Representing trends and trend dependencies with multiple granularities*

Carlo Combi and Rosalba Rossato
Dipartimento di Informatica - Università degli Studi di Verona
{*combi,rossato*}@*sci.univr.it*

Abstract

In this paper we propose a formal framework allowing the expression of temporal trends which involve multiple granularities. We first discuss the representation of trend dependencies and then we propose a characterization of trends according to different temporal features considering also the relationships with the specific temporal granularities involved in the trend definition. A suitable temporal logic is introduced to formally represent time-evolving relations, granularities and granular trends.

1 Introduction

Every day, a huge amount of data is collected in the form of event time sequences [1]: common examples are the recording of values of stock shares during a day, bank transactions, accesses to a computer from a network, and the events related to failures in an industrial plant. These sequences represent significant sources of information not only to search for a particular value or event at a specific time, but also to analyze the frequency and the regularity of certain patterns into temporal data, and to discover sets of events related by particular temporal relationships, according to several time partitions, *i.e., time granularities* [1].

In several applications it is interesting to describe temporal properties or temporal trends having specific time features; for example, in a database containing information about clinical treatments of a set of patients, it could be useful to monitor that the weight of a patient changes during a cycle of treatment (or during a month) but the difference between two consecutive values of the patient's weight does not exceed a given value.

Even though trend detection and analysis are a research topic considered in several research areas [9], at the best of our knowledge, few efforts have been devoted to providing a general framework allowing the specification of different trend features with respect to several time granularities.

In this paper we focus our attention on the definition of a formal framework, based on a suitable temporal logic, which allows one to represent time-evolving relations and to express temporal trends involving different time granularities. We introduce different and orthogonal features for classifying and characterizing *granular trends* (*i.e.,* trends defined according to a specific granularity). To do that, we extend the notion of *labeled linear-time structure* in order to represent time-evolving relations and different time granularities. Temporal constraints and relationships between the represented tuples can be expressed by means of logical formulae.

2 Related Work

In general, the notion of temporal trends can be related to the main topic of deriving implicit information from raw data with data mining techniques [11], query techniques and languages [8], or abstraction-based techniques [9]. The process of discovering temporal patterns of events usually starts with a description of variables which represent events and of temporal constraints between these variables: the goal of data mining is to find temporal patterns, *i.e.,* instantiations of the variables in the structure which frequently appear in the temporal data sequence [1, 11].

Frequently, it could be useful to describe the specific time unit associated to an event; this information can be represented by means of a time granularity [2, 4]: as an example, in [1] the authors proposed a formal framework which allows one to express data mining tasks involving time granularities. In the context of temporal databases, a significant work related to the problem of describing temporal trends which involve multiple granularities is proposed in [11]. In this work, Wijsen introduces a new kind of temporal functional dependency (TDs) which can be considered as a generalization of the common functional dependencies because it allows the comparison of tuples over the time by means of comparison operators. A basic kind of temporal trends (but not related to specific time granularities) has been proposed

*This work has been partially supported by contributions from MIUR under the program COFIN-PRIN "Intelligent analysis of hemodialysis data for the improvement of care processes".

by Shahar in [9]. In this work, the author investigated the specific task of context-sensitive abstraction and interpretation of time-stamped data, focusing his attention on several subdomains of clinical medicine, in which the task of abstraction of data over time occurs frequently.

3 A logical formalism for temporal databases with granularities

In this section we first recall the notion of *labeled linear-time structure* to represent time granularities [4] and then we introduce an extension which allows one to represent temporal relations possibly related to time granularities.

3.1 Labeled linear-time structures

We model time granularities according to the following definition [4], which specializes the more general definition of granularity given in [2]. Let T be the *time domain* and I be the domain of a granularity G, called *index set*. Informally, a granularity is a special kind of mapping from the index set to subsets of the time domain.

In the following formal definition, we assume that both the index set I and the time domain T are the linear discrete domain $\mathbb{N}$ ordered by a relationship $\leq$ (denoted $(\mathbb{N}, \leq)$).

Definition 1. *A granularity is a mapping $G : \mathbb{N} \rightarrow 2^{\mathbb{N}}$ such that:*

1. *for all $i < j$, for any $n \in G(i)$ and $m \in G(j)$, $n < m$;*
2. *for all $i < j$, if $G(j) \neq \emptyset$, then $G(i) \neq \emptyset$.*

The elements of the codomain of G are called *granules*. The first condition of the previous definition states that granules in a granularity do not overlap and that their order is the same as their time domain order. The second condition states that the subset of the index set that maps to nonempty granules forms an initial segment.

In [4] the authors propose *a labeled linear time structure* in order to define possibly infinite sets of granularities by means of suitable linear-time formulae in the propositional temporal linear logic (PPLTL). Let $\mathcal{G} = \{G_1, G_2, \ldots, G_n\}$ be a finite set of granularities (*calendar*) and let $\mathcal{P}_{\mathcal{G}} = \{P_{G_i}, Q_{G_i} | 1 \leq i \leq n\}$ be a set of propositional symbols associated with the calendar $\mathcal{G}$. Given an alphabet of propositional symbols $\mathcal{P} \supseteq \mathcal{P}_{\mathcal{G}}$, the $\mathcal{P}$-labeled (discrete) *linear-time structure* has the form $(\mathbb{N}, \leq, V)$ [5], where $(\mathbb{N}, \leq)$ is the set of natural numbers with the usual ordering modeling the time domain, and $V : \mathbb{N} \rightarrow 2^{\mathcal{P}}$ is a labeling function mapping the natural number set to sets of propositional symbols. The idea is to identify the starting point and the ending point of an arbitrary granule of the granularity G in the structure by means of the propositional symbols P_G and Q_G, respectively.

