From Temporal Rules to Temporal Meta-Rules*

Paul Cotofrei and Kilian Stoffel

University of Neuchatel,
Pierre-a-Mazel 7, 2000 Neuchatel, Switzerland
paul.cotofrei@unine.ch, kilian.stoffel@unine.ch

Abstract. In this article we define a formalism to discover knowledge in the form of *temporal rules*, inferred from databases of *events* with a temporal dimension. The theoretical framework is based on first-order temporal logic and allows the definition of the main temporal data mining notions (event, temporal rule, constraint) in a formal way. The concepts of *consistent linear time structure* and *general interpretation* are fundamentals in the design of algorithms for inferring higher order temporal rules, (called *temporal meta-rules*), from local sets of temporal rules.

1 Introduction

The domain of temporal data mining focuses on the discovery of causal relationships among events that may be ordered in time and may be causally related. The contributions in this domain encompass the discovery of temporal rules, 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.

The main tasks concerning the information extraction from time series database and on which the researchers concentrated their efforts over the last years may be divided in several domains. *Similarity/Pattern Querying* concerns the measure of similarity between two sequences or sub-sequences respectively. Different methods were developed, such as window stitching [1] or dynamic time warping based matching [12]. *Clustering/Classification* concentrates on optimal algorithms for clustering/classifying sub-sequences of time series into groups (classes) of similar sub-sequences. Different techniques were proposed, as Hidden Markov Models (HMM) [13], Dynamic Bayes Networks (DBNs) [7] or Recurrent Neural Networks [8]. *Pattern finding/Prediction* methods regard the search for periodicity patterns (fully or partially periodic) in time series databases. For full periodicity search there is a rich collection of statistic methods, like FFT [14]. For partial periodicity search, different algorithms were developed, which explore properties related to partial periodicity such as the max-subpattern-hit-set property [10] or point-wise periodicity [9]. *Temporal Rules' extraction* approach concentrated to the extraction of explicit rules from time series, like temporal association rules [4] or cyclic association rules [16]. Adaptive methods

* Supported by the Swiss National Science Foundation (grant N ° 2100-063 730.

for finding rules whose conditions refer to patterns in time series were described in [6, 11] and a general methodology for classification and extraction of temporal rules was proposed in [5].

Although there is a rich bibliography concerning formalism for temporal databases, there are very few articles on this topic for temporal data mining. In [2, 3, 15] 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. In this article, we provide an innovative formalism based on first-order temporal logic, which permits an abstract view on temporal rules. The formalism allows also the application of an inference phase in which higher order temporal rules (called temporal meta-rules) are inferred from local temporal rules, the lasts being extracted from different sequences of data. Using this strategy, known in the literature as higher order mining [18], we can guarantee the scalability of our system (the capacity to handle huge databases), by applying statistical and machine learning tools.

The rest of the paper is structured as follows. In the next section, the first-order temporal logic formalism is extensively presented (definitions of the main terms – *event, temporal rules, confidence* – and concepts – *consistent linear time structure, general interpretation*). Section 3 contains a brief description of a general methodology for temporal rules extraction, through the frame of the proposed formalism. The notion of temporal meta-rules and the algorithms for inferring such higher order rules are described in Section 4. Finally, the last section summarizes our work and lists some possible future directions.

2 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. It can be based either on *time points* (instants) or on *intervals* (periods), may have a *discrete* or a *continuous* structure and may be *linear* or *nonlinear* (e.g. acyclic graph). For our methodology, we chose 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 purposes of our methodology we consider a restricted first-order temporal language L which contains only n-ary function symbols ($n \geq 0$), n-ary predicate symbols ($n > 1$, so no proposition symbols), the set of relational symbols $\{=, <, \leq, >, \geq\}$, a single logical connective $\{\wedge\}$ and a temporal connective of the form $X_k,\ k \in \mathbf{Z}$, where k strictly positive means *next k times*, k strictly negative means *last k times* and $k = 0$ means *now*.

