DISCOVERING N-ARY TIMED RELATIONS FROM SEQUENCES

Nabil Benayadi and Marc Le Goc
LSIS Laboratory, University Saint Jerome, Marseille, France

Keywords: Markov Chain, Information Theory, Kullback-Leibler Distance.

Abstract: The goal of this position paper is to show the problems with most used timed data mining techniques for discovering temporal knowledge from a set of timed messages sequences. We will present from a simple example that Apriori-like algorithms for mining sequences as Minepi and Winepi fail for mining a simple sequence generated by a very simple process. Consequently, they cannot be applied to mine sequences generated by complexes process as blast furnace process. We will show also that another technique called TOM4L(Timed Observations Mining for Learning) can be used for mining such sequences and generate significantly better results than produced by Apriori-like techniques. The results obtained with an application on very complex real world system is presented to show the operational character of the TOM4L.

1 INTRODUCTION

When supervising and monitoring dynamic processes, a very large amount of timed messages (alarms or simple records) are generated and collected in databases. Mining these databases allows discovering the underlying relations between the variables that govern the dynamic of the process.

Apriori-like techniques (Roddick and Spiliopoulou, 2002) are probably the most used methods for discovering temporal knowledge in timed messages sequences. The basic principle of these approaches uses a representativeness criterion; typically the support to build the minimal set of sequential patterns that describes the given set of sequences. The support $s(p_i)$ of a pattern p_i is the number of sequences in the set of sequences where the pattern p_i is observed. A frequent pattern is a pattern p_i with a support $s(p_i)$ greater than a user defined thresholds S, $s(p_i) \geq S$. A frequent pattern is interpreted as a regularity or a condensed representation of the given set of sequences.

The Timed Data Mining techniques differ depending on whether the initial set of sequences is a singleton or not. The second case is the simpler because the decision criterion based on the support is directly applicable to a set of sequences. When the initial set of sequences contains a unique sequence, the notion of windows has been introduced to define an adapted notion of support. The first way consists in defining a fixed size of windows that an algorithm like Winepi (Mannila et al., 1997) shifts along the sequence: the sequence becomes then a set of equal length subsequences and the support $s(p_i)$ of a pattern can be computed. The second way consists in building a window for an a priori given pattern p_i. With the Minepi algorithm for example (Mannila et al., 1997), a window $W = [t_s, t_e[$ is a minimal occurrence of p_i if p_i occurs in W and not in any sub-window of W. In practice, a maximal window size parameter *maxwin* must be defined to bound the search space of patterns. Unfortunately, these approaches present two main problems. The first is that the algorithms require the setting of a set of parameters: the discovered patterns depends therefore of the tuning of the algorithms (Mannila, 2002). The second problem is the number of generated patterns that is not linear with threshold value S of the decision criterion $s(p_i) \geq S$. In practice, to obtain an interesting set of frequent pattern, S must be small. Consequently, the number of frequent is huge. Practically, only a very small fraction of the discovered patterns are interesting.

We will show in this paper that the TOM4L approach (Bouché, 2005; Le Goc, 2006) can avoid these two problems with the use of a stochastic representation of a given set of sequences on which an inductive reasoning coupled with a deductive reasoning is applied to reduce the space search.

The next section presents a (very) simple illustrative example and shows these two main problems

Benayadi N. and Le Goc M. (2010).
DISCOVERING N-ARY TIMED RELATIONS FROM SEQUENCES.
In *Proceedings of the 2nd International Conference on Agents and Artificial Intelligence - Artificial Intelligence*, pages 428-433
DOI: 10.5220/0002762304280433

for the most used algorithms for discovering temporal knowledge in timed messages sequences, Winepi and Minepi algorithms. Next, section 3 introduces the basis of the TOM4L process and the section 4 describes the results obtained with an application of the TOM4L on very complex real system monitored with a large scale knowledge based system, the Sachem system of the Arcelor-Mittal Steel group. Section 5 concludes the paper.