3.2 Representing relations on the time domain

Let $\mathcal{A}$ be a finite set of (names of atemporal) attributes $\mathcal{A} = \{A_1, A_2, \ldots\}$ and $\mathcal{D}$ a finite set of domains each of them containing atomic values for the attributes; the domain D_j contains the possible values for the attribute A_j. A *relation schema* is defined by a relation name and a set of attributes; in general, $R(A_1, A_2, \ldots, A_n)$ describes the schema of a relation with name R and set of attributes $A_1, A_2, \ldots, A_n$. A *relation* (or relation state or instance) r of the relation schema $R(A_1, A_2, \ldots, A_n)$, is a finite set of tuples of the form $\langle v_1, v_2, \ldots, v_n \rangle$, where each value v_j, $1 \leq j \leq n$, is an element of D_j.

For sake of simplicity, in the following we assume to describe temporal trends on a single relation with name R and defined on a set U of attributes ($U \subseteq \mathcal{A}$).

In order to represent the history of a relation over the labeled structure, *i.e.,* a finite sequence of relation states, we use a set of symbols which describe, at each time point, the set of valid tuples. Given a relation with schema $R(U)$, where the attribute set $U = \{A_1, A_2, \ldots, A_n\}$, if a tuple $t = \langle v_1, v_2, \ldots, v_n \rangle$ (where v_j is the value of the attribute A_j) is valid at day i, then it is represented by associating to the time point i of the time domain the tuple $R(v_1, v_2, \ldots, v_n)$. $\mathcal{L}$ is the extended alphabet adopted in our framework that contains the propositional symbols to denote the starting and ending points of the granules and symbols which describe the valid tuples of temporal relations. The *$\mathcal{L}$-labeled (discrete) linear-time structure* $\mathcal{M}$ has the form $(\mathbb{N}, \leq, L)$, where $L : \mathbb{N} \rightarrow 2^{\mathcal{L}}$ is a labeling function, extending the previously mentioned function V, which maps natural numbers to sets of symbols denoting either granules or valid tuples. If $i \in \mathbb{N}$, then the set of objects which can be associated to any time point i is composed by propositional symbols for granularities and tuples, as in following:

$$L(i) = \begin{cases} P_G, Q_G & \text{where G is a defined granularity} \\ R(v_1, \ldots, v_n) & \text{where } R \text{ is the name of a relation,} \\ & v_j \text{ is the value associated to the} \\ & j\text{-}th \text{ attribute, } j \in [1, n] \end{cases}$$

As we are dealing with finite databases, we assume that after a time point $n \in \mathbb{N}$, $L(i)$ can only return propositional symbols for granularities.

3.3 The proposed logic

Granular trends can be easily expressed by a formula in the temporal linear logic, suitably extended to deal with attribute values and with algebraic and comparison operators, adopting a notation analogous to that of the attribute-oriented (domain) relational calculus [10].

1530-1311/06 $20.00 © 2006 IEEE

The proposed logic $\mathcal{TL}og$ is a *temporal logic* where expressions are built up from *variables*, used to denote the values of the attributes of the considered tuples.

Notation and Symbols.

We assume to have:

- a finite set $\mathcal{V}$ of *variables* which refer to attributes. We define a function $var : U \to \mathcal{V}$ as a map that associates a variable to an attribute name;
- a set of *propositional symbols* $\mathcal{P}_{\mathcal{G}}$ associated to the calendar $\mathcal{G}$;
- a finite set $\Theta = \{<, =, >, \leq, \geq, \neq\}$ of *comparison operators* and a finite set $\mathcal{OP} = \{+, -\}$ of *algebraic operators*. Note that we assume that each domain D_j is provided with algebraic and comparison operators;
- *boolean operators*: $\wedge, \neg$;
- *quantifiers over the variables*: $\forall$ and $\exists$ for any variable $x \in \mathcal{V}$;
- *temporal operators*: $\mathbf{X}$ (next), $\mathbf{U}$ (until), $\mathbf{X^{-1}}$ (prec) and $\mathbf{S}$ (since). Operators $\mathbf{F}$ (sometime), $\mathbf{G}$ (always), $\mathbf{P}$ (sometime in the past), and $\mathbf{H}$ (always in the past) as standard abbreviations [5];
- *valuation*: if x is a variable associated to the attribute A_j, then the valuation $\varphi(x)$ returns a value in the domain D_j.

Syntax and semantics of $\mathcal{TL}og$ come straightforwardly from the standard FOTL syntax and semantics [5], to consider variables for relation attributes.

Definition 2. **($\mathcal{TL}$og formulae)** *The set of formulae is inductively defined as follows:*

$$\begin{aligned} F \quad := \quad & p \mid R(x, y, \ldots, w, z) \mid x\theta y \mid x\theta\ y\ op\ z \mid \\ & x\theta\ y\ op\ c \mid F \wedge F \mid \neg F \mid \mathbf{X}F \mid F\mathbf{U}F \mid \\ & \mathbf{X^{-1}}F \mid F\mathbf{S}F \mid \forall x.F \mid \exists x.F \end{aligned}$$

where $x, y, \ldots, w, z \in \mathcal{V}$, $p \in \mathcal{P}_{\mathcal{G}}$, $\theta \in \Theta$ is a comparison operator, op $\in \{+, -\}$, and c is a constant value.