The syntax of L defines terms, atomic formulae and compound formulae. A Horn clause is a formula of the form $B_1 \wedge \cdots \wedge B_m \rightarrow B_{m+1}$ where each B_i is a positive (non-negated) atom. The atoms $B_i,\ i = 1, \ldots, m$ 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 defined. However, to allow the description of rules, which formally look like a Horn clause, we introduce a new logical connective, $\mapsto$, which practically will represent a rewrite of the connective $\wedge$. Therefore, a formula in L of the form $p \mapsto q$ is syntactically equivalent with 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* $(n \geq 1)$ $E(t_1, t_2, \ldots, t_n)$*. The first term of the tuple,* t_1*, is a constant representing the name of the event and all others terms are function symbols. A short temporal atom (or the event's head) is the atom* $E(t_1)$.

Definition 2 *A constraint formula for the event* $E(t_1, t_2, \ldots t_n)$ *is a conjunctive compound formula,* $C_1 \wedge C_2 \wedge \cdots \wedge C_k$*, where each* C_j *is a relation implying one of the terms* t_i.

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

Definition 3 *A temporal rule is a formula of the form* $H_1 \wedge \cdots \wedge H_m \mapsto H_{m+1}$*, where* H_{m+1} *is a short constraint formula and* H_i *are constraint formulae, prefixed by the temporal connectives* X_{-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*.

Practically, the only atoms constructed in L are temporal atoms and the only formulae constructed in L are constraint formulae and temporal rules. As a consequence of the definition 3, a conjunction of relations $C_1 \wedge C_2 \wedge \cdots \wedge C_n$, each relation prefixed by temporal connectives X_{-k}, $k \geq 0$, may be rewritten as $C_{\sigma(1)} \wedge \cdots \wedge C_{\sigma(n-1)} \mapsto C_{\sigma(n)}$, — σ being a permutation of $\{1..n\}$ — if and only if there is a short constraint formula $C_{\sigma(n)}$ prefixed by X_0.

The semantics of L is provided by an interpretation I over a domain D. 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 (function symbols $t_2, \cdots, t_n$). 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.

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 $M = (S, x, \Im)$, where S is a set of states, $x : \mathbf{N} \rightarrow S$ is an infinite sequence of states $(s_0, s_1, \ldots, s_n, \ldots)$ and $\Im$ is a function that associates to each state s an interpretation $\Im(s)$ of all symbols defined at s.*

In the framework of temporal data mining, the function $\Im$ is a constant and it is equal to the interpretation I. In fact, the meaning of the events, constraint formulae and temporal rules is not changing over time. What is changing over time is the value of the meaning. Given a first order time structure M, we denote the instant i (or equivalently, the state s_i) for which $I(P) = true$ by $i \Rightarrow P$, i.e. at time instant i the formula P is true. Therefore, $i \Rightarrow E(t_1, \ldots, t_n)$ means that at time i an event with the name t_1 and characterized by the global features $t_2, \ldots, t_n$ started. A constraint formula is true at time i if and only if all relations are true at time i. A temporal rule is true at time i if and only if $i \Rightarrow H_{m+1}$ and $i \Rightarrow (H_1 \wedge \cdots \wedge H_m)$. (*Remark*: $i \Rightarrow P \wedge Q$ if and only if $i \Rightarrow P$ and $i \Rightarrow Q$; $i \Rightarrow X_k P$ if and only if $i + k \Rightarrow P$).

Now suppose that the following assumptions are true:

A. For each formula P in L, there is an algorithm that calculates the value of the interpretation $I(P)$ 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 $(s_0, s_1, \ldots, s_n, \ldots)$.

The first assumption expresses the calculability of the interpretation I. The second assumption expresses the situation when only the body of a temporal rule can be evaluated at 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, from 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 "$\Im$ *is a function that associates to* `almost` *each state s an interpretation* $\Im(s)$ *of all symbols defined at s* ").

