
Time Granularity in Temporal Data Mining

Paul Cotofrei¹ and Kilian Stoffel²

¹ Information Management Institute, University of Neuchâtel, Pierre-à-Mazel, 7, 2000, Neuchâtel, Switzerland, paul.cotofrei@unine.ch

² Information Management Institute, University of Neuchâtel, Pierre-à-Mazel, 7, 2000, Neuchâtel, Switzerland, kilian.stoffel@unine.ch

Summary. In this chapter, a formalism for a specific temporal data mining task (the discovery of rules, inferred from databases of events having a temporal dimension), is defined. The proposed theoretical framework, based on first-order temporal logic, allows the definition of the main notions (*event*, *temporal rule*, *confidence*) in a formal way. This formalism is then extended to include the notion of temporal granularity and a detailed study is made to investigate the formal relationships between the support measures of the same event in linear time structures with different granularities. Finally, based on the concept of *consistency*, a strong result concerning the independence of the confidence measure for a temporal rule, over the worlds with different granularities, is proved.

1 Introduction

The domain of temporal data mining focuses on the discovery of causal relationships among events that are ordered in time and may be causally related. The contributions in this domain encompass the discovery of temporal rule, of sequences and of patterns. However, in many respects this is just a terminological heterogeneity among researchers that are, nevertheless, addressing the same problem, albeit from different starting points and domains.

Although there is a rich bibliography concerning the formalism for temporal databases, there are very few articles on this topic for temporal data mining. In [1, 5, 21] general frameworks for temporal mining are proposed, but usually the researches on causal and temporal rules are more concentrated on the methodological/algorithmic aspect, and less on the theoretical aspect. Based on a methodology for temporal rule extraction, described in [9], we proposed in [10, 11] an innovative formalism based on first-order temporal logic, which permits an abstract view on temporal rules. An important concept defined in this formalism is the property of *consistency*, which guarantees the preserving over time of the confidence/support of a temporal rule. The formalism is developed around a time model for which the events are those that describe system evolution (event-based temporal logic). Each formula

expresses what the system does at each event, events are referring to other events, and so on: this results in specifying relationships of precedence and cause-effect among events. But the real systems are systems whose components (events) have dynamic behavior regulated by very different – even by orders of magnitude – time granularities. Analyzing such systems (hereinafter granular systems) means to approach theories, methodologies, techniques and tools that make use of granules (or groups, classes, clusters of a universe) in the process of problem solving. Granular computing (the label which covers this approach) is a way of thinking that relies on our ability to perceive the real world under various grain sizes, to abstract and to consider only those things that serve our present interest, and to switch among different granularities. By focusing on different levels of granularities, one can obtain various levels of knowledge, as well as inherent knowledge structure. Granular computing is essential to human problem solving, and hence has a very significant impact on the design and implementation of intelligent systems [28, 27, 29, 20].

2 State of Art

The notions of granularity and abstraction are used in many subfields of artificial intelligence. The granulation of time and space leads naturally to temporal and spatial granularities. They play an important role in temporal and spatial reasoning [13, 18, 26]. Based on granularity and abstraction, many authors studied some fundamental topics of artificial intelligence, such as knowledge representation [30], theorem proving [15], search [31], planning [19], natural language understanding [22], intelligent tutoring systems [23], machine learning [25], and data mining [16].

Despite the widespread recognition of its relevance in the fields of formal specifications, knowledge representation and temporal databases, there is a lack of a systematic framework for time granularity. Hobbs [17] proposed a formal characterization of the general notion of granularity, but gives no special attention to time granularity. Clifford et al. [8] provide a set-theoretic formalization of time granularity, but they do not attempt to relate the truth value of assertions to time granularity. Extensions to existing languages for formal specifications, knowledge representation and temporal databases that support a limited concept of time granularity are proposed in [24, 14, 7]. Finally, Bettini et al. [2, 4] provide a formal framework for expressing data mining tasks involving time granularities, investigate the formal relationships among event structures that have temporal constraints, define the pattern-discovery problem with these structures and study effective algorithms to solve it.

The purpose of this chapter is to extend our formalism to include the concept of time granularity. We define the process by which a given structure of time granules μ (called temporal type) induces a first-order linear time structure M_μ (called granular world) on the basic (or absolute) linear time

structure M . The major change for the temporal logic based on M_μ is at the semantic level: for a formula p , the interpretation does not assign a meaning of truth (one of the values $\{true, false\}$), but a degree of truth (a real value from $[0, 1]$). Consequently, we can give an answer to the following question: if the temporal type μ is *finer than* temporal type ν , what is the relationship between the support of the same temporal rule T_p in the linear time structures M_μ and M_ν . We also study the variation process for the set of satisfiable events (degree of truth equal one) during the transition between two time structures with different granularity. By an extension at the syntactic and semantic level we are able to define an aggregation mechanism for events, reflecting the following intuitive phenomenon: in a coarser world, not all events inherited from a finer world are satisfied, but in exchange there are new events which become satisfiable. Finally, using an extension of the concept of consistency for a granular time structure M_μ , we prove a strong result concerning the invariance of the confidence measure for a temporal rule during the process of information transfer between worlds with different granularities.

The rest of the chapter is structured as follows. In the next section, the first-order temporal logic formalism is extensively described (the main terms and concepts). The definitions and theorems concerning the extension of the formalism towards a temporal granular logic are presented in Sect. 4. Finally, the last section summarizes our work, followed by an appendix containing the proofs of the theorems in the chapter.

3 Formalism of Temporal Rules

Time is ubiquitous in information systems, but the mode of representation/perception varies in function of the purpose of the analysis [6, 12]. Firstly, there is a choice of a *temporal ontology*, which can be based either on *time points* (instants) or on *intervals* (periods). Secondly, time may have a *discrete* or a *continuous* structure. Finally, there is a choice of *linear* vs. *nonlinear* time (e.g. acyclic graph). Our selection, imposed by the discrete representation of all databases, is a temporal domain represented by linearly ordered discrete instants.

Databases being first-order structures, the first-order logic represents a natural formalism for their description. Consequently, the first-order temporal logic expresses the formalism of temporal databases. For the purpose of our approach we consider a restricted first-order temporal language L which contains only constant symbols $\{c, d, ..\}$, n -ary ($n \geq 1$) function symbols $\{f, g, ..\}$, variable symbols $\{y_1, y_2, ...\}$, n -ary predicate symbols ($n \geq 1$, so no proposition symbols), the set of relational symbols $\{=, <, \leq, >, \geq\}$, the logical connective \wedge and a temporal connective of the form ∇_k , $k \in \mathbb{Z}$, where k strictly positive means *after k time instants*, k strictly negative means *before k time instant* and $k = 0$ means *now*.

The syntax of L defines terms, atomic formulae and compound formulae. The *terms* of L are defined inductively by the following rules:

- T1. Each constant is a term.
- T2. Each variable is a term.
- T3. If t_1, t_2, \dots, t_n are terms and f is an n -ary function symbol then $f(t_1, \dots, t_n)$ is a term.

The atomic formulae (or atoms) of L are defined by the following rules:

- A1. If t_1, \dots, t_n are terms and P is an n -ary predicate symbol then $P(t_1, \dots, t_n)$ is an atom.
- A2. If t_1, t_2 are terms and ρ is a relational symbol then $t_1 \rho t_2$ is an atom (also called relational atom).

Finally, the (compound) formulae of L are defined inductively as follow:

- F1. Each atomic formula is a formula.
- F2. If p, q are formulae then $(p \wedge q), \nabla_k p$ are formulae.

A Horn clause is a formula of the form $B_1 \wedge \dots \wedge B_m \rightarrow B_{m+1}$ where each B_i is a positive (non-negated) atom. The atoms $B_i, i = 1, \dots, m$ are called implication clauses, whereas B_{m+1} is known as the implicated clause. Syntactically, we cannot express Horn clauses in our language L because the logical connective \rightarrow is not included. However, to allow the description of rules, which formally look like a Horn clause, we introduce a new logical connective, \mapsto , representing practically a rewrite of the connective \wedge . Therefore, a formula in L of the form $p \mapsto q$ is syntactically equivalent to the formula $p \wedge q$. When and under what conditions we may use the new connective, is explained in the next definitions.

Definition 1. *An event (or temporal atom) is an atom formed by the predicate symbol E followed by a bracketed n -tuple of terms ($n \geq 1$) $E(t_1, t_2, \dots, t_n)$. The first term of the tuple, t_1 , is a constant symbol representing the name of the event and all others terms are expressed according to the rule $t_i = f(t_{i1}, \dots, t_{ik_i})$. A short temporal atom (or the event's head) is the atom $E(t_1)$.*

Definition 2. *A constraint formula for the event $E(t_1, t_2, \dots, t_n)$ is a conjunctive compound formula, $E(t_1, t_2, \dots, t_n) \wedge C_1 \wedge C_2 \wedge \dots \wedge C_k$. Each C_j is a relational atom $t \rho c$, where the first term t is one of the terms $t_i, i = 1 \dots n$ and the second term is a constant symbol.*

For a short temporal atom $E(t_1)$, the only constraint formula that is permitted is $E(t_1) \wedge (t_1 = c)$. We denote such constraint formula as *short constraint formula*.

Definition 3. *A temporal rule is a formula of the form $H_1 \wedge \dots \wedge H_m \mapsto H_{m+1}$, where H_{m+1} is a short constraint formula and $H_i, i = 1..m$ are constraint formulae, prefixed by the temporal connectives $\nabla_{-k}, k \geq 0$. The maximum value of the index k is called the time window of the temporal rule.*

Remark. The reason for which we did not permit the expression of the implication connective in our language is related to the truth table for a formula $p \rightarrow q$: even if p is false, the formula is still true, which is unacceptable for a temporal rationing of the form *cause* \rightarrow *effect*.

If we change in Definition 1 the conditions imposed on the terms $t_i, i = 1 \dots n$, into "each term t_i is a variable symbol", we obtain the definition of a temporal atom template. We denote such a template as $E(y_1, \dots, y_n)$. Following the same rationing, a constraint formula template for $E(y_1, \dots, y_n)$ is defined as a conjunctive compound formula, $C_1 \wedge C_2 \wedge \dots \wedge C_k$, where the first term of each relational atom C_j is one of the variables $y_i, i = 1 \dots n$. Consequently, a short constraint formula template is the relational atom $y_1 = c$. Finally, by replacing in Definition 3 the notion "constraint formula" with "constraint formula template" we obtain the definition of a temporal rule template. Practically, the only formulae constructed in L are temporal atoms, constraint formulae, temporal rules and the corresponding templates.

The semantics of L is provided by an interpretation I over a domain D (in our formalism, D is always a linearly ordered domain). The interpretation assigns an appropriate meaning over D to the (non-logical) symbols of L. Usually, the domain D is imposed during the discretisation phase, which is a pre-processing phase used in almost all knowledge extraction methodologies. Based on Definition 1, an event can be seen as a labelled (constant symbol t_1) sequence of points extracted from raw data and characterized by a finite set of features (terms t_2, \dots, t_n). Consequently, the domain D is the union $D_e \cup D_f$, where the set D_e contains all the strings used as event names and the set D_f represents the union of all domains corresponding to chosen features.

Example 1. Consider a database containing daily price variations of a given stock. Suppose that a particular methodology for event detection was applied, which revealed three types of events (shape patterns in this case), potentially useful for a final user. Each event is labelled with one of the strings from the set $\{peak, flat, valley\}$ and is characterized by two features, f_1 and f_2 , representing the output of the statistical functions mean and standard error. These statistics are calculated using daily prices, supposed to be subsequences of length $w = 12$. In the frame of our formalism the language L will include a 3-ary predicate symbol E , three variable symbols $y_i, i = 1..3$, two 12-ary function symbols f and g , two sets of constant symbols – $\{d_1, \dots, d_3\}$ and $\{c_1, \dots, c_n\}$ – and the usual set of relational symbols and logical(temporal) connectives. Consequently, a temporal atom in L is defined as $E(d_i, f(c_{j_1}, \dots, c_{j_{12}}), g(c_{k_1}, \dots, c_{k_{12}}))$, whereas an event template is defined as $E(y_1, y_2, y_3)$. Finally, the domain D is the union of the set $D_e = \{peak, flat, valley\}$ and of the set $D_f = \mathbb{R}^+$ (as the stock prices are positives real numbers and the features are statistical functions).