The semantics of formulae is defined using the usual satisfaction relation $\models$ that links temporal logical formulae with labeled structure $\mathcal{M} = (\mathbb{N}, \leq, L)$, valuation φ, and the evaluation point i. In particular, the formula $R(x, y, \ldots, w, z)$ is satisfied by the structure $\mathcal{M}$ at time i, with respect to the valuation φ, if $R(\varphi(x), \varphi(y), \ldots, \varphi(w), \varphi(z)) \in L(i)$, *i.e.,* it is evaluated on a tuple of the relation R valid at time i.

Besides the notation $R(x, y, \ldots, w, z)$, in the following we will use a concise notation to define general tuples of variables. Given $X \subseteq U$, with $X = \{A_i, A_j, A_k\}$, $i < j < k$, we denote with $\overline{X}$ variables associated to attributes X: that is, $\overline{X} = \{var(A_i), var(A_j), var(A_k)\}$. Moreover, with respect to the schema $R(U) = R(A_1, \ldots, A_n)$, we will use the concise notation $R(\overline{X})$ to represent the tuple of variables $R(-, var(A_i), -, var(A_j), -, var(A_k), -)$, where "-" is a shorthand to denote distinct variable names.

We extend this notation to the subsets $X, Z \subseteq U$. $R(\overline{X}, \overline{Z})$ is the tuple of variables where variable names associated to the attributes X and Z are defined as $\overline{X}$ and $\overline{Z}$, respectively. As an example, given two tuples of variables $R(\overline{X}, \overline{Z})$ and $R(\overline{X}, \overline{Z_1})$, the usage of the same variable name $\overline{X}$ in both of them implies that they have to be evaluated on tuples of r assuming the same values for the attributes X.

4 Expressing trend dependencies in $\mathcal{TL}$og

In [11], Wijsen proposes a generalization of the concept of temporal functional dependencies by comparing time-varying attributes with any operator in the set **OP** that is the power set of $\{<, >, =\}$. Time is represented by the set of natural numbers $\mathbb{N}$ and the concept of *time accessibility relation* (TAR) is used to indicate which time points have to be grouped one with each other. Time granularities are modeled by a restricted class of TARs, called *chronologies*. A temporal relation is viewed as an infinite sequence $I = \langle I_1, I_2, \ldots \rangle$ of conventional timeslice relations, all over the same set of attributes; I_i is the set of tuples valid at time i and it is always empty after some $n \in \mathbb{N}$. Wijsen assumes the existence of a totally ordered infinite set of constants (**dom**, $\leq$) containing the values for attributes. The notion of *directed attribute set* (DAS) over a set of attributes, allows one to describe the attributes which have to be compared and the associated comparing operators. We now report the formal definition of these concepts [11].

Definition 3. *Let U be a set of attributes. A direct attribute set (DAS) over U is a total function from U to **OP**. Let Φ be a DAS and $\theta_1, \ldots, \theta_n \in$ **OP**; the domain of Φ is denoted as atts(Φ). That is:*

$$\begin{aligned} & \textit{atts}(\{(A_1, \theta_1), (A_2, \theta_2), \ldots, (A_n, \theta_n)\}) \\ = \quad & \{A_1, A_2, \ldots, A_n\} \end{aligned}$$

Let s, t be tuples over U. Let Φ be a DAS over some subset X of U. The tuple pair (s, t) satisfies Φ, denoted $\Phi^*(s, t)$, if and only if, $\forall A \in X$, $s(A)\theta_A t(A)$, where θ_A is a shorthand for $\Phi(A)$.

TDs are defined as in the following [11].

Definition 4. *A trend dependency (TD) over U is a statement $\Phi \to_\alpha \Psi$, where α is a TAR and Φ, Ψ are DASs with* *atts*(Φ), *atts*$(\Psi) \subseteq U$. *Let σ denote the TD $\Phi \rightarrow_\alpha \Psi$; then Φ is said the left-hand DAS of σ, Ψ is said the righ-hand DAS, and the TD σ involves the TAR α. The notation $tar(\sigma)$ is used to denote the TAR involved in σ. Let $I = \langle I_1, I_2, \ldots \rangle$ be a temporal relation and let $\Phi \rightarrow_\alpha \Psi$ be a TD (all over U). The TD $\Phi \rightarrow_\alpha \Psi$ is satisfied by I if and only if for every $(i, j) \in \alpha$, for every $s \in I_i$, for every $t \in I_j$, if $\Phi^*(s,t)$, then $\Psi^*(s,t)$.*

4.1 General logical formulae for TDs

In the following, we restrict our attention to the subset of all the possible TDs $\Phi \rightarrow_\alpha \Psi$ where the TAR α is a chronology (*i.e.,* a time granularity). Intuitively, we can say that a TD in the form $\Phi \rightarrow_\alpha \Psi$ can be translated into a logical formula in the form $\mathbf{G}(\Delta \rightarrow \Gamma)$ where the premise Δ has to characterize the tuples to compare, satisfying the TAR α, and the attributes of the considered tuples satisfying the relationships imposed by the DAS ϕ. The consequent Γ describes the relationships that must hold between the attributes referred by the DAS ψ. The following theorem proves the correctness of the logical formulae associated to the special kinds of TDs we are considering. We assume that the (atemporal part of the) temporal relation $I = \langle I_1, I_2, \ldots \rangle$ (according to the notation of Wijsen) is represented in our logical formalism by means of a relation having name R, and the chronology α (*i.e.,* the time granularity) is described by means of a suitable formula γ_α on the labeled linear-time structure $\mathcal{M}_I$, *i.e.,* the structure representing the temporal relation I in our logic $\mathcal{TL}og$.

