threshold is used for computing it with absolute certainty, eliminating uncertainty and leading to potentially unsound conclusions. Third, no special mechanism has been devised for multiple patient abstraction. That is, so far temporal abstraction was performed on a single patient only.

The proposed method, *Probabilistic Temporal Abstraction* (PTA), decomposes the temporal abstraction task into four subtasks, that solve the case of a single patient, and two more subtasks that solve the case of multiple patients. In addition to overcoming the above mentioned subproblems, we also propose a design for a parallel computational model that implements the method.

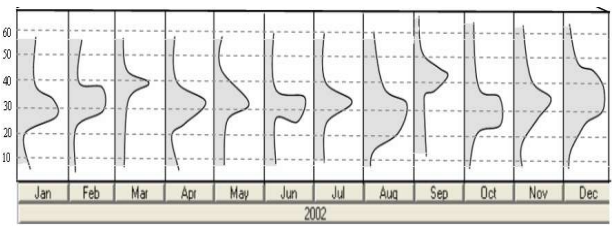

Fig. 1. A typical instance of using the PTA method: the value (*vertical axis*) distribution of a certain medical concept appears for different (in this case consecutive) periods along the time axis. The medical concept, which can be either raw or abstract, and the specification of the set of periods (including the time granularity) are determined by the application using the PTA method.

2 The Subtasks of the PTA Method

Several basic notions in probability theory relate to time, and are important when considering a probabilistic temporal model, task or mechanism. A *stochastic process* $\{X_t : t \in T\}$ is a set of random variables, and may represent a clinical observation, a medical intervention, or an interpretation context of some clinical protocol. The index is often interpreted as time, and thus X_t is referred as the *state* of the process at time t. The set T is called the *index set* of the process. The clinical subtasks specified below are defined in terms of these notions.

2.1 Single-Patient Subtasks

Temporal abstraction for a single patient requires one basic subtask, interpolation, and three interpolation-dependent subtasks – coarsening, transformation and pattern matching.

Temporal Interpolation. Estimating the distribution of a stochastic process state, given the distributions of some of its other states (Fig. 2). For example, estimating the distribution of raw hematological data or derived concepts (such as bone-marrow toxicity grades) during a week in which raw data were not measured, using the distribution of values before and after that week. Applying the interpolation subtask does not increase the abstraction level of the underlying stochastic process, but rather serves the role of a core operation that enables the application of actual temporal abstraction.

Temporal Coarsening. Applying an aggregation function to a stochastic subprocess (Fig. 3). The coarsening subtask abstracts over the time axis and is aimed at the calculation of a stochastic process at a coarser time granularity.

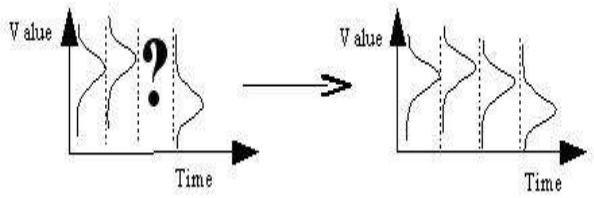

Fig. 2. An illustration of the interpolation subtask. Given the value distribution at several time points, there is a need to calculate an unobserved value distribution. The solution suggested by the PTA considers all value distributions.

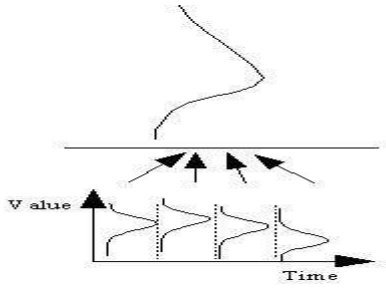

Fig. 3. An illustration of the temporal coarsening subtask. Given the value distribution at several time points, there is a need to calculate an aggregated distribution.

Temporal Transformation. Generating a stochastic process, given stochastic processes of a lower abstraction level. For example, deriving bone-marrow toxicity grade distribution, given the distributions of the raw white blood cell and platelet counts. The transformation subtask abstracts along the (clinical concept) abstraction-level axis.

Temporal Pattern Matching. Locating occurrences of specific values in certain time constraints of high-level predefined temporal variables. As opposed to the temporal transformation subtask, that maps all given data to a higher-level temporal concepts, this subtask is aimed at finding those data sets which complies with the given pattern.

2.2 Multiple-Patient Subtasks

Applying the TA task to multiple patients requires extra subtasks, such as the ones explicated below. However, these subtasks fit also sophisticated needs of abstraction for a single patient.