To define a first-order linear temporal logic based on L, we need a structure having a temporal dimension and capable to capture the relationship between a time moment and the interpretation I at this moment.

Definition 4. Given L and a domain D , a (first order) linear time structure is a triple $M = (S, x, \mathbf{I})$, where S is a set of states, $x : \mathbb{N} \rightarrow S$ is an infinite sequence of states $(s_1, s_2, \dots, s_n, \dots)$ and \mathbf{I} is a function that associates with each state s an interpretation \mathbf{I}_s of all symbols from L .

In the framework of linear temporal logic, the set of symbols is divided into two classes, the class of global symbols and the class of local symbols. Intuitively, a global symbol w has the same interpretation in each state, i.e. $\mathbf{I}_s(w) = \mathbf{I}_{s'}(w) = \mathbf{I}(w)$, for all $s, s' \in S$; the interpretation of a local symbol may vary, depending on the state at which is evaluated. The formalism of temporal rules assumes that all function symbols (including constants) and all relational symbols are global, whereas the predicate symbols and variable symbols are local. Consequently, as the temporal atoms, constraint formulae, temporal rules and the corresponding templates are expressed using the predicate symbol E or the variable symbols y_i , the meaning of truth for these formulae depend on the state at which they are evaluated. Given a first order time structure M and a formula p , we denote the instant i (or equivalently, the state s_i) for which $\mathbf{I}_{s_i}(p) = \text{true}$ by $i \models p$, i.e. at time instant i the formula p is true. Therefore, $i \models E(t_1, \dots, t_n)$ means that at time i an event with the name $\mathbf{I}(t_1)$ and characterized by the global features $\mathbf{I}(t_2), \dots, \mathbf{I}(t_n)$ occurs. Concerning the event template $E(y_1, \dots, y_n)$, the interpretation of the variable symbols y_j at the state s_i , $\mathbf{I}_{s_i}(y_j)$, is chosen such that $i \models E(y_1, \dots, y_n)$ for each time moment i . Because

- $i \models p \wedge q$ if and only if $i \models p$ and $i \models q$, and
- $i \models \nabla_k p$ if and only if $i + k \models p$,

a constraint formula (template) is true at time i if and only if all relational atoms are true at time i and $i \models E(t_1, \dots, t_n)$, whereas a temporal rule (template) is true at time i if and only if $i \models H_{m+1}$ and $i \models (H_1 \wedge \dots \wedge H_m)$.

Remark. The fact that the symbols of language L are divided into two sets (local and global), according to the persistence of their interpretation along the infinite sequence of states s_1, s_2, \dots , is the main reason for the introduction of the notion of *template*. Consider, as example, the temporal atom $E(t_1, t_2, t_3)$ and its corresponding template $E(y_1, y_2, y_3)$. In our vision, the event template is a kind of event pattern and because there is a real event which matches the pattern (an event with name $\mathbf{I}(t_1)$ and features $\mathbf{I}(t_2)$ and $\mathbf{I}(t_3)$), the interpretation of the template must be true at each moment. For this reason we imposed the condition that the interpretation of variable symbols must be chosen such that $i \models E(y_1, y_2, y_3)$ for each time moment i . On the other hand, we expect that in the real world, an event occurs only at certain moments, i.e. the interpretation of the event is evaluated *true* only at these moments. Because the terms t_i , $i = 1..3$ are global symbols (as constant symbols and function symbols) and $\mathbf{I}(E(t_1, t_2, t_3)) = \mathbf{I}(E)(\mathbf{I}(t_1), \mathbf{I}(t_2), \mathbf{I}(t_3))$, the only way to achieve the variability in time for the event interpretation is to include the predicate symbol E in the set of local symbols.

Example 2. The database of events, obtained after applying the specific methodology for event detection (see Example 1), contains tuples with three values, (v_1, v_2, v_3) . For a tuple with a recording index i , the first value expresses the name of the event – *peak*, *flat*, *valley* – which occurs at time moment i and the two other values express the values of the two characterizing features. Therefore, to define a linear time structure $M = (S, x, \mathbf{I})$, we may consider a state s as a triple (v_1, v_2, v_3) , the set S as the set of all tuples from the database and the sequence x as the ordered sequence of tuples in database (see Table 1).

Table 1. The first eighteen states of the linear time structure M (example)

Index	State	Index	State	Index	State
1	(<i>peak</i> , 11.2, 3.91)	7	(<i>peak</i> , 9.15, 4.03)	13	(<i>flat</i> , 7.14, 0.89)
2	(<i>peak</i> , 10.5, 4.87)	8	(<i>peak</i> , 11.52, 3.91)	14	(<i>peak</i> , 10.31, 4.42)
3	(<i>peak</i> , 14.03, 4.23)	9	(<i>flat</i> , 1.5, 1.81)	15	(<i>peak</i> , 12.8, 5.26)
4	(<i>flat</i> , 4.75, 1.42)	10	(<i>valley</i> , 3.08, 1.84)	16	(<i>flat</i> , 3.13, 1.44)
5	(<i>peak</i> , 9.49, 3.18)	11	(<i>valley</i> , 2.72, 1.58)	17	(<i>flat</i> , 5.08, 1.12)
6	(<i>valley</i> , 2.21, 1.12)	12	(<i>valley</i> , 4.42, 2.91)	18	(<i>valley</i> , 3.31, 3.20)

At this stage the interpretation of all symbols (global and local symbols) can be defined. For the global symbols (function symbols and relational symbols), the interpretation is quite intuitive. Therefore, the meaning $\mathbf{I}(d_j)$ is an element of D_e , the meaning $\mathbf{I}(c_j)$, $j = 1..n$, is a positive real number, whereas the meaning $\mathbf{I}(f)$, respectively $\mathbf{I}(g)$, is the function $f : D_f^{12} \rightarrow D_f$, $f(\mathbf{x}) = \bar{\mathbf{x}}$, respectively the function $g : D_f^{12} \rightarrow D_f$, $g(\mathbf{x}) = \text{se}(\mathbf{x})$ – we used the standard notations in statistics for the mean and standard error estimators.

The interpretation of a local symbol (the variable symbols y_i and the predicate symbol E) depends on the state at which is evaluated. For this example suppose that the function $\mathbf{I}_{s_i}(E)$ defined on D^3 with values in $B = \{\text{true}, \text{false}\}$ is provided by a finite algorithm. This algorithm will receive at input at least the state s_i and will provide at output one of the values from B . Therefore, the interpretation of $E(t_1, t_2, t_3)$ evaluated at s_i is defined as:

Algorithm 1 *Temporal atom evaluation*

Consider the state $s_i = (v_1, v_2, v_3)$
 If $(\mathbf{I}_{s_i}(t_1) = v_1)$ and $(\mathbf{I}_{s_i}(t_2) = v_2)$ and $(\mathbf{I}_{s_i}(t_3) = v_3)$
 Then $\mathbf{I}_{s_i}(E(t_1, t_2, t_3)) = \text{true}$
 Else $\mathbf{I}_{s_i}(E(t_1, t_2, t_3)) = \text{false}$

Finally, the interpretation of the variable symbol y_j at the state s_i is given by $\mathbf{I}_{s_i}(y_j) = v_j, j = 1..3$, which satisfies the condition imposed to the interpretation of temporal atom template, which is $\mathbf{I}_{s_i}E(y_1, y_2, y_3) = \text{true}$ for each state s_i . Having well-defined the language L , the syntax and the semantics of L , as well as the linear time structure M , we can construct the temporal atoms evaluated as *true* at time moment i (see Table 2).

Table 2. The temporal atoms evaluated *true* at the states s_2 , s_4 and s_{10} of M (example)

State	Temporal atom
2	$E(\text{peak}, f(3, 5, 7, 9, 15, 18, 19, 14, 12, 9, 8, 7), g(3, 5, 7, 9, 15, 18, 19, 14, 12, 9, 8, 7))$
4	$E(\text{flat}, f(3, 3, 4, 4, 5, 6, 6, 7, 7, 5, 4, 3), g(3, 3, 4, 4, 5, 6, 6, 7, 7, 5, 4, 3))$
10	$E(\text{valley}, f(5, 4, 2, 1, 1, 2, 2, 1, 3, 4, 5, 7), g(5, 4, 2, 1, 1, 2, 2, 1, 3, 4, 5, 7))$

3.1 Consistency

The connection between the restricted first-order temporal logic we defined and the temporal data mining task this logic tries to formalize (temporal rules extraction) is given by the following assumptions:

- A. For each formula p in L , there is an algorithm that calculates the value of the interpretation $\mathbf{I}_s(p)$, for each state s , in a finite number of steps.
- B. There are states (called incomplete states) that do not contain enough information to calculate the interpretation for all formulae defined at these states.
- C. It is possible to establish a measure, (called *general interpretation*) about the degree of truth of a compound formula along the entire sequence of states $(s_0, s_1, \dots, s_n, \dots)$.

The first assumption express the calculability of the interpretation \mathbf{I} (we already considered this assumption in Example 2). The second assumption express the situation when only the body of a temporal rule can be evaluated at a time moment i , but not the head of the rule. Therefore, for the state s_i , we cannot calculate the interpretation of the temporal rule and the only solution is to estimate it using a general interpretation. This solution is expressed by the third assumption. (*Remark:* The second assumption violates the condition about the existence of an interpretation in each state s_i , as defined in Definition 4. But it is well known that in data mining sometimes data is incomplete or is missing. Therefore, we must modify this condition as " \mathbf{I} is a function that associates with *almost* each state s an interpretation \mathbf{I}_s of all symbols from L ").

However, to ensure that this general interpretation is well defined, the linear time structure must present some property of consistency. Practically, this means that if we take any sufficiently large subset of time instants, the conclusions we may infer from this subset are sufficiently close from those inferred from the entire set of time instants. Therefore,

Definition 5. Given L and a linear time structure M , we say that M is a consistent time structure for L if, for every formula p , the limit $\text{supp}(p) = \lim_{n \rightarrow \infty} n^{-1} \#A$ exists, where $\#$ means "cardinality" and $A = \{i = 1..n \mid i \models p\}$. The notation $\text{supp}(p)$ denotes the support (of truth) of p .

Now we define the general interpretation for an n -ary predicate symbol P as:

Definition 6. Given L and a consistent linear time structure M for L , the general interpretation I_G for an n -ary predicate P is a function $D^n \rightarrow [0, 1]$, such that, for each n -tuple of terms $\{t_1, \dots, t_n\}$, $I_G(P(t_1, \dots, t_n)) = \text{supp}(P(t_1, \dots, t_n))$.

The general interpretation is naturally extended to constraint formulae, temporal rules and the corresponding templates. There is another useful measure, called *confidence*, but available only for temporal rules (templates). This measure is calculated as a limit ratio between the number of certain applications (time instants where both the body and the head of the rule are true) and the number of potential applications (time instants where only the body of the rule is true).

Definition 7. The confidence of a temporal rule (template) $H_1 \wedge \dots \wedge H_m \mapsto H_{m+1}$ is the limit $\lim_{n \rightarrow \infty} (\#B)^{-1} \#A$, where $A = \{i = 1 \dots n \mid i \models H_1 \wedge \dots \wedge H_{m+1}\}$ and $B = \{i = 1 \dots n \mid i \models H_1 \wedge \dots \wedge H_m\}$.