Theorem 1. *Let $I = \langle I_1, I_2, \ldots \rangle$ be a temporal relation on the set $U = \{A_1, \ldots, A_n\}$ of atemporal attributes. The TD $\Phi \rightarrow_\alpha \Psi$, with α chronology, is satisfied by I if and only if the following formula is satisfied by the labeled structure $\mathcal{M}_I$:*

$$\begin{aligned}&\forall v_i, v_i', i \in [1, n]\\ &\gamma_\alpha \wedge \mathbf{G}((P_\alpha \wedge \neg Q_\alpha \mathbf{U}(R(v_1, \ldots, v_n) \wedge\\ &(\neg Q_\alpha \mathbf{U}(R(v_1', \ldots, v_n') \wedge \bigwedge_{A_p \in atts(\Phi)} (v_p \, \theta_{A_p} \, v_p')))))\\ &\rightarrow \bigwedge_{A_q \in atts(\Psi)} (v_q \, \theta_{A_q} \, v_q'))\end{aligned}$$

Proof sketch.
According to the Definition 4, a temporal relation I satisfies the TD $\Phi \rightarrow_\alpha \Psi$ if and only if $\forall (i, j) \in \alpha$ and $\forall s \in I_i, \forall t \in I_j$, the following implication holds: $\Phi^*(s,t) \rightarrow \Psi^*(s,t)$. From the particular structure of the considered TAR α, and according to the meaning of the notations $\Phi^*(s,t)$ and $\Psi^*(s,t)$, this means that, $\forall (i,j) \in \alpha$ and $\forall s \in I_i, \forall t \in I_j$, the following implication holds: $\forall A \in atts(\Phi) \; (s[A] \; \theta_A \; t[A]) \rightarrow \forall B \in atts(\Psi) \; (s[B] \; \theta_B \; t[B])$. According to the semantics of the proposed logical formula, the structure $\mathcal{M}_I$ satisfies the formula if and only if, for each couple of tuples s (represented by $R(v_1, \ldots, v_n)$) and t (represented by $R(v_1', \ldots, v_n')$) belonging to the same granule of α, if they satisfy the conditions on the attributes $atts(\Phi)$, then they satisfy also the conditions on the attributes $atts(\Psi)$. The tuples s and t belong to the same granule of α if there exists an index k such that $\alpha(k) \neq \emptyset$, $\alpha(k) = [s_k, e_k]$, with $s_k, e_k \in \mathbb{N}$, and $s \in I_i$, $t \in I_j$, with $s_k \leq i \leq j \leq e_k$. The k-th granule of the granularity α is represented in the structure $\mathcal{M}_I$ by means of the propositional symbols $P_\alpha \in L(s_k)$, $Q_\alpha \in L(e_k)$ and by the condition $P_\alpha, Q_\alpha \notin L(s_h)$ with $s_k < s_h < e_k$. Thus, by definition, the previous conditions holds if and only if I satisfies the TD $\Phi \rightarrow_\alpha \Psi$. $\square$

5 Characterizing granular trends

In our framework, we aim at characterizing and describing specific temporal features possibly related to temporal trends with granularities, such as trends over two or more consecutive granules, or trends lasting all over a granule from trends lasting a limited span within a granule.

5.1 Motivating Example

Let us consider the need of representing granular trends in the context of clinical medicine. Therapy plans deal with the definition of specific granularities according to which it is possible to properly observe the evolution of the patient's state. In the case of chemotherapies for oncological patients, oncology patients undergo several chemotherapy cycles: each cycle can include the administration of several drugs, which the patient has to assume according to a specific temporal pattern. As an example, consider the following chemotherapy recommendation for the treatment of the node-positive breast cancer [7]:

"The recommended CEF regimen consists of 14 days of oral cycloshosphamide and intravenous injection of epirubicin and 5-fluorouracil on days 1 and 8. This is repeated every 28 days for 6 cycles."

According to this scenario, it is possible to identify some definition of time granularities. Let us assume that OC (cyclophosphamide), IE (intravenous epirubicin) and FI (5-fluorouracil) are the granularities corresponding to drugs of the CEF regimen, which corresponds to the overall granularity CEF. Let us now suppose to have a temporal relation *Patient* describing, for each patient, the type of therapy, the patient's identifier, the value of systolic and diastolic blood pressures, the weight, and the assumed drugs. Figure 1 shows the labeled linear-time structure representing the instance of the temporal relation *Patient* related to

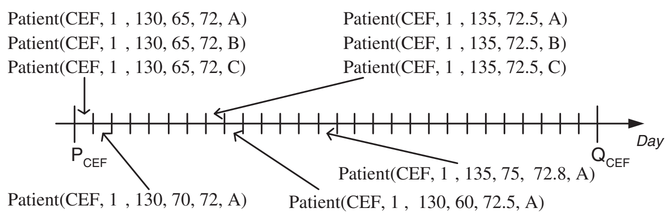

Figure 1. The labeled linear-time structure describing the instance of the relation $Patient$ and the first granule of the granularity CEF.

the first cycle of treatment for the patient with identifier 1. For sake of space we represent only some tuples of the relation. We can note that the symbols reported on the time domain[1] are the propositional symbols for the starting and the ending points of the first granule of CEF (P_{CEF} and Q_{CEF} respectively) and symbols denoting valid tuples of the relation *Patient*.