2 AN ILLUSTRATIVE EXAMPLE

The illustrative example is a simple dynamic SISO (Single Input Single Output) system $y(t) = F \cdot x(t)$ where F is a convolution operator. This example is used through this paper to illustrate the claims.
Let us defining two thresholds ψ_x and ψ_y for the input variable $x(t)$ and the output variable $y(t)$. These two thresholds respectively define two ranges for each of the variables: $rx_0 =]-\infty, \psi_x]$, $rx_1 =]\psi_x, +\infty]$, $ry_0 =]-\infty, \psi_y]$ and $ry_1 =]\psi_y, +\infty]$). Let us suppose that there exists a (very simple) program that writes a constant when a signal enter in a range. Such a program writes the constant 1 (resp. H) when $x(t)$ (resp. $y(t)$) enters in the range rx_1 (resp. ry_1) and 0 (resp. L) when $x(t)$ (resp. $y(t)$) enters in the range rx_0 (resp. ry_0). The evolution of the $x(t)$ in the figure 1 leads to the following sequence: $\Omega = \{(1,t_1), (H,t_2), (0,t_3), (L,t_4), (1,t_5), (H,t_6), (0,t_7), (L,t_8), (1,t_9), (H,t_{10}), (0,t_{11}), (L,t_{12}), (1,t_{13}), (H,t_{14}), (0,t_{15}), (L,t_{16}), (1,t_{17}), (H,t_{18}), (0,t_{19}), (L,t_{20}), (1,t_{21}), (H,t_{22}), (0,t_{23}), (L,t_{24})\}$.

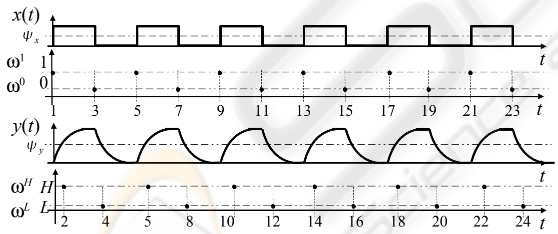

Figure 1: Temporal evolution of variables x and y.

To illustrate the sensitivity of the Winepi and the Minepi algorithms to the parameters, we define two sets of parameters and apply the algorithms to the sequence Ω. In the first set of parameters, the window width w and window movement v for Winepi are both set to 4 (this is the ideal tuning) and for Minepi, the max window is set to 4 and the minimal frequency is fixed to 6 (this is also the ideal tuning). In the second set of parameters, the window width and window movement of Winepi are equal to 8 and the support is equal to 3. The minimal frequency for Minepi is set to 8. The table 1 provides the number of patterns discovered by each algorithm with the two sets of parameters.

These two experimentation show the sensitivity of the Winepi and the Minepi algorithms with the parameters: from the first set to the second, the number of patterns increases of more than 626% for Winepi, and more than 18666% from Minepi. The main problem is the too large number of discovered patterns. The paradox is then the following: to find the ideal set of parameters that minimizes the number of discovered patterns, the user must know the system while this is precisely the global aim of the Data Mining techniques. There is then a crucial need for another type of approach that is able to provide a good solution for such a simple system and provide operational solutions for real world systems. The aim of this paper is to propose such an approach: the TOM4L process (Timed Observation Mining for Learning) finds only four relations with the example without any parameters.

Table 1: Number of Discovered Patterns.

	Winepi	Minepi
First parameter set	15	15
Second parameter set	94	2800

3 BASIS OF THE TOM4L PROCESS

The TOM4L process is based on the Theory of Timed Observations of (Le Goc, 2006) that provides the mathematical foundations of the four steps Timed Data Mining process that reverses the usual Data Mining process in order to minimize the size of the set of the discovered patterns:

1. Stochastic Representation of a set of sequences $\Omega = \{\omega_i\}$. This step produces a set of timed binary relations of the form $R_{i,j}(C^i, C^j, [\tau^-_{i,j}, \tau^+_{i,j}])$.
2. Induction of a minimal set of timed binary relations. This step uses an interestingness criterion based on the BJ-measure describes in the following section.
3. Deduction of a minimal set of n-ary relations. This step uses an abductive reasoning to build a set of n-ary relations that have some interest according to a particular problem.
4. Find the minimal set of n-ary relations being representatives according to the problem. This step corresponds to the usual search step of sequential patterns in a set of sequences in Minepi or Winepi.