The relation between the property of consistency and the existence of the confidence for a temporal rule is expressed in the following lemma.

Lemma 1. If M is a consistent linear time structure for L then every temporal rule (template) $H_1 \wedge \dots \wedge H_m \mapsto H_{m+1}$ for which $\text{supp}(H_1 \wedge \dots \wedge H_m) \neq 0$ has a well-defined confidence.

For different reasons, (the user has not access to the entire sequence of states, or the states he has access to are incomplete), the general interpretation cannot be calculated. A solution is to estimate I_G using a finite linear time structure, i.e. a model.

Definition 8. Given L and a consistent time structure $M = (S, x, \mathbf{I})$, a model for M is a structure $\tilde{M} = (\tilde{T}, \tilde{x})$ where \tilde{T} is a finite temporal domain $\{i_1, \dots, i_n\}$, \tilde{x} is the subsequence of states $\{x_{i_1}, \dots, x_{i_n}\}$ (the restriction of x to the temporal domain \tilde{T}) and for each $i_j, j = 1, \dots, n$, the state x_{i_j} is a complete state.

Now we may define the estimator for a general interpretation:

Definition 9. Given L and a model \tilde{M} for M , an estimator of the general interpretation for an n -ary predicate P , $I_{G(\tilde{M})}(P)$, is a function $D^n \rightarrow [0, 1]$, assigning to each atomic formula $p = P(t_1, \dots, t_n)$ the value defined as the ratio $\frac{\#A}{\#\tilde{T}}$, where $A = \{i \in \tilde{T} \mid i \models p\}$. The notation $\text{supp}(p, \tilde{M})$ will denote the estimated support of p , given \tilde{M} .

The extension of this definition to the other types of formulae in L demands a deeper analysis. Consider, as example, the model \tilde{M} induced by the sequence of $n > 1$ states $\tilde{x} = x_1, \dots, x_n$. The interpretation of a formula $\nabla_1 p$ at the state x_n can not be calculated, because $n \models \nabla_1 p$ if $(n+1) \models p$, but $x_{n+1} \notin \tilde{x}$. Therefore, the cardinality of the set $A = \{i \leq n \mid i \models \nabla_1 p\}$ is strictly smaller than n , which means that, for p a global formula having the meaning of truth *true*, the estimated support is

$$\text{supp}(\nabla_1 p, \tilde{M}) = (n-1)/n \neq 1 = \text{supp}(\nabla_1 p). \quad (1)$$

The fact that the support estimator is biased seems at first glance without importance, especially when, as in this case, the bias (n^{-1}) tends to zero for $n \uparrow \infty$. But considering a formula of type $\nabla_n p$, it is evidently that the interpretation can not be calculated at none of the states from \tilde{x} , and so the support estimator is not even defined. Before indicating how the expression $\frac{\#A}{\#\tilde{T}}$ must be adjusted to avoid this kind of problem, we start by defining the standard form of a formula in L .

Definition 10. A formula $\nabla_{k_1} p_1 \wedge \nabla_{k_2} p_2 \wedge \dots \wedge \nabla_{k_n} p_n$, where $n \geq 1$ and p_i are atoms of L , is in standard form if exists $i_0 \in \{1, \dots, n\}$ such that $k_{i_0} = 0$ and for all $i = 1..n$, $k_i \leq 0$.

For an atomic formula p , it is clearly that its standard form is $\nabla_0 p$. Another example of formula in standard form is a temporal rule (template), where the head of the rule is prefixed by ∇_0 and all other constraint formulae are prefixed by ∇_{-k} , $k \geq 0$. It is obviously that, for M a consistent time structure, the support of a formula does not change if it is prefixed with a temporal connective ∇_k , $k \in \mathbb{Z}$. Therefore, to each formula p in L corresponds an equivalent formula (under the measure supp) having a standard form (denoted $\mathcal{F}(p)$). Based on this concept, we can now give a non equivocal definition for *time windows*:

Definition 11. Let be p a formula in L having the standard form $\nabla_{k_1} p_1 \wedge \nabla_{k_2} p_2 \wedge \dots \wedge \nabla_{k_n} p_n$. The time window of p – denoted $w(p)$ – is defined as $\max\{|k_i| : i = 1..n\}$

In the following, a formula having a time window equal zero will be called *temporal free formula*, whereas a formula with a strictly positive time window will be called a *temporal formula*. The concept of time window allows us to define a non biased estimator for the support measure.

Definition 12. Given L and a model \tilde{M} for M , the estimator of the support for a formula p in L and having $w(p) < \#\tilde{T} = m$, denoted $\text{supp}(p, \tilde{M})$, is the ratio

$$\frac{\#A}{m - w(p)}, \text{ where } A = \{i \in \tilde{T} \mid i \models \mathcal{F}(p)\}. \quad (2)$$

According to this definition, if $w(p) \geq m$ then the estimator $\text{supp}(p, \tilde{M})$ is not defined. The use of the standard form of a formula, in the construction of the set A , eliminates the interpretation problem for a formula of type $\nabla_k p$, $k \geq m$. Moreover, it is easy to see that $\text{supp}(\nabla_k p, \tilde{M}) = \text{supp}(p, \tilde{M})$, for all $k \in \mathbb{Z}$.

Definition 13. Given L and a model \tilde{M} for M , an estimate of the general interpretation for a formula p is given by

$$I_{G(\tilde{M})}(p) = \begin{cases} \text{supp}(p, \tilde{M}), & \text{if } w(p) < \#\tilde{T}, \\ 0 & \text{if } w(p) \geq \#\tilde{T} \end{cases} \quad (3)$$

Once again, the estimation of the confidence for a temporal rule (template) is defined as:

Definition 14. Given a model $\tilde{M} = (\tilde{T}, \tilde{x})$ for M , the estimation of the confidence for the temporal rule (template) $H_1 \wedge \dots \wedge H_m \mapsto H_{m+1}$ is the ratio $\frac{\#A}{\#B}$, where $A = \{i \in \tilde{T} \mid i \models H_1 \wedge \dots \wedge H_m \wedge H_{m+1}\}$ and $B = \{i \in \tilde{T} \mid i \models H_1 \wedge \dots \wedge H_m\}$. The notation $\text{conf}(H, \tilde{M})$ will denote the estimated confidence of the temporal rule (template) H given \tilde{M} .

According to the same arguments used in the definition of a correct support estimator, the existence of a confidence estimator for a temporal rule H is guaranteed only for models having a number of states greater than the time window of the rule. Moreover, if $\tilde{\tilde{T}}$ is the set obtained from \tilde{T} by deleting the first $w(H_1 \wedge \dots \wedge H_{m+1}) - w(H_1 \wedge \dots \wedge H_m)$ states, then we can obtain a non biased confidence estimator if in the expression of the set $B = \{i \in \tilde{T} \mid i \models H_1 \wedge \dots \wedge H_m\}$ the set \tilde{T} is replaced with $\tilde{\tilde{T}}$.

Example 3. Consider the following temporal rule template T (for the moment we are not concerned on how it was discovered):

$$\nabla_{-2}(y_1 = \text{peak}) \wedge \nabla_{-2}(y_2 < 11) \wedge \nabla_{-1}(y_1 = \text{peak}) \wedge \nabla_{-1}(y_3 > 3) \mapsto \nabla_0(y_1 = \text{flat})$$

which may be "translated" in a natural language as:

IF at time $t - 2$ an event "peak" occurred with a mean less than 11 AND at time $t - 1$ another event "peak" occurred with a standard error greater than 3 THEN at time t an event type "flat" occurs.

If M is a consistent linear time structure for the temporal language L defined in Example 1, then the model \tilde{M} given by the sequence of states $\{s_1, \dots, s_{18}\}$ (see Table 1) can be used to estimate the confidence of the temporal rule T . If the local support for the rule is 0.125 (*true* at states 2 and 14, among 17 states) and the local support for the body of the rule is 0.176 (*true* at states 2, 7 and 14, among 16 states) then the estimated confidence for the

rule, based on model \tilde{M} , is $0.125/0.176 = 0.71$. And, due to the consistency property, this estimation is a reliable information about the success rate for this rule when applied on future data.

4 The Granularity Model

We start with the concept of a temporal type to formalize the notion of time granularities, as described in [3].

Definition 15. Let $(\mathcal{T}, <)$ (index) be a linearly ordered temporal domain isomorphic to a subset of integers with the usual order relation, and let $(\mathcal{A}, <)$ (absolute time) be a linearly ordered set. Then, a temporal type on $(\mathcal{T}, \mathcal{A})$ is a mapping μ from \mathcal{T} to $2^{\mathcal{A}}$ such that

1. $\mu(i) \neq \emptyset$ and $\mu(j) \neq \emptyset$, where $i < j$, imply that each element in $\mu(i)$ is less than all the elements in $\mu(j)$, and
2. for all $i < j$, if $\mu(i) \neq \emptyset$ and $\mu(j) \neq \emptyset$, then $\forall k, i < k < j$ implies $\mu(k) \neq \emptyset$.

Each set $\mu(i)$, if non-empty, is called a granule of μ . Property (1) says that granules do not overlap and that the order on indexes follows the order on the corresponding granules. Property (2) disallows an empty set to be the value of a mapping for a certain index value if a lower index and a higher index are mapped to non-empty sets.

When considering a particular application or formal context, we can specialize this very general model along the following dimensions:

- choice of the index set \mathcal{T} ,
- choice of the absolute time set \mathcal{A} ,
- restrictions on the structure of granules,
- restrictions on the temporal types by using relationships.

We call the resulting formalization a temporal type system. Consider some possibilities for each of the above four dimensions. Convenient choices for the index set are natural numbers, integers, and any finite subset of them. The choice for absolute time is typically between dense and discrete. In general, if the application imposes a fixed basic granularity, then a discrete absolute time in terms of the basic granularity is probably the appropriate choice. However, if one is interested in being able to represent arbitrary finer temporal types, a dense absolute time is required. In both cases, specific applications could impose left/right boundedness on the absolute time set. The structure of ticks could be restricted in several ways:

- (1) disallow types with *gaps* in a granule,
- (2) disallow types with *non-contiguous* granules,
- (3) disallow types whose granules do *not cover all* the absolute time, or
- (4) disallow types with *nonuniform* granules (only types with granules having the same size are allowed).

4.1 Relationships and formal properties

Following Bettini et al. [3], we define a number of interesting relationships among temporal types.

Definition 16. Let μ and ν be temporal types on $(\mathcal{T}, \mathcal{A})$.

- *Finer-than*: μ is said to be finer than ν , denoted $\mu \preceq \nu$, if for each $i \in \mathcal{T}$, there exists $j \in \mathcal{T}$ such that $\mu(i) \subseteq \nu(j)$.
- *Groups-into*: μ is said to group into ν , denoted $\mu \trianglelefteq \nu$, if for each non-empty granule $\nu(j)$, there is a subset S of \mathcal{T} such that $\nu(j) = \bigcup_{i \in S} \mu(i)$.
- *Subtype*: μ is said to be a subtype of ν , denoted $\mu \sqsubseteq \nu$, if for each $i \in \mathcal{T}$, there exists $j \in \mathcal{T}$ such that $\mu(i) = \nu(j)$.
- *Shifting*: μ and ν are said to be shifting equivalent, denoted $\mu_1 \rightleftharpoons \mu_2$, if $\mu \sqsubseteq \nu$ and $\nu \sqsubseteq \mu$.