5.2 Simple trends

As mentioned before, it is interesting to describe temporal properties or temporal trends having specific time features. In the following we consider *granular trends*, *i.e.,* trends defined according to specific time granularities. Simple examples of trends, common in the database area [11], in the artificial intelligence area [9], and also in clinical decision support systems [3, 9] are:

- *Increase(Relation, KeySet, Attribute, Rate)*, *i.e.,* the value of the attribute *Attribute* of the tuples of the relation *Relation*, assuming the same values for the attributes *KeySet*, increases over the time and the difference between a value and the next value (for the attribute *Attribute*) is greater than or equal to the value *Rate*.
- *Decrease(Relation, KeySet, Attribute, Rate)*, *i.e.,* the value of the attribute *Attribute* of the tuples of the relation *Relation*, assuming the same values for the attributes *KeySet*, decreases over the time and the difference between a value and the next value (for the attribute *Attribute*) is smaller than the value *Rate*.
- *State(Relation, KeySet, Attribute, Constant)*, *i.e.,* the value of the attribute *Attribute* of the tuples of the relation *Relation*, assuming the same values for the attributes *KeySet* is equal to *Constant* for a given interval.
- *Stationary(Relation, KeySet, Attribute, Threshold)*, *i.e.,* the value of the attribute *Attribute* of the tuples of the relation *Relation*, assuming the same values for the attributes *KeySet*, over the time is inside a range such that the module of the difference between a value and the next one (for the attribute *Attribute*) is not greater than a given threshold value *Threshold*.

[1] We assume that the bottom granularity used for the time domain is the granularity of day.

In the context of temporal databases with multiple granularities, the description of temporal trends which take into account different granularities should be considered [11]. For example, with respect to the temporal relation *Patient*, shown in Figure 1, it is possible to require that: (i) the weight of a patient can change during a cycle of chemotherapy but the difference between two consecutive values of patient's weight does not exceed a given threshold value, and (ii) the values of blood pressures can increase during a chemotherapy cycle. In our work we propose a formalism for the description of granular trends related to tuples associated to time points. For sake of simplicity, we do not consider tuples associated to granules and focus explicitly on tuples related to single time points; thus, we do not consider tuples containing aggregate values (such as average) over temporal granules.

5.3 A characterization of temporal trends

When we describe temporal trends which involve specific temporal granularities, we can distinguish two kinds of trend, called *intragranule trend* and *intergranule trend*, respectively. In the first case, temporal properties expressed by means of a trend have to be satisfied inside a given granule while in the second one the properties involve different granules. Both for intragranule and intergranule trends, we can discriminate two kinds of satisfiability, called *local* and *global*, respectively. The *local* satisfiability allows one to express that a trend must be valid over some time points of the considered granule(s), while a global trend must be valid over all the time points of the considered granule(s). Moreover, both for local and global satisfiability, it is possible to specify *existential* or *universal* temporal quantification with respect to the time domain: the temporal operator **F** is used in order to express an existential temporal quantification, while the temporal operator **G** is used to describe an universal one. The proposed formalism allows one to express some metric properties related to a trend such as the concept of *duration*. The *duration* specifies the temporal length of the required properties; in the case of intragranule trend, the duration describes the *number of samples* composing the required trend inside a specific granule, while in the case of intergranule trend the concept of duration represents the number of distinct granules during which the specific trend has to be satisfied. Other temporal features related to a trend are the number of *exceptions* in a sequence of tuples, and the temporal window (*granule(s) frame*) inside a granule or within several granules during which the trend is satisfied. The concept of exception allows one to express the maximum number of allowed violations to the required trend. The granule(s) frame specifies that the required trend must be verified in a given temporal window inside a granule or in a given set of contiguous granules. Moreover, it is possible to combine simple temporal trends in order to obtain complex ones. In the following, we propose general formulae for the description of granular trends on a temporal relation r, represented by means of a suitable labeled structure $\mathcal{M}_r$. Granular trends are related to the value assumed by the attribute $Y \in U$ and involve tuples of r assuming the same values on the attributes X (*i.e.,* $X \subseteq U$ is the $KeySet$). In particular, we focus here on the characterization of intragranule/intergranule trends, on local and global satisfiability, and on quantification on granules. We will explain the meaning of these features by considering some examples related to the relation *Patient*.