The discovered n-ary relations discovered in the last step are called signatures. The next section provides the basic definitions of the Timed Observations Theory.

3.1 Basic Definitions (Le Goc, 2006)

A discrete event e_i is a couple (x_i, δ_i) where x_i is the name of a variable and δ_i is a constant. The constant δ_i denotes an abstract value that can be assigned to the variable x_i. The illustrative example allows the definition of a set E of four discrete events: $E = \{e_1 \equiv (x,1)$, $e_2 \equiv (x,0)$, $e_3 \equiv (y,H)$, $e_4 \equiv (y,L)\}$. A discrete event class $C^i = \{e_i\}$ is an arbitrary set of discrete event $e_i = (x_i, \delta_i)$. Generally, and this will be true in the suite of the paper, the discrete event classes are defined as singletons because when the constants δ_i are independent, two discrete event classes $C^i = \{(x_i, \delta_i)\}$ and $C^j = \{(x_j, \delta_j)\}$ are only linked with the variables x_i and x_j. The illustrative example allows the definition of a set Cl of four discrete event classes: $Cl = \{C^1 = \{e_1\}, C^0 = \{e_0\}, C^L = \{e_L\}, C^H = \{e_H\}\}$.
An occurrence $o(k)$ of a discrete event class $C^i = \{e_i\}$, $e_i = (x_i, \delta_i)$, is a triple (x_i, δ_i, t_k) where t_k is the time of the occurrence. When useful, the rewriting rule $o(k) \equiv (x_i, \delta_i, t_k) \equiv C^i(k)$ will be used in the following. A sequence $\Omega = \{o(k)\}_{k=1\ldots n}$, is an ordered set of n occurrences $C^i(k) \equiv (x_i, \delta_i, t_k)$. The illustrative example defines the following sequence: $\Omega = \{(C^1(1)$, $C^H(2)$, $C^0(3)$, $C^L(4)$, $C^1(5)$, $C^H(6)$, $C^0(7)$, $C^L(8)$, $C^1(9)$, $C^H(10)$, $C^0(11)$, $C^L(12)$, $C^1(13)$, $C^H(14)$, $C^0(15)$, $C^L(16)$, $C^1(17)$, $C^H(18)$, $C^0(19)$, $C^L(20)$, $C^1(21)$, $C^H(22)$, $C^0(23)$, $C^L(24)\}$.

Le Goc (Le Goc, 2006) shows that when the constants $\delta_i \in \Delta$ are independent, a sequence $\Omega = \{o(k)\}$ defining a set $Cl = \{C^i\}$ of m classes is the superposition of m sequences $\omega_i = \{C^i(k)\}$:

$$\Omega = \{o(k)\} = \bigcup_{i=1\ldots m} \omega^i = \{C^i(k)\} \tag{1}$$

The Ω sequence of the illustrative example is then the superposition of four sequences $\omega_i = \{C^i(k)\}$:
$\omega^1 = \{C^1(1), C^1(5), C^1(9), C^1(13), C^1(17), C^1(21)\}$
$\omega^0 = \{C^0(3), C^0(7), C^0(11), C^0(15), C^0(19), C^0(23)\}$
$\omega^L = \{C^L(4), C^L(8), C^L(12), C^L(16), C^L(20), C^L(24)\}$
$\omega^H = \{C^H(2), C^H(6), C^H(10), C^H(14), C^H(18), C^H(22)\}$