When a temporal type μ is finer than a temporal type ν , we also say that ν is *coarser* than μ . The *finer-than* relationship formalizes the notion of finer partitions of the absolute time. By definition, this relation is reflexive, i.e. $\mu \preceq \mu$ for each temporal type μ . Furthermore, the finer-than relation is obviously transitive. However, if no restrictions are given, it is not antisymmetric, and hence it is not a partial order. Indeed, $\mu \preceq \nu$ and $\nu \preceq \mu$ do not imply $\mu = \nu$, but only $\mu \rightleftharpoons \nu$. Considering the *groups-into* relation, $\mu \trianglelefteq \nu$ ensures that for each granule of μ there exists a set of granules of ν covering exactly the same span of time. The relation is useful, for example, in applications where attribute values are associated with time granules; sometimes it is possible to obtain the value associated with a granule of ν from the values associated with the granules of μ whose union covers the same time. The groups-into relation has the same two properties as the finer-than relation, but generally $\mu \preceq \nu$ does not imply $\mu \trianglelefteq \nu$ or viceversa. The *subtype* relation intuitively identifies a type corresponding to subsets of granules of another type. Similar to the two previous relations, subtype is reflexive and transitive, and satisfies $\mu \sqsubseteq \nu \Rightarrow \mu \preceq \nu$. Finally, *shifting* is clearly an equivalence relation. Concerning this last relation, an equivalent, more useful and practical definition, is:

Definition 17. Two temporal types μ_1 and μ_2 are said to be shifting equivalent (denoted $\mu_1 \rightleftharpoons \mu_2$) if there is a bijective function $h : \mathcal{T} \rightarrow \mathcal{T}$ such that $\mu_1(i) = \mu_2(h(i))$, for all $i \in \mathcal{T}$.

In the following we will consider only temporal type systems which satisfy the restriction that no pair of different types can be shifting equivalent, i.e.

$$\mu_1 \rightleftharpoons \mu_2 \Rightarrow \mu_1 = \mu_2. \quad (4)$$

For this class of systems, the three relationships \preceq , \trianglelefteq and \sqsubseteq are reflexive, transitive and antisymmetric and, hence, each relationship is a partial order. Therefore, for the relation we are particularly interested in, *finer-than*, there exists a unique least upper bound of the set of all temporal types, denoted by

μ_\top , and a unique greatest lower bound, denoted by μ_\perp . These top and bottom elements are defined as follows: $\mu_\top(i) = \mathcal{A}$ for some $i \in \mathcal{T}$ and $\mu_\top(j) = \emptyset$ for each $j \neq i$, and $\mu_\perp(i) = \emptyset$ for each $i \in \mathcal{T}$. Moreover, for each pair of temporal types μ_1, μ_2 , there exist a unique least upper bound (μ_1, μ_2) and a unique greatest lower bound (μ_1, μ_2) of the two types, with respect to \preceq . We formalize this result in the following theorem, proved by Bettini et al. [3]:

Theorem 1. *Any temporal type system having an infinite index, and satisfying (4), is a lattice with respect to the finer-than relationship.*

Let denote \mathcal{G}_0 the set of temporal types for which the index set and the absolute time set are isomorphic with the set of positive natural numbers, i.e. $\mathcal{A} = \mathcal{T} = \mathbb{N}$. Consider now the following particular subsets of \mathcal{G}_0 , represented by temporal types with a) non-empty granules, b) with granules covering all the absolute time and c) with constant size granules:

$$\mathcal{G}_1 = \{\mu \in \mathcal{G}_0 \mid \forall i \in \mathbb{N}, 0 < \#\mu(i)\} \quad (5)$$

$$\mathcal{G}_2 = \{\mu \in \mathcal{G}_1 \mid \forall i \in \mathbb{N}, \mu(i)^{-1} \neq \emptyset\} \quad (6)$$

$$\mathcal{G}_3 = \{\mu \in \mathcal{G}_2 \mid \forall i \in \mathbb{N}, \mu(i) = c_\mu\} \quad (7)$$

The membership of a temporal type defined by one of these subsets implies very useful properties, a first result being expressed in the following lemma.

Lemma 2. *If μ_1, μ_2 are temporal types from \mathcal{G}_1 , then $\mu_1 \preceq \mu_2 \Rightarrow \mu_1 = \mu_2$.*

Therefore, the set \mathcal{G}_1 of temporal types is a lattice with respect to the finer-than relationship. The temporal type system \mathcal{G}_1 is not closed, because (μ_1, μ_2) is not always in \mathcal{G}_1 . In exchange it can be shown that the temporal type system \mathcal{G}_2 (given by (6), where $\mu^{-1}(i) = \{j \in \mathbb{N} : i \in \mu(j)\}$) is a closed system having a unique greatest lower bound, $\mu_\perp(i) = i, \forall i \in \mathbb{N}$, but no least upper bound μ_\top . Furthermore, the membership of \mathcal{G}_2 is a sufficient condition for the equivalence of the relationships *finer-than* and *groups-into*, according to the following lemma:

Lemma 3. *If μ and ν are temporal types from \mathcal{G}_2 , then $\mu \preceq \nu \Leftrightarrow \mu \leq \nu$.*

4.2 Linear Granular Time Structure

If $M = (S, x, \mathbf{I})$ is a first-order linear time structure, then let the absolute time \mathcal{A} be given by the sequence x , by identifying the time moment i with the state $s_{(i)}$ (on the i^{th} position in the sequence). If μ is a temporal type from \mathcal{G}_2 , then the temporal granule $\mu(i)$ may be identified with the set $\{s_j \in S \mid j \in \mu(i)\}$. Therefore, the temporal type μ induces a new sequence, x_μ , defined as $x_\mu : \mathbb{N} \rightarrow 2^S, x_\mu(i) = \mu(i)$. (*Remark:* In the following the set $\mu(i)$ will be considered, depending of the context, either as a set of states or as a set of natural numbers, the indexes of these states).

Consider now the linear time structure derived from M , $M_\mu = (2^S, x_\mu, \mathbf{I}^\mu)$. To be well defined, we must give the interpretation $\mathbf{I}_{\mu(i)}^\mu$, for each $i \in \mathbb{N}$. Because for a fixed i the set $\mu(i)$ is a finite sequence of states, it defines (if all the states are complete states) a model $\tilde{M}_{\mu(i)}$ for M . Therefore the estimated general interpretation $I_{G(\tilde{M}_{\mu(i)})}$ is well defined and we consider, by definition, that for a temporal free formula (e.g. a temporal atom) p in L ,

$$\mathbf{I}_{\mu(i)}^\mu(p) = I_{G(\tilde{M}_{\mu(i)})}(p) = \text{supp}(p, \tilde{M}_{\mu(i)}) \quad (8)$$

This interpretation is extended to any temporal formula in L according to the rule:

$$\mathbf{I}_{\mu(i)}^\mu(\nabla_{k_1} p_1 \wedge \dots \wedge \nabla_{k_n} p_n) = \frac{1}{n} \sum_{j=1}^n \mathbf{I}_{\mu(i+k_j)}^\mu(p_j) \quad (9)$$

where p_i are temporal free formulae and $k_i \in \mathbb{Z}, i = 1 \dots n$.

Definition 18. If $M = (S, x, \mathbf{I})$ is a first-order linear time structure and μ is a temporal type from \mathcal{G}_2 , then the linear granular time structure induced by μ on M is the triple $M_\mu = (2^S, x_\mu, \mathbf{I}^\mu)$, where $x_\mu : \mathbb{N} \rightarrow 2^S$, $x_\mu(i) = \mu(i)$ and \mathbf{I}^μ is a function that associates with almost each set of states $\mu(i)$ an interpretation $\mathbf{I}_{\mu(i)}^\mu$ according to the rules (8) and (9).

Of a particular interest is the linear granular time structure induced by the greatest lower bound temporal type of the lattice \mathcal{G}_2 , $\mu_\perp(i) = i$. In this case, $M_{\mu_\perp} = (S, x, \mathbf{I}^{\mu_\perp})$, where the only difference from the initial time structure M is at the interpretation level: for $p = P(t_1, \dots, t_n)$ a formula in L , if the interpretation $\mathbf{I}_s(p)$ is a function defined on D^n with values in $\{true, false\}$ – giving so the meaning of truth – the interpretation $\mathbf{I}_s^{\mu_\perp}(p)$ is a function defined on D^n with values in $[0, 1]$ – giving so the degree of truth. The relation linking the two interpretations is given by $\mathbf{I}_s(p) = true$ if and only if $\mathbf{I}_s^{\mu_\perp}(p) = 1$. Indeed, supposing the state $s_{(i)}$ is a complete state, it defines the model $\tilde{M}_i = (i, s_{(i)})$ and we have, for p a temporal free formula,

$$\mathbf{I}_{\mu_\perp(i)}^{\mu_\perp}(p) = \text{supp}(p, \tilde{M}_i) = \begin{cases} 1, & \text{if } \mathbf{I}_{s(i)}(p) = true, \\ 0, & \text{if } \mathbf{I}_{s(i)}(p) = false \end{cases} \quad (10)$$

For a formula $\pi = \nabla_{k_1} p_1 \wedge \dots \wedge \nabla_{k_n} p_n$, we have $\mathbf{I}_{s(i)}(\pi) = true$ iff $\forall j \in \{1 \dots n\}, i + k_j \models p_j$, which is equivalent with

$$\begin{aligned} \text{supp}(p_1, \tilde{M}_{i+k_1}) = \dots = \text{supp}(p_n, \tilde{M}_{i+k_n}) = 1 \\ \Leftrightarrow \frac{1}{n} \sum_{j=1}^n \mathbf{I}_{\mu_\perp(i+k_j)}^{\mu_\perp}(p_j) = \mathbf{I}_{\mu_\perp(i)}^{\mu_\perp}(\pi) = 1. \end{aligned} \quad (11)$$

Consequently, the linear granular time structure M_{μ_\perp} can be seen as an extension, at the interpretation level, of the classical linear time structure M .

Example 4. Consider the first three granules of the linear granular time structure M_μ , which cover the first nine states of the linear time structure M (represented graphically in Fig. 1).

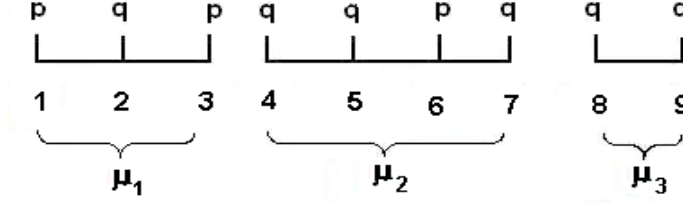


Fig. 1. Graphical representation of the first nine states from the time structure M and of the first three granules of temporal type μ

If p and q are two temporal free formulae in L , interpreted as *true* in the states $\{1, 3, 6\}$ (respectively $\{2, 4, 5, 7, 8, 9\}$) from M , then the following relations occur:

$$\begin{aligned}
 \mathbf{I}_{\mu(1)}(p) &= \text{supp}(p, \tilde{M}_{\mu_1}) = \frac{2}{3}, & \mathbf{I}_{\mu(2)}(p) &= \text{supp}(p, \tilde{M}_{\mu_2}) = \frac{1}{4} \\
 \mathbf{I}_{\mu(3)}(p) &= \text{supp}(p, \tilde{M}_{\mu_3}) = \frac{0}{2}, & \mathbf{I}_{\mu(1)}(q) &= \text{supp}(q, \tilde{M}_{\mu_1}) = \frac{1}{3} \\
 \mathbf{I}_{\mu(2)}(q) &= \text{supp}(q, \tilde{M}_{\mu_2}) = \frac{3}{4}, & \mathbf{I}_{\mu(3)}(q) &= \text{supp}(q, \tilde{M}_{\mu_3}) = \frac{2}{2} \\
 \mathbf{I}_{\mu(1)}(p \wedge \nabla q) &= \frac{1}{2} (\mathbf{I}_{\mu(1)}(p) + \mathbf{I}_{\mu(2)}(q)) = \frac{17}{24} \\
 \mathbf{I}_{\mu(1)}(q \wedge \nabla p \wedge \nabla_2 q) &= \frac{1}{3} (\mathbf{I}_{\mu(1)}(q) + \mathbf{I}_{\mu(2)}(p) + \mathbf{I}_{\mu(3)}(q)) = \frac{11}{36}
 \end{aligned}$$

4.3 Linking two Granular Time Structures

All the granular time structures induced by a temporal type have in common interpretations which take values in $[0, 1]$ if applied on predicate symbols in L . This observation allows us to establish the relation linking the interpretations \mathbf{I}^μ and \mathbf{I}^ν , from two linear granular time structures induced by μ and ν , when there exists a relationship *finer-than* between these two temporal types. According to the lemma 3, for each $i \in \mathbb{N}$ there is a subset $N_i \subset \mathbb{N}$ such that $\nu(i) = \bigcup_{j \in N_i} \mu(j)$. If p is a temporal free formula in L , then the interpretation \mathbf{I}^ν for p at $\nu(i)$ is the weighted sum of the interpretations $\mathbf{I}_{\mu(j)}^\mu(p)$, where $j \in N_i$. We formalize this result in the following theorem:

Theorem 2. *If μ, ν are temporal types from \mathcal{G}_2 , such that $\mu \preceq \nu$, and $\mathbf{I}^\mu, \mathbf{I}^\nu$ are the interpretations from the induced linear time structures M_μ and M_ν on M , then for each $i \in \mathbb{N}$,*

$$\mathbf{I}_{\nu(i)}^\nu(p) = \frac{1}{\#\nu(i)} \sum_{j \in N_i} \#\mu(j) \mathbf{I}_{\mu(j)}^\mu(p), \quad (12)$$

where N_i is the subset of \mathbb{N} which satisfies $\nu(i) = \bigcup_{j \in N_i} \mu(j)$ and p is a temporal free formula in L .