5.3.1 Intragranule and Intergranule trends.

An *intragranule trend* allows one to express a relationship between tuples which are valid in the same granule. This kind of trend could require that the given property has to be satisfied in all time points of a specific granule or in some ones. The general formula that allows one to check if there exists an intragranule increasing trend, with rate $Rate$, for the attribute Y between two consecutive tuples of the relation r (valid inside the same granule of the granularity ω) assuming the same values on the X is the following:

$$\begin{aligned}&\exists \overline{X}, \overline{Y}, \overline{Y_1}.\\ &\mathbf{F}(P_\omega \ \wedge\ \neg Q_\omega \mathbf{U}(R(\overline{X}, \overline{Y}) \ \wedge\ \neg Q_\omega \mathbf{U} R(\overline{X}, \overline{Y_1}) \ \wedge\\ &\quad (\overline{Y_1} \geq \overline{Y} + \mathit{Rate})))\end{aligned}$$

In the case we want to describe a decreasing trend, a state trend, or a stationary trend with the same features of the previous proposed trend, the general formula describing it admits the same construction with the only one difference that $(\overline{Y_1} \leq \overline{Y} + \mathit{Rate})$, is replaced by the suitable expressions corresponding to the considered trend.

Example 1. *We want to verify whether a patient exists having a decrease trend of the parameter SBP during a cycle of chemotherapy. In order to describe the required temporal condition for the property, i.e., the increasing during a specific granule of the* CEF *granularity, we can use the following formula:*

$$\begin{aligned}&\exists Id, Ch, \ldots\\ &\mathbf{F}(P_{CEF} \ \wedge\ \neg Q_{CEF} \mathbf{U}(Patient(Ch, Id, Sbp, Dbp, W, D)\\ &\quad \wedge \neg Q_{CEF} \mathbf{U}(Patient(Ch, Id, Sbp_1, Dbp_1, W_1, D_1))\\ &\quad \wedge (\mathit{Sbp}_1 < \mathit{Sbp} + \mathit{Rate})))\end{aligned}$$

The previous formula is satisfied when *exists at least* one decrease trend between two consecutive values of the SBP parameter, related to the same patient. The first tuple ($Patient(Ch, Id, Sbp, Dbp, W, D)$) is valid at some time i of some granule of CEF, the second one ($Patient(Ch, Id, Sbp_1, Dbp_1, W_1, D_1)$) is valid at time j ($j > i$) and they refer to the same patient and the same chemotherapy (*i.e.,* same variables Id and Ch). These tuples are valid during the same granule of CEF because they must be after P_{CEF} and before Q_{CEF}. Figure 2 reports the diagram related to the values of the SBP parameter during the first cycle of the treatment CEF for the patient with identifier 1 (cf. Figure 1). The broken lines indicate possible values of the parameter that are not reported in the corresponding instance (cf. Figure 1). As highlighted in the figure, the values of SBP associated to the patient 1 satisfy the previous formula.

An *intergranule trend* allows one to express a relationship between tuples which are valid in distinct granules (of the same granularity). Also in this case it is possible to require that the property is valid either in several consecutive granules or for all granules. The general formula that allows one to check if there exists an intergranule increasing trend, with rate $Rate$, for the attribute Y ($Y \in U$) between two tuples assuming the same values on the $KeySet\ X$ ($X \subseteq U$), where the first one is valid during the granule i and the second one is valid during the granule $i + n$ of the granularity ω, is the following one:

$$\begin{aligned}&\exists \overline{X}, \overline{Y}, \overline{Y_1}.\\ &\mathbf{F}(R(\overline{X}, \overline{Y}) \ \wedge\ \mathbf{X}^{\mathbf{n}}_\omega(R(\overline{X}, \overline{Y_1})) \ \wedge\ (\overline{Y_1} \geq \overline{Y} + \mathit{Rate}))\end{aligned}$$

The formula $\mathbf{X^n}_\omega(f)$ is the logical formula which states that, in the $n\text{-}th$ granule (w.r.t. the current time point) of the granularity ω, the formula f holds at least in one time point.

Example 2. *We want to verify whether a stationary trend exists for the weight of a patient between two consecutive cycles of chemotherapy with a specific threshold value ϵ. In order to describe the required temporal condition for the property, i.e., the stationary trend between two consecutive cycles of chemotherapy, we can use the following formula:*

$$\begin{aligned}&\exists Id, Ch, \ldots\\ &\mathbf{F}(Patient(Ch, Id, Sbp, Dbp, W, D) \ \wedge\\ &\quad \mathbf{X}^{\mathbf{1}}_{\mathbf{CEF}}(Patient(Ch, Id, Sbp_1, Dbp_1, W_1, D_1)) \ \wedge\\ &\quad ((W_1 > W \ \wedge (W + \epsilon) \geq W_1) \ \vee\\ &\quad (W \geq W_1 \ \wedge\ (W_1 + \epsilon) \geq W)))\end{aligned}$$

The formula is satisfied when exists a stationary trend between the values of the parameter *Weight* associated to two tuples t_1 and t_2 (related to the same patient and the same chemotherapy), where the first is valid in the *k*-th granule of the granularity CEF and the second is valid in the *k+1*-th one, as described by means of the formula $\mathbf{X}^{\mathbf{1}}_{\mathbf{CEF}}(\ldots)$. Figure 3 reports the diagram related to the (possible) values of the $Weight$ parameter during two consecutive cycles of the CEF treatment for the patient with identifier 1. In the case that the threshold ϵ is equal to 0.5, then the part highlighted in the figure represents all the possible couples of samples satisfying the stationary trend for the weight.