3.2 Stochastic Representation

The stochastic representation transforms a set of sequences $\omega_i = \{o(k)\}$ in a Markov chain $X = (X(t_k); k > 0)$ where the state space $Q = \{q_i\}$, $i = 1\ldots m$, of X is confused with the set of m classes $Cl = \{C^i\}$ of $\Omega = \bigcup_i \omega_i$.
Consequently, two successive occurrences $(C^i(k-1)$, $C^j(k))$ correspond to a state transition in X: $X(t_{k-1}) = q_i \longrightarrow X(t_k) = q_j$. The conditional probability $P\left[X(t_k) = q_j | X(t_{k-1}) = q_i\right]$ of the transition from a state q_i to a state q_j in X corresponds then to the conditional probability $P\left[C^j(k) \in \Omega | C^i(k-1) \in \Omega\right]$ of observing an occurrence of the class C^j at time t_k knowing that an occurrence of a class C^i at time t_{k-1} has been observed: The transition probability matrix $P = [p_{i,j}]$ of X is computed from the contingency table $N = [n_{i,j}]$, where $n_{i,j} \in N$ is the number of couples $(C^i(k), C^j(k+1))$ in Ω. For example, the table 2 is the contingency table N of the sequence Ω of the illustrative example.

Table 2: Contingency table $N = [n_{i,j}]$ of Ω.

	C^1	C^0	C^H	C^L	Total
C^1	0	0	6	0	6
C^0	0	0	0	6	6
C^H	0	6	0	0	6
C^L	5	0	0	0	5
Total	5	6	6	6	23

The stochastic representation of a given set Ω of sequences is then the definition of a set $R = \{R_{i,j}(C^i, C^j, [\tau^-_{ij}, \tau^+_{ij}])\}$ where each the conditional probability $p_{i,j} = P\left[C^j(k) \in \Omega | C^i(k-1) \in \Omega\right]$ of each binary relation $R_{i,j}(C^i, C^j, [\tau^-_{ij}, \tau^+_{ij}])$ is not null. The timed constrains $[\tau^-_{ij}, \tau^+_{ij}]$ is provided by a function of the set D of delays $D = \{d_{ij}\} = \{(t_{k_j} - t_{k_i})\}$ computed from the binary superposition of the sequences $\omega^{i,j} = \omega^i \cup \omega^j$: $\tau^-_{ij} = f^-(D), \tau^+_{ij} = f^+(D)$. For example, the authors of (Bouché, 2005) use the properties of the Poisson law to compute the timed constraints: $\tau^-_{ij} = 0, \tau^+_{ij} = \frac{1}{\lambda_{i,j}}$ where $\lambda_{i,j}$ is the Poisson rate (the exponential intensity) of the exponential law that is the average delay $d^{ij}_{moy} = \frac{\sum(d_{ij})}{Card(D)}$.
The set R of the illustrative example is the following: $R = \{R_{1,H}(C^1, C^H, [\tau^-_{1,H}, \tau^+_{1,H}])$, $R_{0,L}(C^0, C^L, [\tau^-_{0,L}, \tau^+_{0,L}])$, $R_{H,0}(C^H, C^0, [\tau^-_{H,0}, \tau^+_{H,0}])$, $R_{L,1}(C^L, C^1, [\tau^-_{L,1}, \tau^+_{L,1}])\}$.

3.3 Induction of Binary Relations

The induction step in TOM4L consists to select a subset I of R, $I \subseteq R$, where each relation $R_{i,j}(C^i, C^j)$ in I presents a potential interest. This selection is based on the definition of an interestingness measure of temporal binary relation $R_{i,j}(C^i, C^j)$, called BJ-Measure. The BJ-Measure of timed binary relation $R_{i,j}(C^i, C^j)$ is the adaptation of the kullback-Leibler distance to timed data (Benayadi and Goc, 2008). The BJ-Measure evaluates the information amount car-