If we consider $\mu = \mu_\perp \preceq \nu$ then $\#\mu(j) = 1$, for all $j \in \mathbb{N}$ and $\mathbf{I}_{\mu(j)}^\mu(p) = \text{supp}(p, \tilde{M}_j)$. Therefore,

$$\begin{aligned} \mathbf{I}_{\nu(i)}^\nu(p) &= \frac{1}{\#\nu(i)} \sum_{j \in \nu(i)} \text{supp}(p, \tilde{M}_j) \\ &= \frac{1}{\#\nu(i)} \#\{j \in \nu(i) \mid j \models p\} = \text{supp}(p, \tilde{M}_{\nu(i)}) = \mathbf{I}_{G(\tilde{M}_{\nu(i)})}(p) \end{aligned} \quad (13)$$

result which is consistent with Definition 18. But the significance of the theorem 2 is revealed in a particular context. If $\mu, \nu \in \mathcal{G}_3$ and $\mu \preceq \nu$, it can be shown that $\#N_i = \frac{c_\nu}{c_\mu}, \forall i \in \mathbb{N}$ and so the relation (12) becomes

$$\mathbf{I}_{\nu(i)}^\nu(p) = \frac{1}{\#N_i} \sum_{j \in N_i} \mathbf{I}_{\mu(j)}^\mu(p). \quad (14)$$

Generally speaking, consider three worlds, W_1, W_2 and W_3 – defined as sets of granules of information – where W_1 is finer than W_2 which is finer than W_3 . Suppose also that the conversion between granules from two different worlds is given by a constant factor. If the **independent** part of information in each granule is transferred from W_1 to W_2 and then the world W_1 is "lost", the theorem 2 under the form (14) affirms that it is possible to transfer the independent information from W_2 to W_3 and to obtain the same result as for the transfer from W_1 to W_3 .

Example 5. : Consider a linear time structure M (here, the world W_1) and a temporal free formula p such that, for the first six time moments, we have $i \models p$ for $i \in \{1, 3, 5, 6\}$. The concept of independence, in this example, means that the interpretation of p in the state s_i does not depend on the interpretation of p in the state s_j . Let be $\mu, \nu \in \mathcal{G}_3$, $\mu \preceq \nu$, with $\mu(i) = \{2i - 1, 2i\}$ and $\nu(i) = \{6i - 5, \dots, 6i\}$. Therefore, $\nu(1) = \mu(1) \cup \mu(2) \cup \mu(3)$. According to the definition 18, $\mathbf{I}_{\mu(1)}^\mu(p) = \text{supp}(p, \{1, 2\}) = 0.5$, $\mathbf{I}_{\mu(2)}^\mu(p) = \text{supp}(p, \{3, 4\}) = 0.5$, $\mathbf{I}_{\mu(3)}^\mu(p) = \text{supp}(p, \{5, 6\}) = 1$, whereas $\mathbf{I}_{\nu(1)}^\nu(p) = \text{supp}(p, \{1, \dots, 6\}) = 0.66$. If the linear time structure M is "lost", the temporal types μ and ν are "lost" too (we don't know the absolute time \mathcal{A} given by M). But if we know the induced time structure M_μ (world W_2) and the relation between μ and ν

$$\nu(k) = \mu(3k - 2) \cup \mu(3k - 1) \cup \mu(3k), \forall k \in \mathbb{N}$$

then we can completely deduce the time structure M_ν (world W_3). As example, according to (14), $\mathbf{I}_{\nu(1)}^\nu(p) = \frac{1}{3} \sum_{i=1}^3 \mathbf{I}_{\mu(i)}^\mu(p) = 0.66$. The condition

about a constant conversion factor between granules is necessary because the information about the size of granules, as it appears in expression 12, is "lost" when the time structure M is "lost"

The theorem 2 is not effective for temporal formulae (which can be seen as the dependent part of the information of a temporal granule). In this case we can prove that the interpretation, in the coarser world, of a temporal formula with a given time window is linked with the interpretation, in the finer world, of a similar formula but having a larger time window.

Theorem 3. *If μ and ν are temporal types from \mathcal{G}_3 such that $\mu \preceq \nu$ and $\mathbf{I}^\mu, \mathbf{I}^\nu$ are the interpretations from the induced linear time structures M_μ and M_ν on M , then for each $i \in \mathbb{N}$,*

$$\mathbf{I}_{\nu(i)}^\nu(p \wedge \nabla q) = \frac{1}{k} \sum_{j \in N_i} \mathbf{I}_{\mu(j)}^\mu(p \wedge \nabla_k q) \quad (15)$$

where $k = c_\nu/c_\mu$, $\nu(i) = \bigcup_{j \in N_i} \mu(j)$ and p, q are temporal free formulae in L .

If we define the operator $zoom_k$ over the set of formulae in L as

$$zoom_k(\nabla_{k_1} p_1 \wedge \dots \wedge \nabla_{k_n} p_n) = \nabla_{k \cdot k_1} p_1 \wedge \dots \wedge \nabla_{k \cdot k_n} p_n$$

then an obvious corollary of this theorem is

Corollary 1. *If μ and ν are temporal types from \mathcal{G}_3 such that $\mu \preceq \nu$ and $\mathbf{I}^\mu, \mathbf{I}^\nu$ are the interpretations from the induced linear time structures M_μ and M_ν on M , then for each $i \in \mathbb{N}$,*

$$\mathbf{I}_{\nu(i)}^\nu(\nabla_{k_1} p_1 \wedge \dots \wedge \nabla_{k_n} p_n) = \frac{1}{k} \sum_{j \in N_i} \mathbf{I}_{\mu(j)}^\mu(zoom_k(\nabla_{k_1} p_1 \wedge \dots \wedge \nabla_{k_n} p_n)) \quad (16)$$

where $k = c_\nu/c_\mu$, $\nu(i) = \bigcup_{j \in N_i} \mu(j)$, $k_i \in \mathbb{N}$ and $p_i, i = 1..n$ are temporal free formulae in L .

According to this corollary, if we know the degree of truth of a temporal rule (template) in the world W_1 , we can say nothing about the degree of truth of the same rule in the world W_2 , coarser than W_1 . The information is only transferred from the temporal rule $zoom_k(H)$ in W_1 (which has a time window greater than $k-1$) to the temporal rule H in W_2 , where k is the coefficient of conversion between the two worlds. Consequently, all the information related to temporal formulae having a time window less than k is lost during the transition to the coarser world W_2 .

4.4 The Consistency Problem

The importance of the concepts of consistency, support and confidence, (see Subsect. 3.1), for the process of information transfer between worlds with different granularity may be highlighted by analyzing the analogous expressions for a linear granular time structure M_μ .

Definition 19. Given L and a linear granular time structure M_μ on M , we say that M_μ is a consistent granular time structure for L if, for every formula p , the limit

$$\text{supp}(p, M_\mu) = \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n \mathbf{I}_{\mu(i)}^\mu(p)}{n} \quad (17)$$

exists. The notation $\text{supp}(p, M_\mu)$ denotes the support (degree of truth) of p under M_μ .

A natural question concern the inheritance of the consistency property from the basic linear time structure M by the induced time structure M_μ . The answer is formalized in the following theorem.

Theorem 4. If M is a consistent linear time structure and $\mu \in \mathcal{G}_3$ then the granular time structure M_μ is also consistent.

The proof of the theorem (see Appendix 6) is based on the relation between the support of a formula p in M , respectively in M_μ , which is:

$$\text{supp}(p, M_\mu) = \text{supp}(p, M) \quad (18)$$

$$\text{supp}(\nabla_{k_1} p_1 \wedge \dots \wedge \nabla_{k_m} p_m, M_\mu) = \frac{1}{m} \sum_{j=1}^m \text{supp}(p_j, M) \quad (19)$$

The implications of Theorem 4 are extremely important. It is easy to show, starting from Definition 14, that the confidence of a temporal rule (template) may be expressed using only the support measure if the linear time structure M is consistent. Therefore, considering that by definition the confidence of a temporal rule (template) \mathcal{H} , $H_1 \wedge \dots \wedge H_m \mapsto H_{m+1}$, giving a consistent granular time structure M_μ , is

$$\text{conf}(\mathcal{H}, M_\mu) = \begin{cases} \frac{\text{supp}(H_1 \wedge \dots \wedge H_m \wedge H_{m+1}, M_\mu)}{\text{supp}(H_1 \wedge \dots \wedge H_m, M_\mu)} & \text{if } \text{supp}(H_1 \wedge \dots \wedge H_m, M_\mu) > 0, \\ 0 & \text{if not} \end{cases} \quad (20)$$

we can deduce, by applying Theorem 4 and the relations (18) and (19), that the confidence of \mathcal{H} , for a granular time structure M_μ induced on a consistent time structure M by a temporal type $\mu \in \mathcal{G}_3$, is independent of μ . In other words,

"The property of consistency is a sufficient condition for the independence of the measure of support/confidence, during the process of information transfer between worlds with different granularities, all derived from an absolute world using constant conversion factors. In practice, this means that even we are not able to find, for a given world M_μ , the granules where a temporal rule \mathcal{H} apply (according to Theorem 3), we are sure that the confidence of \mathcal{H} is the same in each world $M_\mu, \forall \mu \in \mathcal{G}_3$."

4.5 Event Aggregation

All the deduction processes made until now were conducted to obtain an answer to the following question: how is changing the degree of truth of a formula p if we pass from a linear time structure with a given granularity to a coarser one. And we proved that we can give a proper expression if we impose some restrictions on the temporal types which induce these time structures. But there is another phenomenon which follows the process of transition between two real worlds with different time granularities: new kinds of events appear, some kinds of events disappear.

Definition 20. *An event type (denoted $E[t]$) is the set of all temporal atoms from L having the same name (or head).*

All the temporal atoms of a given type $E[t]$ are constructed using the same symbol predicate and we denote by $N[t]$ the arity of this symbol. Consider $E(t, t_2, \dots, t_n) \in E[t]$ (where $n = N[t]$). According to Definition 1, a term $t_i, i \in \{2, \dots, n\}$ has the form $t_i = f(t_{i1}, \dots, t_{ik_i})$. Suppose now that for each index i the function symbol f from the expression of t_i belongs to a family of function symbols with different arities, denoted $\mathcal{F}_i[t]$ (so different sets for different event types $E[t]$ and different index i). This family has the property that the interpretation for each of its member is given by a real functions which

- is applied on a variable number of arguments, and
- is invariant in the order of the arguments.