5.3.2 Local and global satisfiability.

From a trend validity point of view, we can distinguish two kinds of satisfiability, called *local* and *global*. A *local* satisfiability allows one to describe the validity of a (intragranule or intergranule) trend only for some time points in the considered granule(s), while the *global* satisfiability allows one to require that the trend must be valid all over the time points of the considered granule(s). The previous general formulae for intragranule or intergranule trends consider the case of local satisfiability; we now introduce the formulae that allow one to characterize the global satisfiability in the case of intragranule and intergranule trends. For sake of simplicity, we consider the increasing trend but the same formulae can be suitably adapted to the other kinds of trends. The above logical formula describing the intragranule increasing trend (for the parameter Y) that is valid, during each time point inside a granule of the granularity ω, between a tuple and its consecutive one (assuming the same values on the $KeySet\ X$) is the following:

$$\begin{aligned}\exists\overline{X}.\quad & \mathbf{F}(P_\omega \ \wedge\ \neg(\neg Q_\omega\ \mathbf{U}(\exists\overline{Y}.R(\overline{X},\overline{Y})\ \wedge\\ & \quad \neg Q_\omega \mathbf{U}(\exists\overline{Y_1}.R(\overline{X},\overline{Y_1}))\ \wedge\ (\overline{Y_1} < \overline{Y} + Rate)))\ \wedge\\ & \quad \neg Q_\omega\ \mathbf{U}(\exists\overline{Y_2}.R(\overline{X},\overline{Y_2})))\end{aligned}$$

The logical formula we propose for the satisfaction of an intragranule and global trend it is satisfied whether an increase trend for the value of the parameter Y, during a whole granule of the granularity ω exists, *i.e.,* it is not possible to find a counterexample of the required trend.

Example 3. *We want to verify whether there is a patient having, for a given treatment, an increase trend of the parameter SBP during a whole granule of the granularity CEF.*

$$\begin{aligned}& \exists Id, Ch.\\ & \mathbf{F}(P_{CEF}\ \wedge\ \neg(\neg Q_{CEF}\ \mathbf{U}(\exists Bg, Sbp, Dbp, W, D.\\ & \quad Patient(Ch, Id, Bg, Sbp, Dbp, W, D)\ \wedge\\ & \quad \neg Q_{CEF}\mathbf{U}(\exists Bg_1, Sbp_1, Dbp_1, W_1, D_1.\\ & \quad Patient(Ch, Id, Bg_1, Sbp_1, Dbp_1, W_1, D_1))\ \wedge\\ & \quad (Sbp_1 < Sbp + Rate)))\ \wedge\\ & \quad \neg Q_{CEF}\ \mathbf{U}(\exists Bg_2, Sbp_2, Dbp_2, W_2, D_2.\\ & \quad Patient(Ch, Id, Bg_2, Sbp_2, Dbp_2, W_2, D_2)))\end{aligned}$$

The logical formula we propose for the satisfaction of an *intergranule and global trend* requires that, for each time point of a granule ω where a valid tuple t exists, for each corresponding tuple s valid in the next granule, the considered trend between t and s (with respect to the values they assume on the attribute Y) is satisfied. In the case the required trend is the increasing, the formula is the following:

$$\begin{aligned}\exists\overline{X}.\quad & \mathbf{F}(P_\omega\ \wedge\ \neg(\neg Q_\omega \mathbf{U}(\exists\overline{Y}.R(\overline{X},\overline{Y})\ \wedge\\ & \quad \mathbf{X}^{\mathbf{1}}_\omega(\exists\overline{Y_1}.R(\overline{X},\overline{Y_1}))\ \wedge\ (\overline{Y_1} < \overline{Y} + Rate)))\ \wedge\\ & \quad \neg Q_\omega \mathbf{U}(\exists\overline{Y_2}.R(\overline{X},\overline{Y_2}))\ \wedge\\ & \quad \mathbf{X}^{\mathbf{1}}_\omega(\neg Q_\omega \mathbf{U}(\exists\overline{Y_3}.R(\overline{X},\overline{Y_3}))))\end{aligned}$$

Thus, for each tuple t valid inside the first granule, the formula checks that no counterexamples exist (*i.e.,* there are no tuples in the next granule that do not satisfy the increasing with respect to the value of t in Y).

5.3.3 Quantifying over the time domain.

Both for local and global satisfiability, it is possible to require an *existential* or *universal* temporal quantification with respect to the time domain. The temporal operator **F** is used in order to express an existential temporal quantification while the temporal operator **G** is used to describe an universal one. In the previous examples, we have shown trends which require an existential temporal quantification: more restrictive properties can require that the trends involving a given granularity ω are always valid and, for this reason, the temporal operator **G** is used instead of **F**.

Example 4. *We want to check whether there is a patient having an increase trend for the parameter SBP which is valid in the* whole *granule and* for each *granule of CEF granularity.*

$$\begin{aligned}\exists Id, Ch.\quad & \mathbf{G}(P_{CEF} \rightarrow (\neg(\neg Q_{CEF}\mathbf{U}(\exists Bg, Sbp, Dbp, W, D.\\ & \quad Patient(Ch, Id, Bg, Sbp, Dbp, W, D)\ \wedge\\ & \quad \neg Q_{CEF}\mathbf{U}(\exists Bg_1, Sbp_1, Dbp_1, W_1, D_1.\\ & \quad Patient(Ch, Id, Bg_1, Sbp_1, Dbp_1, W_1, D_1))\ \wedge\\ & \quad (Sbp_1 < Sbp + Rate)))\ \wedge\\ & \quad \neg Q_{CEF}\ \mathbf{U}(\exists Bg_2, Sbp_2, Dbp_2, W_2, D_2.\\ & \quad Patient(Ch, Id, Bg_2, Sbp_2, Dbp_2, W_2, D_2))))\end{aligned}$$

The formula checks whether inside each granule of the granularity CEF, the parameter SBP increases in the whole granule. The usage of the temporal operator **G** makes sure that the trend is valid in each granule of the CEF granularity, when data related to the considered patient exist.