ried by an observation of the class C^i to an observation of the class C^j. The principle of this adaptation is to consider the set of n binary relations as set of binary communication channels without memory (in the sense of Shannon, see Figure 2). Each binary relation $R_{i,j}(C^i,C^j) \in R$ is represented by a discrete binary memoryless channel linking two abstract binary variables X and Y, where $X(t_k) \in \{C^i,\overline{C^i}\}$ and $Y(t_{k+1}) \in \{C^j,\overline{C^j}\}$. The class $\overline{C^i}$ (resp. $\overline{C^j}$) represents all classes in Cl except C^i : $Cl - C^i$ (resp. $Cl - C^j$) where its occurrence number is the average of occurrence number of each classes in $\overline{C^i}$ (resp. $\overline{C^j}$). Therefore, any sequence $(C^i(k),C^j(k+1)) \in \Omega$ is an example of $R_{i,j}(C^i,C^j)$ and each $(C^i(k),\overline{C^j}(k+1)) \in \Omega$ or $(\overline{C^i}(k),C^j(k+1)) \in \Omega$ is counter-example.
Two adaptations of Kullback-Liebler distance are proposed to calculate the interestingness of the binary relation $R_{i,j}(C^i,C^j) \in R$, the first is from the point of view of class C^i, the BJ-measure in length, the other is from the point of view of C^j, the BJ-measure in width. The BJ-measure in length of a binary relation $R_{i,j}(C^i,C^j) \in R$, noted $BJL(C^i,C^j)$, measures the amount of information provided by an occurrence of the class C^i on an occurrence of class C^j or class $\overline{C^j}$:

- if $p(j|i) \geq p(j)$ then $BJL(C^i,C^j) = D(p(Y|C^i)\|p(Y))$
- else $BJL(C^i,C^j) = 0$

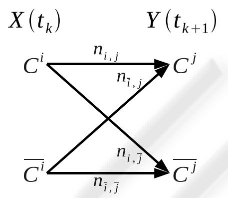

Figure 2: Two abstract binary variables connected by a discrete memoryless channel.

Symmetrically, the BJ-measure in width of a binary relation $R_{i,j}(C^i,C^j) \in R$, noted $BJW(C^i,C^j)$, measures the amount of information provided by an occurrence of the class C^i or $\overline{C^i}$ on an occurrence of the class C^j. The values of both measures are high when the informational contribution of the class C^i on C^j is strong. To have a global vision of the informational contribution of the class C^i on C^j, the two measures are combined into a single measure noted M. The measure M is simply the norm of the vector $\begin{pmatrix} BJL(C^i,C^j) \\ BJW(C^i,C^j) \end{pmatrix}$ normalized between 0 and 1.
For example, the values of the M-measure of the set $R = \{R_{1,H}(C^1,C^H,[\tau^-_{1,H},\tau^+_{1,H}]),\ R_{0,L}(C^0,C^L,[\tau^-_{0,L},\tau^+_{0,L}]),\ R_{H,0}(C^H,C^0,[\tau^-_{H,0},\tau^+_{H,0}]),\ R_{L,1}(C^L,C^1,[\tau^-_{L,1},\tau^+_{L,1}])\}$ of the illustrative example are given in table 3.

Table 3: Matrix M.

	C^1	C^0	C^H	C^L
C^1	0	0	1	0
C^0	0	0	0	1
C^H	0	1	0	0
C^L	1	0	0	0

Table 4: The M values evolution with different λ_{err}.

λ_{err}	$R(C^1,C^H)$	$R(C^H,C^0)$	$R(C^0,C^L)$	$R(C^L,C^1)$
0	1	1	1	1
6	0.75	0.56	1	0.63
12	0.78	0	1	0
18	0.61	0	0.79	0
24	0.55	0	0.55	0
30	0	0	0	0

In this example, the relations $R_{1,H}(C^1,C^H,[\tau^-_{1,H},\tau^+_{1,H}])$ and $R_{0,L}(C^0,C^L,[\tau^-_{0,L},\tau^+_{0,L}])$ have not the same meaning as the relations $R_{H,0}(C^H,C^0,[\tau^-_{H,0},\tau^+_{H,0}])$ $R_{L,1}(C^L,C^1,[\tau^-_{L,1},\tau^+_{L,1}])$: only the two first are linked with the system $y(t) = Fx(t)$, the two latters being only sequential relation (the system computes the values of $y(t)$, not the values of $x(t)$).