Temporal Aggregation. Generating an aggregation of stochastic processes. The Aggregation subtask abstracts along the patient axis. This subtask is aimed at the application of aggregation functions, such as minimum, maximum, average, etc. on stochastic processes.

Temporal Correlation. Calculating the correlation between two stochastic processes. The correlation subtask compares two temporal abstractions. This subtask is intended to mainly compare two patient populations, but should work the same when comparing different time periods of the same patient.

3 The PTA Property

The central property of the PTA method is based on the notion of *temporal field*, as defined below. Following this definition, the property states, that each unobserved state of some stochastic process is a linear combination of the temporal fields of the observed states of the process. Thus, the unobserved distribution of bone-marrow toxicity grades is a linear combination of all of the observed distributions, before and after it. A proper basis that will fit the requirements of the PTA property could be found in the following two known definitions.

Let $\{w_{ij}\}_{1 \le i \le m, 1 \le j \le n}$ and $\{\mu_i\}_{1 \le i \le m}$ be constants. The random variables X_i are said to have *multivariate normal distribution*, if:

$$X_i = \sum_j w_{ij} \cdot Z_j + \mu_i \quad , Z_j \sim Normal(0,1)$$

A stochastic process $\{X_i : i \ge 0\}$ is called *Gaussian process* if each state X_i of the process has a multivariate normal distribution.

Uncertain Observations. Observed states of stochastic processes are distributed as a function of the clinical test taken and the clinical data itself. Typically, where states of stochastic processes have a normal distribution, the mean (expectation) of the state is the value sampled, and the variance is determined by the reliability or the precision of the test taken.

Temporal Fields. Calculating a dependent variable given the independent variables as they appear in a multivariate distribution may imply a temporal *persistence* of the independent variables. However, allowing the observed states to induce a *field*[1] over its temporal environment could express temporal knowledge about the stochastic process in question, such as a *periodic behavior* or *change* of the observed states. Thus, for each stochastic process, a temporal field is induced by a time index, which formally means a function that maps time points to states of the stochastic process, as follows:

$$field_{\vec{X}}(t): T \to \mathbb{R}^{\Omega} \quad , X_t: \Omega \to \mathbb{R}$$

For example, suppose a stochastic process with a periodic behavior and cycle length *c*. The temporal field of an observed state of such stochastic process could be as follows:

$$(field_{\vec{X}}(t_s))(t_i) = \sin\left(\frac{\pi}{c} \cdot (|t_i - t_s| \bmod c)\right) \cdot X_{t_s}$$

Temporal Weighting. A specific choice for the selection of the weights of the independent variables can be suggested. These weights should express the notion that the closer-in-time the observed state is – the more relevant it is. That is, the absolute time difference between a dependent state and its observed state should be inversely proportional to the weight of the latter when estimating the former. Therefore, there is a need to choose a monotonic decreasing function of absolute time differences between a dependent state and its inducing observed states. The weighting function is of the following form:

$$w_{\vec{X}}: \Delta T \to \mathbb{R}$$

A natural choice for the monotonic decreasing weighting function would be a normal density, where its variance (σ^2) determines the temporal tolerance of observed states of the stochastic process. Thus, *w* may hold:

$$w_{\vec{X}}(\Delta t) = f_W(\Delta t) \quad , W \sim Normal(0, \sigma^2)$$

Prior Knowledge. Each stochastic process may have a prior knowledge of its typical state. Prior distribution is expressed by giving it the $-\infty$ time index for the temporal field inducer argument, as well as the temporal field argument.

4 Mechanisms of the PTA Method

The main computational concept in our methodology is the PTA chain. A *PTA chain* is defined as the application of any subset of the following composition of subtasks, while preserving the relative order among them:

$$Coarsen \circ Correlate \circ Aggregate \circ Transform \circ Interpolate(data)$$

Temporal Interpolation. The subtask of interpolation is solved by the application of the PTA property. Given the temporal weighting function of a stochastic process, its values need to be normalized to ensure they sum to unity. The subset of sampled states which participate in the calculation process of each unobserved state determines the precision of its distribution, and could be determined given the temporal weighting function. If we interpolate in t_i and have all of the points that are known t_s sampled, then:

$$X_{t_i} = \frac{1}{\sum_{t_s} w_{\vec{X}}(t_i - t_s)} \sum_{t_s} w_{\vec{X}}(t_i - t_s) \cdot (field_{\vec{X}}(t_s))(t_i)$$

For each temporal gap between sampled data, the procedure *Interpolate* generally computes the value distribution of missing states starting at one extreme point (an observed state) until either reaching the prior value distribution (and then doing the same in the other direction) or the other extreme. This leaves out states in which prior value distribution is expected, in order to reduce costs in time and space. For the case in which updates to the underlying clinical data occur, we consider a hierarchical system of states, where each unobserved state has a set of observed parent states, as depicted by Pearl [1987]. In case the sample is updated, propagating the new piece of evidence we are viewing as the perturbation that propagated through a Bayesian network via message-passing between neighboring processors.