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 n-ary predicate symbol P, the limit*

$co(P) = \lim_{n\to\infty} \frac{\#A}{n}$ *exists, where* $A = \{i \in \{0, \ldots, n\} | i \Rightarrow P\}$ *and* $\#$ *means "cardinality". The notation* $co(P)$ *denotes the confidence 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* $D^n \to true \times [0, 1]$, $I_G(P) = (true, co(P))$.

The general interpretation is naturally extended to constraint formulae, prefixed or not by temporal connectives. There is only one exception: for temporal rules the confidence is calculated as the limit of the 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). The reason for this choice is related to the presence of incomplete states, where the interpretation for the implicated clause cannot be calculated.

Definition 7 *The confidence of a temporal rule* $H_1 \wedge \cdots \wedge H_m \mapsto H_{m+1}$ *is the limit* $\lim_{n\to\infty} \frac{\#A}{\#B}$, *where* $A = \{i \in \{0, \ldots, n\} | i \Rightarrow H_1 \wedge \cdots \wedge H_m \wedge H_{m+1}\}$ *and* $B = \{i \in \{0, \ldots, n\} | i \Rightarrow H_1 \wedge \cdots \wedge H_m\}$.

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* $M = (S, x, \Im)$, *a model for M is a structure* $\tilde{M} = (\tilde{T}, \tilde{x})$ *where* $\tilde{T}$ *is a finite temporal domain* $\{i_1, \ldots, i_n\}$, $\tilde{x}$ *is the subsequence of states* $\{x_{i_1}, \ldots, x_{i_n}\}$ *(the restriction of* x *to the temporal domain* $\tilde{T}$*) and for each* $i_j, j = 1, \ldots, n$, *the state* x_{i_j} *is a complete state.*

Now we may define the estimation 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,* $\tilde{I}_G(P)$, *is a function* $D^n \to true \times [0, 1]$, *assigning to P the value* true *with a confidence equal to the ratio* $\frac{\#A}{\#\tilde{T}}$, *where* $A = \{i \in \tilde{T} | i \Rightarrow P\}$. *The notation* $co(P, \tilde{M})$ *will denote the estimated confidence of P, given* $\tilde{M}$.

The estimation of the confidence of a temporal rule, giving a model, is defined as:

Definition 10 *Given a model* $\tilde{M} = (\tilde{T}, \tilde{x})$ *for M, the estimation of the confidence of the temporal rule* $H_1 \wedge \cdots \wedge H_m \mapsto H_{m+1}$ *is the ratio* $\frac{\#A}{\#B}$, *where* $A = \{i \in \tilde{T} | i \Rightarrow H_1 \wedge \cdots \wedge H_m \wedge H_{m+1}\}$ *and* $B = \{i \in \tilde{T} | i \Rightarrow H_1 \wedge \cdots \wedge H_m\}$.

3 A General Methodology

A general methodology for temporal rules extraction may be structured in two phases. The first, called discretisation phase, transforms sequential raw data into sequences of events. Practically, this means to establish the set of events, identified by names (constant symbol t_1), and the set of features, common for all events (function symbols $t_2, \ldots, t_n$). As example, consider a database containing the daily price variations of a given stock. After the application of the first phase we obtain an ordered sequence of events. Each event has the form $(name, v_1,\ v_2)$, where the name is one of the strings {*peak, plateau, valley*} and $v_1,\ v_2$ represents the maximum variation, respectively, the standard error of the daily prices.

In the frame of our formalism, during this phase, we establish the set of temporal atoms which can be defined syntactically in L. In the above example, an event is defined as $E(t_1, t_2, t_3)$. Semantically, the domain D_e is defined as {*peak, plateau, valley*} and the domain D_f as $\Re^+$ (the prices are positives real numbers and the features are statistical functions). Furthermore, during this phase, a linear time structure M is defined: at each time moment i, the state s_i contains as information the set of events started at i (see Table 1).