A good example of a such real function is a statistical function, e.g. $\text{mean}(x_1, \dots, x_n)$. Based on the set $\mathcal{F}_i[t]$ we consider the set of terms expressed as $f_k(c_1, \dots, c_k)$, where f_k is a k -ary function symbol from $\mathcal{F}_i[t]$ and $c_i, i = 1..k$ are constant symbols. We denote such a set as $T_i[t]$. Consider now the operator \oplus defined on $T_i[t] \times T_i[t] \rightarrow T_i[t]$ such that

$$f_n(c_1, \dots, c_n) \oplus f_m(d_1, \dots, d_m) = f_{n+m}(c_1, \dots, c_n, d_1, \dots, d_m)$$

Of course, because the interpretation of any function symbol from $\mathcal{F}_i[t]$ is invariant in the order of arguments, we have

$$f_n(c_1, \dots, c_n) \oplus f_m(d_1, \dots, d_m) = f_n(c_{\sigma(1)}, \dots, c_{\sigma(n)}) \oplus f_m(d_{\varphi(1)}, \dots, d_{\varphi(m)})$$

where σ (respectively φ) is a permutation of the set $\{1, \dots, n\}$ (respectively $\{1, \dots, m\}$). Furthermore, it is evident that the operator \oplus is commutative and associative.

We introduce now a new operator (denoted \boxplus) defined on $E[t] \times E[t] \rightarrow E[t]$, such that, for $E(t, t_2, \dots, t_i, \dots, t_n) \in E[t]$, $E(t, t'_2, \dots, t'_i, \dots, t'_n) \in E[t]$ we have:

$$E(t, t_2, \dots, t_n) \boxplus E(t, t'_2, \dots, t'_n) = E(t, t_2 \oplus t'_2, \dots, t_n \oplus t'_n) \quad (21)$$

Once again, it is obviously that the operator \boxplus is commutative and associative. Therefore, we can apply this operator on a subset \mathcal{E} of temporal atoms from $E[t]$ and we denote the result as $\boxplus_{e_i \in \mathcal{E}} e_i$.

By definition, a formula p is satisfied by a linear time structure $M = (S, x, \mathbf{I})$ (respectively by a model $\tilde{M} = (\tilde{T}, \tilde{x})$ of M) if there is at least a state $s_i \in x$ (respectively $s_i \in \tilde{x}$) such that $\mathbf{I}_{s_i}(p) = \text{true}$. Therefore, the set of events of type t satisfied by M (respectively \tilde{M}) is given by:

$$E[t]_M = \{e \in E[t] \mid \exists s_i \in x \text{ such that } \mathbf{I}_{s_i}(e) = \text{true}\} \quad (22)$$

respectively by:

$$E[t]_{\tilde{M}} = \{e \in E[t] \mid \exists s_i \in \tilde{x} \text{ such that } \mathbf{I}_{s_i}(e) = \text{true}\} \quad (23)$$

If we consider now M_μ , the linear time structure induced by the temporal type μ on M , the definition of $E[t]_{M_\mu}$ is derived from (22) by changing the condition $\mathbf{I}_{s_i}(e) = \text{true}$ with $\mathbf{I}_{\mu(i)}^\mu(e) = 1$. Of course, only for $\mu = \mu_\perp$ we have $E[t]_M = E[t]_{M_\mu}$ (we proved that $\mathbf{I}_{s_i}(p) = \text{true} \Leftrightarrow \mathbf{I}_{\mu_\perp(i)}^{\mu_\perp}(p) = 1$). Generally $E[t]_M \supset E[t]_{M_\mu} \supset E[t]_{M_\nu}$, for $\mu \preceq \nu$, which is a consequence of the fact that a coarser world satisfies less temporal events than a finer one.

Example 6. : If M is a linear time structure such that for the event $e \in E[t]$ we have $i \models e$ if and only if i is odd, and μ is a temporal type given by $\mu(i) = \{2i - 1, 2i\}$, then it is obviously that $e \in E[t]_M$ but $e \notin E[t]_{M_\mu}$ (for all $i \in \mathbb{N}$, $\mathbf{I}_{\mu(i)}^\mu(e) = \text{supp}(e, \{2i - 1, 2i\}) = 0.5$).

In the same time a coarser world may satisfies new events, representing a kind of aggregation of local, "finer" events.

Definition 21. If μ is a temporal type from \mathcal{G}_2 , we call the aggregate event of type t induced by the granule $\mu(i)$ (denoted $e[t]_{\mu(i)}$) the event obtained by applying the operator \boxplus on the set of events of type t which are satisfied by the model $\tilde{M}_{\mu(i)}$, i.e.

$$e[t]_{\mu(i)} = \boxplus_{e_i \in E[t]_{\tilde{M}_{\mu(i)}}} e_i \quad (24)$$

According to (8), the interpretation of an event e in any world M_μ depends on the interpretation of the same event in M . Therefore, if e is not satisfied by M it is obvious that $\mathbf{I}_{\mu(i)}^\mu(e) = 0$, for all μ and all $i \in \mathbb{N}$. Because an aggregate event (conceived of a new, "federative" event), usually is not satisfied by M , the relation (8) is not appropriate to give the degree of truth for $e[t]_{\mu(i)}$. But before to give the rule expressing the interpretation for an aggregate temporal atom, we must impose on M the following restriction: two different events of type t can not be evaluated as *true* at the same state $s \in S$, or:

$$\exists h : E[t]_M \rightarrow S, h \text{ injective, such that } h(e) = s \text{ where } \mathbf{I}_s(e) = \text{true} \quad (25)$$

Definition 22. If M_ν is a linear granular time structure and $e[t]_{\mu(i_0)}$ is an aggregate event induced by the granule μ_{i_0} ($\mu, \nu \in \mathcal{G}_2$), then the interpretation of $e[t]_{\mu(i_0)}$ in the state $\nu(i)$ is defined as:

$$\mathbf{I}_{\nu(i)}^\nu(e[t]_{\mu(i_0)}) = \frac{\#(\mathcal{E}_i \cap \mathcal{E})}{\#\mathcal{E}} \sum_{e_j \in \mathcal{E}} \mathbf{I}_{\nu(i)}^\nu(e_j) \quad (26)$$

where $\mathcal{E} = E[t]_{\tilde{M}_{\mu(i_0)}}$, $\mathcal{E}_i = E[t]_{\tilde{M}_{\nu(i)}}$.

The restriction (25) is given to assure that $\sum_{e_j \in \mathcal{E}} \mathbf{I}_{\nu(i)}^\nu(e_j) \leq 1$, for all $i, i_0 \in \mathbb{N}$. Indeed, let be e_1, \dots, e_n the events from \mathcal{E} . If $h(e_j) = s_j, j = 1..n$, then consider the sets $A_j = \{k \in \nu(i) \mid k \models e_j\} = \{s \in \nu(i) \mid s = s_j\}$. The function h being injective, the sets A_j are disjoint and therefore $\sum_{j=1}^n \#A_j \leq \#\nu(i)$. Consequently, we have

$$\sum_{j=1}^n \mathbf{I}_{\nu(i)}^\nu(e_j) = \sum_{j=1}^n \text{supp}(e_j, M_{\nu(i)}) = \frac{1}{\#\nu(i)} \sum_{j=1}^n \#A_j \leq \frac{\#\nu(i)}{\#\nu(i)} = 1. \quad (27)$$

The relation (27) and the fact that the coefficient $\frac{\#(\mathcal{E}_i \cap \mathcal{E})}{\#\mathcal{E}}$ is less or equal one guarantee that the interpretation of an aggregate event is well-defined, i.e. $\mathbf{I}_{\nu(i)}^\nu(e[t]_{\mu(i_0)}) \leq 1$. Furthermore, the interpretation is equal one if and only if:

$$(i) \quad \frac{\#(\mathcal{E}_i \cap \mathcal{E})}{\#\mathcal{E}} = 1 \Leftrightarrow \mathcal{E} = \mathcal{E}_i \quad (28)$$

meaning that all the events of type t satisfied by $\tilde{M}_{\mu_{i_0}}$ are also satisfied by \tilde{M}_{ν_i} , and

$$(ii) \quad \sum_{j=1}^n \mathbf{I}_{\nu(i)}^\nu(e_j) = 1 \Leftrightarrow \frac{1}{\#\nu(i)} \sum_{j=1}^n \#A_j = 1 \Leftrightarrow \sum_{j=1}^n \#A_j = \#\nu(i) \quad (29)$$

meaning that the sets A_j form a partition of $\nu(i)$ (or equivalently $h^{-1}(\nu(i)) = \mathcal{E}$).

Example 7. Let be M a linear time structure, $e_1, e_2, e_3 \in E[t]$ such that (see Fig. 2)

$$\begin{aligned} 1 &\models e_1, 4 \models e_1 \text{ and } i \models e_1 \text{ for } i \in \{6k-2, 6k-1, 6k \mid k \geq 2\} \\ 3 &\models e_2, 5 \models e_2 \text{ and } i \models e_2 \text{ for } i \in \{6k+1, 6k+2, 6k+3 \mid k \geq 1\} \\ 6 &\models e_3 \end{aligned}$$

Consider two temporal types $\mu, \nu \in \mathcal{G}_3$ such that $\mu(i) = \{3i-2, 3i-1, 3i \mid i \geq 1\}$ and $\nu(i) = \{6i-5, \dots, 6i \mid i \geq 1\}$. The different aggregate events induced by granules of temporal type μ and ν are:

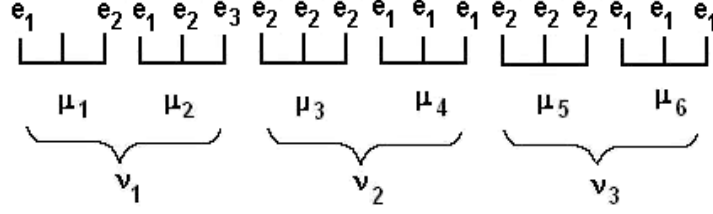


Fig. 2. Graphical representation of the first eighteen states of the time structure M and of the first six (respectively three) granules of temporal types μ and ν

$$e[t]_{\mu(1)} = e_1 \boxplus e_2, e[t]_{\mu(2)} = e_1 \boxplus e_2 \boxplus e_3, e[t]_{\mu(3)} = e_2, e[t]_{\mu(4)} = e_1 \\ e[t]_{\nu(1)} = e_1 \boxplus e_2 \boxplus e_3, e[t]_{\nu(i)} = e_1 \boxplus e_2 \text{ (for } i > 1)$$