The knowledge required for the application of the interpolation subtask includes for each type of PTA chain the definitions of temporal fields (the default is set to persistence of the inducer state), temporal weighting (the default is set to normal density function with mean 0), prior distribution of a typical state (no default is set), and a function that maps each pair of clinical test taken and datum (sampled value) to the distribution of the field inducing state (default sets sampled value to the state's mean).

Temporal Coarsening. The procedure *Coarsen* transforms a given PTA chain to one with a coarser time granularity. The value of such application to a subchain in the requested time-granularity length is a stochastic state, according to the following formula:

$$X_{[t_i, t_j]} = \frac{1}{j - i + 1} \cdot \sum_{k=i}^{j} X_{t_k}$$

Temporal Transformation. The procedure *Transform* returns the application of the given transformation function to the given PTA chains according to the following formula:

$$Y_t = (g(\overrightarrow{X_1}, \ldots, \overrightarrow{X_n}))(t)$$

If *g* has the following form, then |*g*| is called a *rate* transformation, and *sgn(g)* (positive, negative or zero) is called a *gradient* transformation:

$$(g(\vec{X}))(t_i) = \frac{X_{t_i} - X_{t_{i-1}}}{t_i - t_{i-1}}$$

For example, in the case of a contemporaneous transformation of several arguments (e.g., height and weight) into a higher-level abstraction (e.g., body-mass index), the time-series of the arguments are the same as the of the abstraction. However, a context of a Bone-Marrow Transplantation (BMT) is defined as the application of the fol-

[1] In the sense of an electromagnetic field.

lowing transformation function to the Boolean day-granularity stochastic process that represents a BMT:

$$(g(\overrightarrow{BMT}))(t) = BMT_{t-3} \vee \ldots \vee BMT_{t+90}$$

Temporal Pattern Matching. The procedure *Match* returns the probability of the occurrence of the given temporal pattern in each subinterval of the given time interval. The temporal patterns are represented by regular expressions, where the concatenation operator stands for temporal succession, Kleene-closure stands for a temporally unbounded repetition and the alphabet Σ_g is the discrete finite vector space spanned over the sample or transformation spaces of random variables composing the temporal pattern, and *g* is the respective time granularity. That is, a letter $\sigma \in \Sigma_g$ is a vector, which its *i*-th coordinate is some possible discrete value of the *i*-th variable composing the pattern. For example, a pattern of platelet half-life is composed of bone-marrow transplantation (first coordinate) and platelet state (second coordinate), using ϕ to represent all value possibilities, and an hour-granularity:

$$\langle true, \phi \rangle \langle \phi, high \rangle$$
$$(\langle \phi, high \rangle \cup \langle \phi, normal \rangle) * \langle \phi, low \rangle$$

The probabilistic nature of the underlying data requires the temporal matching mechanism to compute the conditional probability of the occurrence of each letter given the occurrence of the subpattern preceding it. In order to identify the data used for the probability computation of the preceding subpattern, one needs to find the time-series of the transformation arguments of each coordinate in the preceding letters. This is accomplished by the definition and application of functions of the following form:

$$h_{\vec{Y}}(t) = \langle \overrightarrow{T}_1, \ldots, \overrightarrow{T}_n \rangle$$

Given these functions, the interpolation mechanism is used only for the resulting time intervals as well as given the already computed subpattern-match probability. The probability the occurrence of some pattern in a given time interval is thus the joint probability of its letters, i.e., the multiplication of their conditional probabilities. Computing the value distribution of some letter coordinate given its conditional distribution (when matching a new interval, that is not conditioned with the time-points given in the former interval matched) is done by removing the weighted temporal fields from the (interpolated) conditional distribution. The matching process continues until the probability for the occurrence of some letter's coordinate equals or lesser than its prior probability, or until the pattern was fully matched.