The second phase, called inference phase, extracts temporal rules from the ordered sequence of events. This phase is divided in two steps. The first step consists in applying an induction process on a model $\tilde{M}$ to obtain temporal rules. This step can be repeated for different models $\tilde{M}_i$ (see Table 2). Different approaches – Association Rules, Inductive Logic Programming, Classification Trees [5] – can be applied to extract rules from a database of events.

The second step consists in applying an inference process, using the previously inferred temporal rules, to obtain the final set of temporal meta-rules. These rules are temporal rules in accordance with the Definition 3, but supposed to have a small variability of the estimated confidence among different models. Therefore, such a rule may be applied with the same confidence in any state, complete or incomplete. The process of inferring temporal meta-rules is related to a new approach in data mining, called higher order mining, i.e. mining from the results of previous mining results. The formalism we proposed does not impose what approach must be use to discover first order temporal rules (rules generated by the induction process). As long as these rules may be expressed according to the Definition 3, the strategy (including algorithms, criterions, statistical methods) developed to infer temporal meta-rules remains valid.

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

State	Events	State	Events	State	Events
s_1	$E(peak, 10, 1.5)$	s_3	$E(plateau, 3, 0.2)$	s_5	$E(valley, 15, 1.9)$
s_2	$E(peak, 12, 1.2)$	s_4	$E(plateau, 1, 0.5)$	s_6	$E(peak, 6, 1.1)$

Table 2. Two temporal rules extracted from two models $\tilde{M}$ using the induction process

Model	Temporal Rule
$s_1 \dots s_{100}$	$X_{-4}(t_1 = peak) \wedge X_{-4}(t_2 < 11) \wedge X_{-3}(t_1 = peak) \mapsto X_0(t_1 = valley)$
$s_{300} \dots s_{399}$	$X_{-2}(t_1 = peak) \wedge X_{-2}(t_3 < 1.1) \wedge X_{-1}(t_1 = plateau) \wedge \mapsto X_0(t_1 = valley)$

4 Temporal Meta-rules

Suppose that we dispose of a set of temporal rules corresponding to a given model $\tilde{M}$. It is very likely that some temporal rules contain constraint formulae that are irrelevant, i.e. by deleting these relations, the general interpretation of the rules remain unchanged. In the frame of a consistent time structure M, it is obvious that we cannot delete a relation from a temporal rule (noted TR) if the resulting temporal rule (noted TR^-) has a lower confidence. But for a given model $\tilde{M}$, we calculate an estimate of $co(TR)$, which is $co(TR, \tilde{M})$. Because this estimator has a binomial distribution, we may calculate a confidence interval for $co(TR)$ and, consequently, we accept to delete a relation from TR if and only if the lower confidence limit of $co(TR^-, \tilde{M})$ is greater than the lower confidence limit of $co(TR, \tilde{M})$.

The estimator $co(TR, \tilde{M})$ being a ratio, $\#A/\#B$, a confidence interval for this value is constructed using a normal distribution with mean $\pi = \#A/\#B$ and variance $\sigma^2 = \pi(1-\pi)/\#B$). The lower limit of the interval is $L_\alpha(A, B) = \pi - z_\alpha \sigma$, where z_α is the quantile of the standard normal distribution for a given confidence level α (usually, 0.95). The following algorithm generalizes a single temporal rule TR, by deleting one relation.

Algorithm 1 ***Generalization 1-delete***

Step 1 *Let* $TR = H_1 \wedge \dots \wedge H_m \mapsto H_{m+1}$. *Let* $\aleph = \cup C_j$, *where* C_j *are all relations that appear in the constraint formulae of the implication clauses. Rewrite TR, by an abuse of notation, as* $\aleph \mapsto H_{m+1}$. *If* $n = \#\aleph$, *denote by* $C_1, \dots, C_n$ *the list of all relations from* $\aleph$.