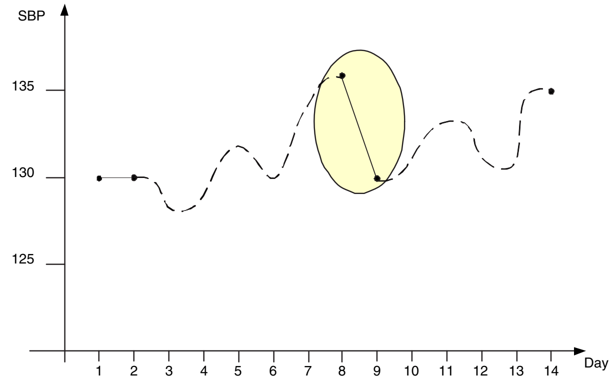

Figure 2. Values assumed by the parameter SBP during the first cycle of the CEF treatment for the patient 1.

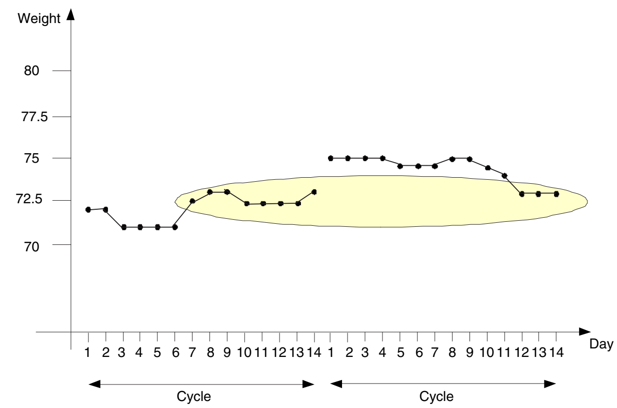

Figure 3. Values assumed by the parameter *Weight* during two consecutive cycles of the CEF treatment for the patient 1.

6 Discussion and Conclusions

We have proposed a general framework for expressing trends over different granularities, highlighting different orthogonal features, as local and global satisfiability, intragranule and intergranule trends, temporal quantification, and so on. Different kinds of temporal trends which involve multiple granularities can be represented in this model with suitable formulae in the context of a temporal linear logic. From a computability point of view, in this work we have considered a labeled linear-time structure on which first order logical formulae are evaluated. Moreover, we have assumed to deal with discrete sets of numerical values for the trend attributes. According to the decidability results of the Presburger arithmetic, which may be doubly exponential [6], we have considered as operation for our numerical domains only the addition (+); the difference between two numerical values is defined by using the addition operation. As we are interested in the model checking problem on a finite database, the logical fragment we have defined is decidable for model checking, even though the complexity, without assuming the usage of index structures, may be quadratic in the dimension of the database. As for future work, we aim to complete the proposed framework with model checking algorithms for evaluating whether a given trend is satisfied on a database. Moreover, an interesting research direction is to design temporal data mining techniques taking into account the framework we introduced for trends.

References

[1] C. Bettini, S. Jajodia, J. Lin, and X.S. Wang. Discovering frequent event patterns with multiple granularities in time sequences. *IEEE TKDE*, 10(2):222–237, 1998.

[2] C. Bettini, S. Jajodia, and X.S. Wang. *Time granularities in Databases, Data Mining, and Temporal Reasoning*. Springer, 2000.

[3] C. Combi and L. Chittaro. Abstraction on clinical data sequences: an object-oriented data model and a query language based on the event calculus. *Artificial Intelligence in Medicine*, 17(3):271–301, 1999.

[4] C. Combi, M. Franceschet, and A. Peron. Representing and reasoning about temporal granularities. *Journal of Logic and Computation*, 14(1):51–77, 2004.

[5] E.A. Emerson. Temporal and modal logic. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume B, chapter 16, pages 995–1072. MIT Press, 1990.

[6] M.J. Fischer and M.O.Rabin. Super-exponential complexity of presburger arithmetic. In *SIAM-AMS Symposium in Applied Mathematics*, volume 7, pages 27–41, 1974.

[7] M. Levine, C. Sawka, and D. Bowman. Clinical practice guidelines for the care and treatment of breast cancer: 8. adjuvant systemic therapy for women with node-positive breast cancer (2001 update). *Canadian Medical Association journal*, page 164, 2001.

[8] R. Sadri, C. Zaniolo, A. M. Zarkesh, and J. Adibi. Optimization of sequence queries in database systems. In *Proceedings of the Twentieth ACM Symposium on Principles of Database Systems*, pages 73–83. ACM, 2001.

[9] Yuval Shahar. A framework for knowledge-based temporal abstraction. *Artificial Intelligence*, 90(1-2):79–133, 1997.

[10] J.D. Ullman. *Principles of Database and Knowledge-Base Systems*. Computer Science Press, 1988.

[11] J. Wijsen. Trends in databases: Reasoning and mining. *IEEE TKDE*, 13(3):426–438, 2001.

Proceedings of the Thirteenth International Symposium on Temporal Representation and Reasoning (TIME'06)
1530-1311/06 $20.00 © 2006 IEEE