Temporal Aggregation. Applied to stochastic processes of the same sample space and independent patients, resulting in a new stochastic process. This measure is computed as a new PTA chain, where each of its state is the application of some aggregative function (minimum, maximum, average, etc.) to the corresponding states of the given PTA chains. Suppose *agg* is some aggregation function and t_i is some time-point, then:

$$agg_{t_i}(\overrightarrow{X}_1, \ldots, \overrightarrow{X}_n) = agg(X_{1t_i}, \ldots, X_{nt_i})$$

Temporal Correlation. Applied to stochastic processes of different sample spaces, independent patients or same, resulting in a series of correlation factors. This measure is computed as a time series of correlations between corresponding states of the given PTA chains:

$$\rho(X_{t_i}, Y_{t_j}) = \frac{Cov(X_{t_i}, Y_{t_j})}{\sqrt{Var(X_{t_i}) \cdot Var(Y_{t_j})}}$$

An example for a single patient would be the contemporaneous correlations between height and weight or correlation of height during different periods for the same person.

6 The Parallel Computational Model

The computational model used to compute a PTA chain is *goal-driven, bottom-up* and *knowledge-based* (the pattern matching mechanism is *top-down*, however, as explicated above). The main algorithm is thus required to compute the result of a PTA chain (the goal), given the transformation and interpolation functions (the temporal knowledge) as well as the access to the clinical data, beginning at the raw (lowest abstraction level) clinical data. The computational model is parallelized in three orthogonal aspects: (1) Time, during the calculation of the PTA chains' states; (2) Transformation, during the calculation of the transformation arguments; and (3) Patient, during the calculation of the PTA chains for multiple patients.

The Main Algorithm. A parallel algorithm is typically presented in terms of a theoretical model for parallel computing: the *Parallel Random-Access Machine* (PRAM) [Brent, 1974]. In its basic architecture, the PRAM model includes *p* serial processors that have a shared memory. We shall assume the PRAM supports *concurrent-read*, i.e., multiple processors can read from the same location of shared memory at the same time.

The following procedure computes the PTA chain for the given patient, goal and index set. First, it retrieves the goal's transformation function. In case it does not exist, it retrieves the raw clinical data, interpolates the missing clinical data, and may change in parallel the time granularity. If the transformation function was found, its arguments are retrieved, and the transformation is applied in parallel.

Complexity of the Computation. The results of asymptotic run-time analysis for parallel combinatorial circuits (Brent's theorem) [Brent, 1974] can be applied to such analysis of the overall algorithm. The description of the different PTA mechanisms suggests parallelizing the interpolation subtask and the temporal coarsening subtask on the time axis, and the transformation subtask on its arguments axis. Multiple-patients subtasks are parallelized on the time axis as well as on the patient axis. Let *args* be the the maximal number of arguments in all transformation, let ΔT be the temporal length of the requested PTA chain, let *p* be the number of processors, and let *level* be the number of transformations applied until the requested goal is reached, then the corresponding PTA chain are created in:

$$O(args^{level+1} \cdot \Delta T \cdot |subjects| / p + level)$$

7 Implementation

The PTA architecture is in the process of fully being implemented using the C++ programming language, the Standard Template Library (STL), and the MPICH2 implementation of the Message-Passing Interface (MPI)[2], an international parallel programming standard. The implementation is thus object-oriented and platform-independent. The implementation is in the process of fully integrated into the IDAN system [Boaz and Shahar, 2005], which satisfies the need to access medical knowledge and clinical data sources.

8 Discussion

In this paper, we proposed a probabilistic method to solve the task of abstraction of longitudinal clinical records, and described a *scalable* [Hwang and Xu, 1998] parallel computational model that implements it. The new method has removed several limitations of former methods. First, the use of PTA chains enables the expression of uncertainty in the underlying clinical data. Second, two mechanisms were developed for temporal abstraction of the clinical data of multiple patients. Third, the interpolation mechanism was shown to be sound and complete. However, the previous model's assumptions were replaced with those of the other's: observed clinical data are assumed to be independently distributed. This assumption could be easily removed, provided the extra medical knowledge of conditional distribution functions for the underlying stochastic processes available.

When dealing with probability of events that occur over time, it is not unusual to assume the Markovian property. This property states that the conditional distribution of any future state, given the present state and all past states, depends only on the present state and is independent of the past. Our probabilistic temporal model, however, cannot assume this known property for a couple of reasons. First, the property does not hold for temporal chains, in which past states help in forecasting future states. Second, the assumption that is actually needed is one that would explicitly state the influence of future states on *interpreting* past states, and in particular on interpolating the present state, given past and future states.