Step 2 *For each* $i = 1, \dots, n$ *do*

$\aleph^- = \aleph - C_i, \quad TR_i^- = \aleph^- \mapsto H_{m+1}$

$A = \{i \in \tilde{T} | i \Rightarrow \aleph \wedge H_{m+1}\}, \quad B = \{i \in \tilde{T} | i \Rightarrow \aleph\}$

$A^- = \{i \in \tilde{T} | i \Rightarrow \aleph^- \wedge H_{m+1}\}, \quad B^- = \{i \in \tilde{T} | i \Rightarrow \aleph^-\}$

$co(TR, \tilde{M}) = \#A/\#B, \quad co(TR_i^-, \tilde{M}) = \#A^-/\#B^-$

If $L_\alpha(A, B) \leq L_\alpha(A^-, B^-)$ *then store* TR_i^-

Step 3 *Keep the generalized temporal rule* TR_i^- *for which* $L_\alpha(A^-, B^-)$ *is maximal.*

The complexity of this algorithm is linear in n. Using the criterion of lower confidence limit, (or LCL), we define the temporal meta-rule inferred from TR as the temporal rule with a maximum set of relations deleted from $\aleph$ and having the maximum lower confidence limit greater than $L_\alpha(A, B)$. An algorithm

Table 3. Parameters calculated in Step 2 of Algorithm 1 by deleting one relation

Deleted relation	$\#A^-$	$\#B^-$	$co(TR_i^-, \tilde{M})$	$L_\alpha(A^-, B^-)$
$X_{-4}(t_1 = peak)$	24	49	0.489	0.349
$X_{-4}(t_2 < 11)$	30	50	0.60	0.464
$X_{-3}(t_1 = peak)$	22	48	0.458	0.317

designed to find the largest subset of relations that can be deleted will have an exponential complexity. A solution is to use Algorithm 1 in successive steps until no more deletion is possible, but without having the guarantee to get the global maximum.

As example, consider the first temporal rule of Table 1 and suppose that $\#A = 20$ and $\#B = 40$ (giving an estimate $co(TR, \tilde{M}) = 0.5$, with $L_{0.95}(0.5) = 0.345$). Looking at Table 3, we find two relations which could be deleted (the first and the second) with a maximum $L_\alpha(A^-, B^-)$ given by the second relation.

Suppose now that we dispose of two models, $\tilde{M}_1 = (\tilde{T}_1, \tilde{x}_1)$ and $\tilde{M}_2 = (\tilde{T}_2, \tilde{x}_2)$, and for each model we have a set of temporal rules with the same implicated clause H (sets denoted S_1, respectively S_2). Let S be a subset of the union $S_1 \cup S_2$. If $TR_j \in S$, $j = 1, \ldots, n$, $TR_j = H_1 \wedge \cdots \wedge H_{m_j} \mapsto H$, then consider the sets $A_j = \{i \in \tilde{T}_1 \cup \tilde{T}_2 | i \Rightarrow H_1 \wedge \ldots \wedge H_{m_j} \wedge H\}$, $\mathbf{A} = \cup A_j$, $B_j = \{i \in \tilde{T}_1 \cup \tilde{T}_2 | i \Rightarrow H_1 \wedge \ldots \wedge H_{m_j}\}$, $\mathbf{B} = \cup B_j$ and $\mathbf{C} = \{i \in \tilde{T}_1 \cup \tilde{T}_2 | i \Rightarrow H\}$. The performance of the subset S can be summarized by the number of *false positives* (time instants where the implication clauses of each temporal rule from S are true, but not the clause H) and the number of *false negatives* (time instants where the clause H is true, but not at least one of the implication clauses of the temporal rules from S). Practically, the number of *false positives* is $fp = \#(\mathbf{B} - \mathbf{A})$ and the number of *false negatives* is $fn = \#(\mathbf{C} - \mathbf{B})$. The worth of the subset S of temporal rules is assessed using the Minimum Description Length Principle (MDLP)[17]. This provides a basis for offsetting the accuracy of a theory (here, a subset of temporal rules) against its complexity. The principle is simple: a Sender and a Receiver have both the same models $\tilde{M}_1$ and $\tilde{M}_2$, but the states of the models of the Receiver are incomplete states (the interpretation of the implicated clause cannot be calculated). The sender must communicate the missing information to the Receiver by transmitting a theory together with the exceptions to this theory. He may choose either a simple theory with a great number of exceptions or a complex theory with fewer exceptions. The MLD Principle states that the best theory will minimize the number of bits required to encode the total message consisting of the theory together with its associated exceptions.