To distinguish between these two kinds of relations, the idea in the induction step is to add noise in the initial set of sequences. To this aim, we defined the "noisy" observation class C^{err} the occurrences of which are randomly timed. If a relation $R_{i,j}(C^i,C^j,[\tau^-_{i,j},\tau^+_{i,j}])$ is a property of the system, then the time interval between the occurrences of the C^i and C^j classes will be more regular than if this relation is a purely sequential relation. The table 4 shows the values of the M-measures of the relations $R(C^1,C^H)$, $R(C^H,C^0)$, $R(C^0,C^L)$ and $R(C^L,C^1)$ with different rate $\lambda_{err} = \dfrac{n_{err}}{t_{24}-t_0}$ of noisy occurrences added in Ω. The table 4 shows that when $\lambda_{err} \in \{12,24\}$, the binary relations $R(C^H,C^0)$ and $R(C^L,C^1)$ disappears. Naturally, when the noise is too strong ($\lambda_{err} = 30$), all the relations disappear: this means that at least one occurrence $C^{err}(k)$ is systematically inserted between two occurrences of the initial sequence Ω.
This example leads also to an operational property of the M-measure: when $\theta_{i,j} \gg 1$ or $\theta_{i,j} \ll 1$, one class plays the same role of a noisy class for the other. This situation arises in the two following cases:

- $n_{i,\bar{j}} \geq n_{i,j} \Rightarrow p(\bar{j}|i) \geq 0.5$. The $\overline{C^j}$ plays the role of a noisy class for the class C^i.

- $n_{\bar{i},j} \geq n_{i,j} \Rightarrow p(j|\bar{i}) \geq 0.5$. The $\overline{C^i}$ plays the role of a noisy class for the class C^j.

These two conditions are both evaluated when comparing the product $p(j|i) \cdot p(i|j)$ with $\frac{1}{2} \cdot \frac{1}{2}$: when $p(j|i) \cdot p(i|j) \leq \frac{1}{4}$, $M(C^i,C^j) \leq 0.5$ and the relation $R_{i,j}(C^i,C^j)$ cannot be justified with the M-measure. Inversely, when $p(j|i) \cdot p(i|j) > \frac{1}{4}$, $M(C^i,C^j) > 0.5$ and the relation $R_{i,j}(C^i,C^j)$ has some interest from the point of view of the M-measure. This leads to the following simple inducing rule that uses the M-measure as interestingness criteria:

$$M(C^i,C^j) > 0.5 \Rightarrow R_{i,j}(C^i,C^j) \in I \qquad (2)$$

So, the set I of induced binary relations contains only two binary relations : $I = \{R_{1,H}(C^1,C^H,[\tau^-_{1,H},\tau^+_{1,H}]), R_{0,L}(C^0,C^L,[\tau^-_{0,L},\tau^+_{0,L}])\}$

3.4 Deduction of n-ary Relations

The set I of binary relations contains then the minimal subset of R where each relation $R_{i,j}(C^i,C^j)$ presents a potential interest. From this set, the objective of the deduction step consists to deduce from I a small set $M = \{m^{k-1,n}\}$ of n-ary relations $m^{k-1,n}$ so that a search algorithm can be used effectively to identify the most representative relations $m^{k-1,n}$. To this aim, an heuristic $h(m^{i,n})$ is to select a minimal set $M = \{m^{k,n}\}$ of n-ary relations of the form $m^{k,n} = \{R_{i,i+1}(C^i,C^{i+1})\}, i = k,\cdots,n-1$, that is to say paths leading to a particular final observation class C^n. The heuristic $h(m^{i,n})$ makes a compromise between the generality and the quality of a path $m^{i,n}$:

$$h(m^{i,n}) = card(m^{i,n}) \times BJL(m^{i,n}) \times P(m^{i,n}) \qquad (3)$$

In this equation, $card(m^{i,n})$ is the number of relations in $m^{i,n}$, $BJL(m^{i,n})$ is the sum of the BJL-measures $BJL(C^{k-1},C^k)$ of each relation $R_{k-1,k}(C^{k-1},C^k)$ in $m^{i,n}$ and $P(m^{i,n})$ is the product of the probabilities associated with each relation in $m^{i,n}$.