Lets denote e_{12} the aggregate event induced by $\mu(1)$. Evidently $\mathcal{E} = E[t]_{\tilde{M}_{\mu(1)}} = \{e_1, e_2\}$ so $\#\mathcal{E} = 2$. The interpretation of this event in different granules of types μ and ν can be calculated according to relation (26):

$$\mathbf{I}_{\mu(1)}^{\mu}(e_{12}) = \frac{\#(\mathcal{E} \cap \mathcal{E})}{\#\mathcal{E}} \sum_{j=1}^2 \mathbf{I}_{\mu(1)}^{\mu}(e_j) = \sum_{j=1}^2 \text{supp}(e_j, \tilde{M}_{\mu(1)}) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3} \\ \mathbf{I}_{\mu(2)}^{\mu}(e_{12}) = \frac{\#(\mathcal{E}_2 \cap \mathcal{E})}{\#\mathcal{E}} \sum_{j=1}^2 \mathbf{I}_{\mu(2)}^{\mu}(e_j) = \frac{\#(\{e_1, e_2, e_3\} \cap \{e_1, e_2\})}{2} \sum_{j=1}^2 \text{supp}(e_j, \tilde{M}_{\mu(2)}) \\ = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}$$

$$\mathbf{I}_{\mu(3)}^{\mu}(e_{12}) = \frac{\#(\mathcal{E}_3 \cap \mathcal{E})}{\#\mathcal{E}} \sum_{j=1}^2 \mathbf{I}_{\mu(2)}^{\mu}(e_j) = \frac{\#(\{e_2\} \cap \{e_1, e_2\})}{2} \sum_{j=1}^2 \text{supp}(e_j, \tilde{M}_{\mu(3)}) \\ = \frac{1}{2} \left(\frac{0}{3} + \frac{3}{3} \right) = \frac{1}{2}$$

$$\mathbf{I}_{\mu(4)}^{\mu}(e_{12}) = \frac{\#(\mathcal{E}_4 \cap \mathcal{E})}{\#\mathcal{E}} \sum_{j=1}^2 \mathbf{I}_{\mu(2)}^{\mu}(e_j) = \frac{\#(\{e_1\} \cap \{e_1, e_2\})}{2} \sum_{j=1}^2 \text{supp}(e_j, \tilde{M}_{\mu(4)}) \\ = \frac{1}{2} \left(\frac{3}{3} + \frac{0}{3} \right) = \frac{1}{2}$$

$$\mathbf{I}_{\mu(2k+1)}^{\mu}(e_{12}) = \frac{1}{2}, \mathbf{I}_{\mu(2k+2)}^{\mu}(e_{12}) = \frac{1}{2}, \text{ for all } k \geq 1.$$

According to these results, the event e_{12} obviously is not satisfied by M_{μ} .

$$\begin{aligned}
\mathbf{I}_{\nu(1)}^\nu(e_{12}) &= \frac{\#(\mathcal{E}_1 \cap \mathcal{E})}{\#\mathcal{E}} \sum_{j=1}^2 \mathbf{I}_{\nu(1)}^\nu(e_j) = \frac{\#(\{e_1, e_2, e_3\} \cap \{e_1, e_2\})}{2} \sum_{j=1}^2 \text{supp}(e_j, \tilde{M}_{\nu(1)}) \\
&= \frac{2}{6} + \frac{2}{6} = \frac{2}{3} \\
\mathbf{I}_{\nu(2)}^\nu(e_{12}) &= \frac{\#(\mathcal{E}_2 \cap \mathcal{E})}{\#\mathcal{E}} \sum_{j=1}^2 \mathbf{I}_{\nu(2)}^\nu(e_j) = \frac{\#(\{e_1, e_2\} \cap \{e_1, e_2\})}{2} \sum_{j=1}^2 \text{supp}(e_j, \tilde{M}_{\nu(2)}) \\
&= \frac{3}{6} + \frac{3}{6} = 1 \\
\mathbf{I}_{\nu(k)}^\nu(e_{12}) &= 1 \text{ for all } k \geq 1,
\end{aligned}$$

which means that e_{12} is satisfied by M_ν , a coarser world than M_μ . As a general rule, the degree of truth for an aggregate event e is equal one in a given granule $\mu(i)$ if all individual events composing e (and only these events) are satisfied by $\mu(i)$.

5 Conclusions

In this article we developed a formalism for a specific temporal data mining task: the discovery of knowledge, represented in the form of general Horn clauses, inferred from databases with a temporal dimension. The theoretical framework we proposed, based on first-order temporal logic, permits to define the main notions (*event*, *temporal rule*, *constraint*) in a formal way. The concept of a consistent linear time structure allows us to introduce the notions of *general interpretation*, of *support* and of *confidence*, the last two measure being the expression of the two similar concepts used in data mining.

Starting from the inherent behavior of temporal systems – the perception of events and of their interactions is determined, in a large measure, by the temporal scale – we extended the capability of our formalism to "capture" the concept of time granularity. To keep an unitary viewpoint on the meaning of the same formula at different scales of time, we changed the usual definition of the interpretation \mathbf{I}^μ for a formula in the frame of a first-order temporal granular logic: it return the degree of truth (a real value between zero and one) and not only the meaning of truth (*true* or *false*).

The consequence of the definition for \mathbf{I}^μ is formalized in Theorem 2 : only the independent information (here, the degree of truth for a temporal free formula) may be transferred without loss between worlds with different granularities. Concerning the temporal rules (scale dependent information), we proved that the interpretation of a rule in a coarser world is linked with the interpretation of a similar rule in a finer world, rule obtained by applying the operator $zoom_k$ on the initial temporal rule.

By defining a similar concept of consistency for a granular time structure M_μ , we could proved that this property is inherited from the basic time structure M if the temporal type μ is of type \mathcal{G}_2 (granules with constant size). The

major consequence of Theorem 4 is that the confidence of a temporal rule (template) is preserved in all granular time structures derived from the same consistent time structure.

We defined also a mechanism to aggregate events of the same type, that reflects the following intuitive phenomenon: in a coarser world, not all events inherited from a finer world are satisfied, but in exchange there are new events which become satisfiable. To achieve this we extended the syntax and the semantics of L by allowing "family" of function symbols and by adding two new operators.

In our opinion, the logical next step in our work consists in adding a probabilistic dimension to the formalism. Preliminary results (see [11]) confirm that this approach allows a unified framework including the logical formalism and its granular extension, framework in which the property of consistency becomes a consequence of the capacity of a particular stochastic process to obey the strong law of large numbers.

6 Appendix. Proofs

Proof (of Lemma 2).

Before we start, we introduce the following notation: given two non-empty sets S_1 and S_2 of elements in \mathcal{A} , $S_1 \ll S_2$ holds if each number in S_1 is strictly less than each number in S_2 (formally, $S_1 \ll S_2$ if $\forall x \in S_1 \forall y \in S_2 (x < y)$). Moreover, we say that a set \mathbb{S} of non-empty sets of elements in \mathcal{A} is *monotonic* if for each pair of sets S_1 and S_2 in \mathbb{S} either $S_1 \ll S_2$ or $S_2 \ll S_1$.

The relation $\mu_1 \equiv \mu_2$ is equivalent with the existence of a bijection function $h : \mathbb{N} \rightarrow \mathbb{N}$ such that $\mu_1(i) = \mu_2(h(i))$, for all i . We will prove by induction that $h(i) = i$, which is equivalent with $\mu_1 = \mu_2$.

- $i = 1$: suppose that $h(1) > 1$. If $a = \min(\mu_1(1))$ – the existence of a is ensured by the condition (5) – then $\mu_1(1) = \mu_2(h(1)) \Rightarrow a \in \mu_2(h(1))$. Because $1 < h(1)$ we have $\mu_2(1) \ll \mu_2(h(1))$ (according to Definition 15) and so there is $b \in \mu_2(1)$ such that $b < a$. The inequality $1 < h(1)$ implies $h^{-1}(1) > 1$, and so $\mu_2(1) = \mu_1(h^{-1}(1)) \gg \mu_1(1)$. But the last inequality (\gg) is contradicted by the existence of $b \in \mu_1(h^{-1}(1))$ which is smaller than $a \in \mu_1(1)$. In conclusion, $h(1) = 1$.
- $i = n + 1$: from the induction hypothesis we have $h(i) = i, \forall i \leq n$. Supposing that $h(n+1) \neq n+1$, then the only possibility is $h(n+1) > n+1$. This last relation implies also $h^{-1}(n+1) > n+1$. Using a similar rationing as in the previous case (it's sufficient to replace 1 with $n+1$), we obtain

$$\mu_1(n+1) \ll \mu_1(h^{-1}(n+1)) = \mu_2(n+1) \ll \mu_2(h(n+1)) = \mu_1(n+1)$$

where each of the set from this relation are non-empty, according to (5). The contradiction of the hypothesis, in this case, means that $h(n+1) = n+1$ and, by induction principle, that $h(i) = i, \forall i \in \mathbb{N}$. ■

Proof (of Lemma 3).

Let be $\mu \in \mathcal{G}_2, \nu \in \mathcal{G}_2$.

- $\mu \preceq \nu$: let $j_0 \in \mathbb{N}$. The relation (6) means that for all $k \in \nu(j_0)$, $\mu^{-1}(k) \neq \emptyset$ and so $S = \bigcup_{k \in \nu(j_0)} \{\mu^{-1}(k)\} \neq \emptyset$. It is obviously, according to Definition 15, that the relation *finer-than* implies that for each $i \in \mathbb{N}$ there is a **unique** $j \in \mathbb{N}$ such that $\mu(i) \subseteq \nu(j)$. Consequently, if $\mu \preceq \nu$ and $\mu(i) \cap \nu(j) \neq \emptyset$ then $\mu(i) \subseteq \nu(j)$. Therefore, for all $i \in S$, $\mu(i) \subseteq \nu(j_0)$ which implies $\bigcup_{i \in S} \mu(i) \subseteq \nu(j_0)$ (a). At the same time, $\forall k \in \nu(j_0)$ we have $k \in \mu(\mu^{-1}(k))$ which implies $\nu(j_0) \subseteq \bigcup_{i \in S} \mu(i)$ (b). From (a) and (b) we have $\nu(j_0) = \bigcup_{i \in S} \mu(i)$, which implies $\mu \preceq \nu$.
- $\mu \trianglelefteq \nu$: let $i_0 \in \mathbb{N}$ and let $k \in \mu(i_0)$. According to (6), there exists $j = \nu^{-1}(k)$. Because $\mu \trianglelefteq \nu$ there is a set S such that $\nu(j) = \bigcup_{i \in S} \mu(i)$. Because the sets $\mu(i), i \in S$ are disjunct and $k \in \nu(j) \cap \mu(i_0)$ we have $i_0 \in S$. Therefore, for each i_0 there is $j \in \mathbb{N}$ such that $\mu(i_0) \subseteq \nu(j)$, which implies $\mu \preceq \nu$. ■

Proof (of Theorem 2).

The formula p being a temporal free formula, we have $w(p) = 0$. According to Definition 18 and Definition 12, we have

$$\mathbf{I}_{\nu(i)}^\nu(p) = \text{supp}(p, \tilde{M}_{\nu(i)}) = \frac{\#\{j \in \nu(i) \mid j \models p\}}{\#\nu(i)} \quad (30)$$

On the other hand, because $\nu(i) = \bigcup_{j \in N_i} \mu(j)$, we have also

$$\begin{aligned} \frac{1}{\#\nu(i)} \sum_{j \in N_i} \#\mu(j) \mathbf{I}_{\mu(j)}^\mu(p) &= \frac{1}{\#\nu(i)} \sum_{j \in N_i} \#\mu(j) \text{supp}(p, \tilde{M}_{\mu(j)}) \\ &= \frac{1}{\#\nu(i)} \sum_{j \in N_i} \#\{k \in \mu(j) \mid k \models p\} = \frac{\#\{j \in \nu(i) \mid j \models p\}}{\#\nu(i)} \end{aligned} \quad (31)$$

From (30) and (31) we obtain (12). ■

Proof (of Theorem 3).