To encode a temporal rule from S, we must specify its implication clauses (the clause H is the same for all rules, so is no need to encoded it). Because the order of the implication clauses is not important, the number of required bits is reduced by $\kappa \log_2(m!)$, where m is the number of implication clauses and κ is a constant depending on the encoding procedure. The number of bits required to encode the set S is the sum of encoding length for each temporal rule from S reduced by $\kappa \log_2(n!)$ (the order of the n temporal rules from S is not important). The exceptions are encoded by indicating the sets *false positive* and *false negative*. If $b = \#\mathbf{B}$ and $N = \#(\tilde{T}_1 + \tilde{T}_2)$ then the number of bits required is $\kappa \log_2\left(\binom{b}{fp}\right) + \kappa \log_2\left(\binom{N-b}{fn}\right)$, because we have $\binom{b}{fp}$ possibilities to choose the *false positives* among the cases covered by the rules and $\binom{N-b}{fn}$ possibilities to indicate the *false negatives* among the uncovered cases. The total number of bits required to encode the message is then equal to *theory bits + exceptions bits*.

Using the criterion of MDLP, we define as temporal meta-rules inferred from a set of temporal rules (implying the same clause and extracted from at least two different models), the subset S that minimizes the total encoding length. The algorithm that find this subset S has again an exponential complexity, but in practice we may use different non-optimal strategies (hill-climbing, genetic algorithms, simulated annealing), having a polynomial complexity.

Because the two definitions of temporal meta-rules differ not only in criterion (LCL, respectively MLDP), but also in the number of initial models (one, respectively at least two), the second inference process is applied in two steps. During the first step, temporal meta-rules are inferred from each set of temporal rules based on a single model. During the second step, temporal meta-rules are inferred from each set of temporal rules created during the step one and having the same implicated clause. There is another reason to apply firstly the LCL criterion: the resulted temporal meta-rules are less redundant concerning the set of implication clauses and so the encoding procedures, used by MLDP criterion, don't need an adjustment against this effect.

5 Conclusions

In this article we constructed a theoretical framework to discover knowledge, represented in the form of general Horn clauses, inferred from databases with a temporal dimension. This framework, based on first-order temporal logic, permits to define the main notions (event, temporal rule, constraint) in a formal way. The concept of consistent linear time structure allows us to introduce the notions of *general interpretation* and of *confidence*.

Also included in the proposed framework, the process of inferring temporal meta-rules is related to a new approach in data mining, called *higher order mining*, i.e. mining from the results of previous mining runs. According to this approach, the rules generated by the first induction process are first order rules and those generated by the second inference process (i.e temporal meta-rules) are higher order rules. Our formalism does not impose which methodology must be used to discover first order rules. As long as these rules may be expressed according to Definition 3, the strategy (here including algorithms, criterions, statistical methods), developed to infer temporal meta-rules may be applied.

It is important to mention that the condition of the existence of the limit, in the definition of a consistent linear time structure, is a fundamental one: it expresses the fact that the structure M represents a homogenous model and therefore the conclusions (or inferences) based on a finite model $\tilde{M}$ for M are consistent. However, at this moment, we do not know any methods which may certify that a given model is consistent. In our opinion, the only feasible approach to this problem is the development of methods and procedures for detecting the change points in the model. In this sense, the analysis of the evolution of temporal meta-rules seems to ba a very promising starting point.