$P(m^{i,n})$ corresponds to the Chapmann-Kolmogorov probability of a path in the transition matrix $P = [p(k-1,k)]$ of the Stochastic Representation. The interestingness heuristic $h(m^{i,n})$ being of the form $\phi \cdot \ln(\phi)$, it can be used to build all the paths $m^{i,n}$ where $h(m^{i,n})$ is maximum (Benayadi and Le Goc, 2008). For the illustrative example, the deduction step found a set M of two binary relations $(M = I)$[1].

[1] no paths containing more than one binary relation can be deduced from I

$$Cover\ Rate = \frac{number\ of\ target\ class\ occurrences\ predicted\ by\ the\ model}{Total\ number\ of\ target\ class\ occurrences}$$

$$Anticipation\ Rate = \frac{TP}{TP+FP}$$

$TP = True\ Positive\ Prediction \qquad FP = False\ Positive\ Prediction$

Figure 3: Evaluation Measures for n-ary relation.

3.5 Find Representativeness n-ary Relations

Given a set $M = \{m^{k,n})\}$ of paths $m^{k,n} = \{R_{i,i+1}(C^i,C^{i+1})\}$, $i = k,\cdots,n-1$, the fourth and final step of the discovery process TOM4L, step Find, uses two representativeness criterion (Cover Rate and Anticipation Rate, figure 3) to build the subset $S \subseteq M$ containing the pathes $m^{k,n}$ being representative according the initial set Ω of sequences. These paths are called *Signatures*.

Generally, a threshold equal to 50% is used to discard n-ary relations which have more false prediction than correct prediction. For example, the values of the cover rate and the anticipation rate of both binary relations of M of the illustrative example are 100%. So, $S = M$, $S = \{R_{1,H}(C^1,C^H,[\tau^-_{1,H},\tau^+_{1,H}]), R_{0,L}(C^0,C^L,[\tau^-_{0,L},\tau^+_{0,L}])\}$.
These signatures are the only relations (patterns) that are linked with the system $y(t) = Fx(t)$. Comparing with the set of patterns found by Apriori-like approaches, we can confirm from this illustrative example that TOM4L approach converges towards a minimal set of operational relations, which describe the dynamic of the process. In the next section, we present the application of TOM4L on a sequence generated by a very complex dynamic process, blast furnace process. Due to the process complexity, we can confirm, without experience, that Apriori-like approaches fail to mine this sequence.

4 APPLICATION

Our approach has been applied to sequences generated by knowledge-based system SACHEM developed to monitor, diagnose and control the blast furnace (Le Goc, 2006). We are interested with the *omega* variable that reveals the wrong management of the whole blast furnace. The studied sequence comes from Sachem at Fos-Sur-Mer (France) from 08/01/2001 to 31/12/2001. It contains 7682 occurrences of 45 discrete event classes (i.e. phenomena). For the 1463 class linked to the *omega* variable, the search space contains about $20^5 = 3,200,000$ binary relations. The inductive and the abductive reasoning steps of TOM4L produces a minimal set M of only 166 binary relations from which the set S of signatures of figure 5 have been discovered ($Ta = 50\%$ and $Tc = 10\%$). The set S is made with 50 binary relations.

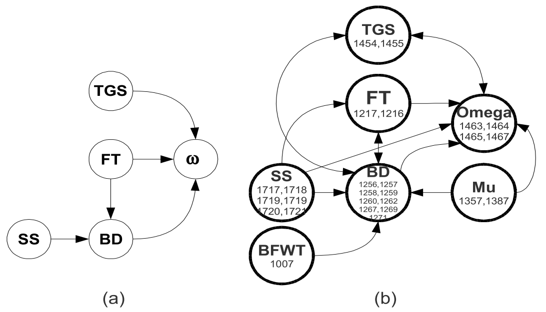

Figure 4: Expert's (1995, a) and discovered relations (2007, b).