Finally, there are two more points that are worth mentioning, while comparing the proposed method to the model used to solve the temporal abstraction task, as part of the KBTA method. First, as it was specified in section 4, the interpolation in the PTA model is performed at the lowest abstraction level only, as opposed to being repeatedly performed at every abstraction level in the former method. Second, the temporal patterns can be acquired in any temporal representation language, such as CAPSUL [Chakravarty and Shahar, 2001] or TAR [Balaban et al., 2004], assuming it is reducible to regular expressions in the temporal semantics attributed above. The expressions of the source language are then compiled to the formal regular expressions beforehand, thus gaining modularity as well as run-time computational speedup.

We are in the process of fully implementing the new architecture and evaluating it on a large longitudinal clinical database.

[2]http://www.mpi-forum.org/

Acknowledgments

This work has been supported in part by NIH award No. LM-06806 and Israeli Ministry of Defense award No. 89357628-01. We would like to thank Denis Klimov and Efrat German of the Medical Informatics Research Center in Ben-Gurion University.

References

[Balaban et al., 2004] Applying Temporal Abstraction in Medical Information Systems. *Annals of Mathematics, Computing & Teleinformatics* 1(1), 54-62.

[Boaz and Shahar] A Framework for Distributed Mediation of Temporal-Abstraction Queries to Clinical Databases. *Artificial Intelligence in Medicine* (in press)

[Brent, 1974] The Parallel Evaluaion of General Arithmetic Expressions. *Journal of the ACM* 21.2: 201-206

[Chakravarty and Shahar, 2001] Specification and Detection of Periodicity in Clinical Data. *Methods of Information in Medicine* 40(5), 410-420.

[Chakravarty and Shahar 2000] A Constraint-Based Specification of Periodic Patterns in Time-Oriented Data. *Annals of Mathematics and Artificial Intelligence* 30 (1-4)

[De Zegher-Geets, 1987] IDEFIX: Intelligent Summarization of a Time-Oriented Medical Database. M.S. Dissertation. *Program in Medical Information Sciences*, Stanford University School of Medicine

[Haimowitz and Kohane, 1993] Automated trend detection with alternate temporal hypotheses. *Proceedings of the 13th International Joint Conference on Artificial Intelligence*. Morgan Kaufmann, San Mateo 146–151.

[Hwang and Xu, 1998] Scalable Parallel Computing, WCB McGraw-Hill.

[Kahn, 1991] Combining physiologic models and symbolic methods to interpret time varying patient data. *Methods of Information in Medicine* **30**(3) 167–178.

[Kohane, 1987] Temporal reasoning in medical expert systems. Technical Report 389, Laboratory of Computer Science, Massachusetts Institute of technology, Cambridge, MA.

[Ladkin, 1986a] Time representation: A Taxonomy of interval relations. In *Proc. of the AAAI*, pages 360-366

[Ladkin, 1986b] Primitives and Units for Time Specification. In *Proc. of the AAAI* pages 354-359

[Miksch et al., 1997] Time-Oriented Analysis of High-Frequency Data in ICU Monitoring. In *Intelligent Data Analysis in Medicine and Pharmacology*, N.

Lavrac, E.T. Keravnou, and B. Zupan, Editors. Kluwer. p. 17-36.

[O'Connor et al., 2001] RASTA: A Distributed Temporal Abstraction System to facilitate Knowledge-Driven Monitoring of Clinical Databases. *MedInfo 2001*, London

[Pearl, 1987] Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann Publishers.

[Russ, 1989] Using hindsight in medical decision making. In *Proceedings of Symposium on Computer Applications in Medical Care*. IEEE Computer Society Press, New York NY 38–44.

[Salatian and Hunter, 1999] Deriving trends in historical and real-time continuously sampled medical data. *Journal of intelligent information systems*. **13**: p. 47-71.

[Shahar, 1997] A Framework for Knowledge-Based Temporal Abstraction. *Artificial Intelligence* 90 79-133

[Shahar, 1999] Knowledge-Based Temporal Interpolation. *Journal of Experimental and Theoretical Artificial Intelligence* 11:123-144

[Spokoiny and Shahar 2001] A Knowledge-based Time-oriented Active Database Approach for Intelligent Abstraction, Querying and Continuous Monitoring of Clinical Data. *MedInfo 2004*:84-8.