References

1. R. Agrawal, K. I. Lin, H. S. Sawhney, and K. Smith. Fast Similarity Search in the Presence of Noise, Scaling and Transaction in Time-Series Databases. In *Proc. of the* 21^{th} *Conf. on VLDB*, pages 490–501, 1995.
2. S. Al-Naemi. A Theoretical Framework for Temporal Knowledge Discovery. In *Proc. of Int. Workshop on Spatio-Temporal Databases*, pages 23–33, Spain, 1994.
3. X. Chen and I. Petrounias. A Framework for Temporal Data Mining. In *Proc. of the* 9^{th} *Int. Conf. DESA*, pages 796–805, L.N.C.S. 1460, 1998.
4. X. Chen and I. Petrounias. Discovering Temporal Association Rules: Algorithms, Language and System. In *Proc. of the* 6^{th} *ICDE*, page 306, San Diego, USA, 2000.
5. P. Cotofrei and K. Stoffel. Classification Rules + Time = Temporal Rules. In *Proc. of ICCS*, pages 572–581. Lecture Notes in Computer Science, 2329, 2002.
6. G. Das, K. Lin, H. Mannila, G. Renganathan, and P. Smyth. Rule Discovery from Time Series. In *Proc. of the* 4^{th} *Conf. KDD*, pages 16–22, 1998.
7. N. Friedman, K. Murphy, and S. Russel. Learning the structure of dynamic probabilistic networks. In *Proc. of the* 14^{th} *Conf. on Uncertainty in AI*, pages 139–147. AAAI Press, 1998.
8. G. Guimares. Temporal Knowledge Discovery for multivariate time series with enhanced self-organizing maps. In *Proceedings of the IEEE-INNS-ENNS Int. Joint Conference on Neural Networks*, pages 165–170. IEEE Computer Society, 2000.
9. J. Han, G. Dong, and Y. Yin. Efficient Mining of Partial Periodic Patterns in Time Series Database. In *Proc. of ICDE*, pages 106–115, Sydeny, Australia, 1999.
10. J. Han, W. Gong, and Y. Yin. Mining Segment-Wise Periodic Patterns in Time-Related Databases. In *Proc. of the* 4^{th} *Conf. KDD*, pages 214–218, 1998.
11. F. Hoppner. Learning Temporal Rules from State Sequences. In *IJCAI Workshop on Learning from Temporal and Spatial Data*, pages 25–31, Seattle, USA, 2001.
12. E. J. Keogh, S. Chu, D. Hart, and M. J. Pazzani. Iterative Deepening Dynamic Time Warping for Time Series. In *Proc. of* 2^{th} *SIAM Conf. on Data Mining*, 2002.
13. W. Lin, M. A. Orgun, and G. J. Williams. Temporal Data Mining using Hidden Markov-Local Polynomial Models. In *LNCS*, volume 2035, pages 324–335, 2001.
14. H. Loether and D. McTavish. *Descriptive and Inferential Statistics: An introduction.* Allyn and Bacon, 1993.
15. D. Malerba, F. Esposito, and F. Lisi. A Logical Framework for Frequent Pattern Discovery in Spatial Data. In *Proc. of* 5^{th} *Conf. KDD*, pages 53–62, 2001.
16. B. Ozden, S. Ramaswamy, and A. Silberschatz. Cyclic association rules. In *Proc. Of Int. Conf. on Data Engineering*, pages 412–421, Orlando, USA, 1998.
17. J. Rissanen. Modelling by shortest data description. *Automatica*, 14:465–471, 1978.
18. M. Spiliopoulou and J. Roddick. Higher Order Mining: Modeling and Mining the Results of Knowledge Discovery. In *Proc. Of the* 2^{nd} *Int. Conf. on Data Mining Methods and Databases*, pages 309–320, UK, 2000.