If $\mu, \nu \in \mathcal{G}_3$ such that $\#\mu(i) = c_\mu$ and $\#\nu(i) = c_\nu$, for all $i \in \mathbb{N}$, it is easy to show that the sets N_i satisfying $\nu(i) = \bigcup_{j \in N_i} \mu(j)$ have all the same cardinality, $\#N_i = c_\nu/c_\mu = k$ and contain successive natural numbers, $N_i = \{j_i, j_i + 1, \dots, j_i + k - 1\}$. From the relations (9) and (14) we have:

$$\begin{aligned}
\mathbf{I}_{\nu(i)}^{\nu}(p \wedge \nabla q) &= \frac{1}{2} \left(\mathbf{I}_{\nu(i)}^{\nu}(p) + \mathbf{I}_{\nu(i+1)}^{\nu}(q) \right) \\
&= \frac{1}{2} \left(\frac{1}{\#N_i} \sum_{j \in N_i} \mathbf{I}_{\mu(j)}^{\mu}(p) + \frac{1}{\#N_{i+1}} \sum_{j \in N_{i+1}} \mathbf{I}_{\mu(j)}^{\mu}(q) \right) \\
&= \frac{1}{2} \left(\frac{1}{k} \sum_{j=j_i}^{j_i+k-1} \mathbf{I}_{\mu(j)}^{\mu}(p) + \frac{1}{k} \sum_{j=j_i+k}^{j_i+2k-1} \mathbf{I}_{\mu(j)}^{\mu}(q) \right) \\
&= \frac{1}{2k} \left(\sum_{j=j_i}^{j_i+k-1} \left(\mathbf{I}_{\mu(j)}^{\mu}(p) + \mathbf{I}_{\mu(j+k)}^{\mu}(q) \right) \right) \\
&= \frac{1}{2k} \left(\sum_{j=j_i}^{j_i+k-1} 2\mathbf{I}_{\mu(j)}^{\mu}(p \wedge \nabla_k q) \right) = \frac{1}{k} \sum_{j \in N_i} \mathbf{I}_{\mu(j)}^{\mu}(p \wedge \nabla_k q) \blacksquare.
\end{aligned}$$

Proof (of Theorem 4).

M being a consistent time structure, for each formula p in L the sequence $x(p)_n = n^{-1} \# \{i \leq n \mid i \models p\}$ has a limit and $\lim_{n \rightarrow \infty} x(p)_n = \text{supp}(p, M)$. In the same time, $\mu \in \mathcal{G}_3$ implies $\#\mu(i) = k$ for all $i \in \mathbb{N}$ and $\mu(i) = \{k(i-1) + 1, k(i-1) + 2, \dots, ki\}$. Consider the following two cases:

- *p temporal free formula* : We have

$$\begin{aligned}
\frac{\sum_{i=1}^n \mathbf{I}_{\mu(i)}^{\mu}(p)}{n} &= \frac{\sum_{i=1}^n \text{supp}(p, M_{\mu(i)})}{n} \\
&= \frac{\sum_{i=1}^n \frac{\#\{j \in \mu(i) \mid j \models p\}}{\#\mu(i)}}{n} = \frac{\sum_{i=1}^n \#\{j \in \mu(i) \mid j \models p\}}{kn} \\
&= \frac{\#\{i \in \bigcup_{i=1}^n \mu(i) \mid i \models p\}}{kn} = \frac{\#\{i \leq kn \mid i \models p\}}{kn} = x(p)_{kn}
\end{aligned}$$

Therefore, there exists the limit $\lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n \mathbf{I}_{\mu(i)}^{\mu}(p)}{n} = \lim_{n \rightarrow \infty} x(p)_{kn}$ and we have

$$\text{supp}(p, M_{\mu}) = \text{supp}(p, M) \text{ for } p \text{ temporal free formula} \quad (32)$$

- *temporal formula* $\pi = \nabla_{k_1} p_1 \wedge \dots \wedge \nabla_{k_m} p_m$: We have

$$\begin{aligned}
& \frac{\sum_{i=1}^n \mathbf{I}_{\mu(i)}^\mu(\nabla_{k_1} p_1 \wedge \dots \wedge \nabla_{k_m} p_m)}{n} = \frac{\sum_{i=1}^n \left(m^{-1} \sum_{j=1}^m \mathbf{I}_{\mu(i+k_j)}^\mu(p_j) \right)}{n} \\
& = \frac{1}{m} \frac{\sum_{i=1}^n \sum_{j=1}^m \text{supp}(p_j, M_{\mu(i+k_j)})}{n} = \frac{1}{mn} \sum_{j=1}^m \sum_{i=1}^n \text{supp}(p_j, M_{\mu(i+k_j)}) \\
& = \frac{1}{mn} \sum_{j=1}^m \sum_{i=1}^n \frac{\#\{h \in \mu(k_j + i) \mid h \models p_j\}}{k} \\
& = \frac{1}{mn} \sum_{j=1}^m \frac{\#\{h \in \bigcup_{i=1}^n \mu(k_j + i) \mid h \models p_j\}}{k} \\
& = \frac{1}{mnk} \sum_{j=1}^m (\#\{h \leq k(k_j + n) \mid h \models p_j\} - \#\{h \leq kk_j \mid h \models p_j\}) \\
& = \frac{1}{mnk} \sum_{j=1}^m (k(k_j + n)x(p_j)_{k(k_j+n)} - kk_j x(p_j)_{kk_j}) \\
& = \frac{1}{m} \sum_{j=1}^m \frac{k_j + n}{n} x(p_j)_{k(k_j+n)} - \frac{1}{m} \sum_{j=1}^m \frac{k_j}{n} x(p_j)_{kk_j} \quad (33)
\end{aligned}$$

By taking $n \rightarrow \infty$ in (33), we obtain

$$\begin{aligned}
& \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n \mathbf{I}_{\mu(i)}^\mu(\nabla_{k_1} p_1 \wedge \dots \wedge \nabla_{k_m} p_m)}{n} \\
& = \lim_{n \rightarrow \infty} \left(\frac{1}{m} \sum_{j=1}^m \frac{k_j + n}{n} x(p_j)_{k(k_j+n)} - \frac{1}{m} \sum_{j=1}^m \frac{k_j}{n} x(p_j)_{kk_j} \right) \\
& = \frac{1}{m} \sum_{j=1}^m \lim_{n \rightarrow \infty} \frac{k_j + n}{n} x(p_j)_{k(k_j+n)} - \frac{1}{m} \sum_{j=1}^m \lim_{n \rightarrow \infty} \frac{k_j}{n} x(p_j)_{kk_j} \\
& = \frac{1}{m} \sum_{j=1}^m \lim_{n \rightarrow \infty} x(p_j)_{k(k_j+n)} = \frac{1}{m} \sum_{j=1}^m \text{supp}(p_j, M) \quad (34)
\end{aligned}$$

and so we have

$$\text{supp}(\nabla_{k_1} p_1 \wedge \dots \wedge \nabla_{k_m} p_m, M_\mu) = \frac{1}{m} \sum_{j=1}^m \text{supp}(p_j, M) \quad (35)$$

From (6) and (35) results the conclusion of the theorem ■.

References

- [1] S. Al-Naemi. A theoretical framework for temporal knowledge discovery. In *Proceedings of International Workshop on Spatio-Temporal Databases*, pages 23–33, Spain, 1994.

- [2] C. Bettini, X. S. Wang, and S. Jajodia. Mining temporal relationships with multiple granularities in time sequences. *Data Engineering Bulletin*, 21(1):32–38, 1998.
- [3] C. Bettini, X. S. Wang, and S. Jajodia. A general framework for time granularity and its application to temporal reasoning. *Ann. Math. Artif. Intell.*, 22(1-2):29–58, 1998.
- [4] C. Bettini, X. S. Wang, S. Jajodia, and J.-L. Lin. Discovering frequent event patterns with multiple granularities in time sequences. *IEEE Trans. Knowl. Data Eng.*, 10(2):222–237, 1998.
- [5] X. Chen and I. Petrounias. A Framework for Temporal Data Mining. *Lecture Notes in Computer Science*, 1460:796–805, 1998.
- [6] J. Chomicki and D. Toman. Temporal Logic in Information Systems. *BRICS Lecture Series*, LS-97-1:1–42, 1997.
- [7] E. Ciapessoni, E. Corsetti, A. Montanari, and P. S. Pietro. Embedding time granularity in a logical specification language for synchronous real-time systems. *Sci. Comput. Program.*, 20(1-2):141–171, 1993.
- [8] J. Clifford and A. Rao. A simple general structure for temporal domains. In *Temporal Aspects of Information Systems*. Elsevier Science, 1988.
- [9] P. Cotofrei and K. Stoffel. Classification Rules + Time = Temporal Rules. In *Lecture Notes in Computer Science, vol 2329*, pages 572–581. Springer Verlag, 2002.
- [10] P. Cotofrei and K. Stoffel. From temporal rules to temporal meta-rules. In *Proceedings of 6th International Conference Data Warehousing and Knowledge Discovery, DaWaK 2004*, Lecture Notes in Computer Science, vol. 3181, pages 169–178, Zaragoza, Spain, 2004.
- [11] P. Cotofrei and K. Stoffel. Stochastic processes and temporal data mining. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 183–190, San Jose, USA, August 2007.
- [12] E. A. Emerson. Temporal and Modal Logic. *Handbook of Theoretical Computer Science*, pages 995–1072, 1990.
- [13] J. Euzenat. An algebraic approach to granularity in qualitative time and space representation. In *IJCAI (1)*, pages 894–900, 1995.
- [14] C. Evans. The macro-event calculus: representing temporal granularity. In *Proceedings of PRICAI*, Japan, 1990.
- [15] F. Giunchiglia and T. Walsh. A theory of abstraction. *Artificial Intelligence*, 56:323–390, 1992.
- [16] J. Han, Y. Cai, and N. Cercone. Data-driven discovery of quantitative rules in databases. *IEEE Transactions on Knowledge and Data Engineering*, 5:29–40, 1993.
- [17] J. Hobbs. Granularity. In *Proceedings of the IJCAI-85*, pages 432 – 435, 1985.
- [18] K. Hornsby. Temporal zooming. *Transactions in GIS*, 5:255–272, 2001.

- [19] C. Knoblock. *Generating Abstraction Hierarchies: an Automated Approach to Reducing Search in Planning*. Kluwer Academic Publishers, 1993.
- [20] T. Y. Lin and E. Louie. Data mining using granular computing: fast algorithms for finding association rules. *Data mining, rough sets and granular computing*, pages 23–45, 2002.
- [21] D. Malerba, F. Esposito, and F. Lisi. A logical framework for frequent pattern discovery in spatial data. In *Proceedings of 5th Conference Knowledge Discovery in Data*, 2001.
- [22] I. Mani. A theory of granularity and its application to problems of polysemy and underspecification of meaning. In *Proceedings of the Sixth International Conference Principles of Knowledge Representation and Reasoning*, pages 245–255, 1998.
- [23] G. McCalla, J. Greer, J. Barrie, and P. Pospisil. Granularity hierarchies. *Computers and Mathematics with Applications*, 23:363–375, 1992.
- [24] G.-C. Roman. Formal specification of geographic data processing requirements. *IEEE Trans. Knowl. Data Eng.*, 2(4):370–380, 1990.
- [25] L. Saitta and J.-D. Zucker. Semantic abstraction for concept representation and learning. In *Proceedings of the Symposium on Abstraction, Reformulation and Approximation*, pages 103–120, 1998.
- [26] J. Stell and M. Worboys. Stratified map spaces: a formal basis for multi-resolution spatial databases. In *Proceedings of the 8th International Symposium on Spatial Data Handling*, pages 180–189, 1998.
- [27] Y. Yao. Granular computing: basic issues and possible solutions. In P. Wang, editor, *Proceedings of the 5th Joint Conference on Information Sciences*, pages 186–189, Atlantic City, New Jersey, USA, 2000.
- [28] Y. Yao and N. Zhong. Potential applications of granular computing in knowledge discovery and data mining. In M. Torres, B. Sanchez, and J. Aguilar, editors, *Proceedings of World Multiconference on Systemics, Cybernetics and Informatics*, pages 573–580, Orlando, Florida, USA, 1999. International Institute of Informatics and Systematics.
- [29] L. A. Zadeh. Information granulation and its centrality in human and machine intelligence. In *Rough Sets and Current Trends in Computing*, pages 35–36, 1998.
- [30] B. Zhang and L. Zhang. *Theory and Applications of Problem Solving*. North-Holland, Amsterdam, 1992.
- [31] L. Zhang and B. Zhang. The quotient space theory of problem solving. In *Proceedings of International Conference on Rough Sets, Fuzzy Set, Data Mining and Granular Computing*, pages 11–15, 2003.