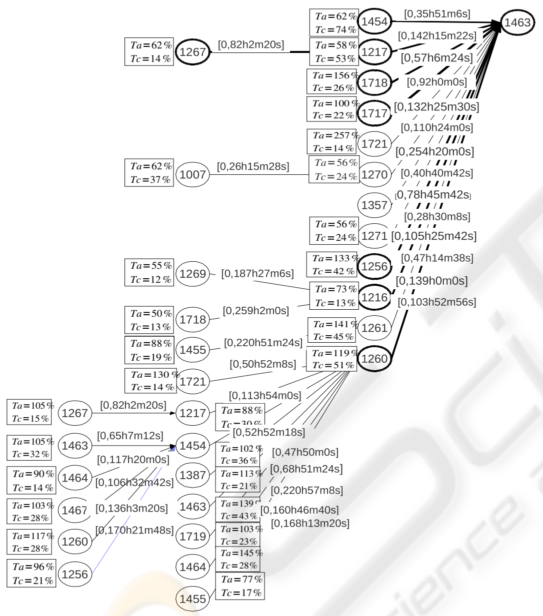

Figure 5: Part of signatures of 1463 class.

When substituting a class with its associated variable (the *omega* variable with the class 1464 for example) and the signatures of Figure 5 becomes the graph (b) of Figure 4 that contains the graph of the Expert's in 1995. Two variables appear in the graph (b): the experts agree that the blast furnace wall temperature $BFWT$ and the gas distribution over the burden *MuGlo* have an influence on the *BD* and the *omega* variables. It is to note that the similar result is obtained with other real world monitored processes.

As with the simple illustrative example of this paper, this result shows that the TOM4L process converges through a minimal set of binary relations with the elimination of the non interesting relations, despite the complexity of the monitored process.

5 CONCLUSIONS

We have presented from an illustrative example that Apriori-like algorithms as Minepi and Winepi can fail for mining sequences generated by a (very) simple process. Furthermore, we argue that these approaches cannot be applied to mine complexes process as blast furnace process. We have presented also that we can avoid all problems by applying TOM4L process, which is based on four steps: (1) a stochastic representation of a given set of sequences that is induced (2) a minimal set of timed binary relations, and an abductive reasoning (3) is then used to build a minimal set of n-ary relations that is used to find (4) the most representative n-ary relations according to the given set of sequences. The results obtained with an application on a very complex real world process (a blast furnace) are presented to show the operational character of the TOM4L process.

REFERENCES

Benayadi, N. and Goc, M. L. (2008). Discovering temporal knowledge from a crisscross of timed observations. In *The proceedings of the 18th European Conference on Artificial Intelligence (ECAI'08)*, University of Patras, Patras, Greece.

Benayadi, N. and Le Goc, M. (2008). Using a measure of the crisscross of series of timed observations to discover timed knowledge. *Proceedings of the 19th International Workshop on Principles of Diagnosis (DX'08)*.

Bouché, P. (2005). *Une approche stochastique de modélisation de séquences d'événements discrets pour le diagnostic des systèmes dynamiques*. These, Faculté des Sciences et Techniques de Saint Jérôme.

Le Goc, M. (2006). *Notion d'observation pour le diagnostic des processus dynamiques: Application à Sachem et à la découverte de connaissances temporelles*. Hdr, Faculté des Sciences et Techniques de Saint Jérôme.

Mannila, H. (2002). Local and global methods in data mining: Basic techniques and open problems. *29th International Colloquium on Automata, Languages and Programming*.

Mannila, H., Toivonen, H., and Verkamo, A. I. (1997). Discovery of frequent episodes in event sequences. *Data Mining and Knowledge Discovery*, 1(3):259–289.

Roddick, F. J. and Spiliopoulou, M. (2002). A survey of temporal knowledge discovery paradigms and methods. *IEEE Transactions on Knowledge and Data Engineering*, (14):750–767.