A GLOBAL MODEL OF SEQUENCES OF DISCRETE EVENT CLASS OCCURRENCES

Philippe Bouché[1], Marc Le Goc[2] and Jérome Coinu[3]
[1]LGECO - INSA de Strasbourg, 24 Bd de la Victoire, 67084 Strasbourg, France

[2]LSIS, UMR CNRS 6168, Université Aix-Marseille, Domaine Universitaire St Jérôme,13397 Marseille cedex 20, France

[3]2i, 9, Rue Gaston CASTEL, ZAC de Saumaty Séon B.P. 127, 13321 Marseille Cedex 16, France

Keywords: Monitored Control Systems, Knowledge Based Systems, Discrete Event Systems, Artificial Intelligence.

Abstract: This paper proposes a global model of a set of alarm sequences that are generated by knowledge based system monitoring a dynamic process. The modelling approach is based on the Stochastic Approach to discover timed relations between discrete event classes from the representation of a set of sequences under the dual form of a homogeneous continuous time Markov chain and a superposition of Poisson processes. An abductive reasoning on these representations allows discovering chronicle models that can be used as diagnosis rules. Such rules subsume a temporal model called the average time sequence that sums up the initial set of sequences. This paper presents this model and the role it play in the analysis of an industrial process monitored with a network of industrial automata.

1 INTRODUCTION

A Knowledge Based System (KBS) for monitoring a dynamic process aims at warning the operator(s) about the occurrences of unsatisfactory behaviors with a sequence of alarms. Such a situation is now the standard framework in most industries and one of the problems is the acquisition of knowledge about alarm correlations in dynamic systems.

The purpose of our work[1] is to define a method for discovering the timed relations between alarms to predict undesirable alarms. The alarms we are concerned with can be very high level alarms like Sachem's alarms, the generic KBS developed by the Arcelor-Mittal Group for monitoring its production tools (Le Goc, 2004), or low level alarms like PLC's alarms (Programmable Logic Controller or industrial automaton) for example. Experts are convinced that such timed relations exist but are not able to provide them because this kind of knowledge is intimately related to the dynamics of a monitored process: tools must then be defined to facilitate the discovery and quantification of the timed relations.

[1] This work has been partly financed by the 2i company under the contract n°120/06/04/2006.

To this aim, we develop the Stochastic Approach for discovering temporal knowledge from a set of sequences of discrete event class occurrences and represent this knowledge with abstract chronicle models. An abstract chronicle model is a set of binary relations between discrete event classes. Such a model is operational when it allows predicting an alarm before it occurs with a minimal confidence. In this case, such a model is called a *signature* of the alarm. This paper shows that this model corresponds to a global model of the set of sequences called the Average Time Sequence that can be used to reason about the couple made with the process and its monitoring KBS. The Average Time Sequence is then a new concept that fills one the mains problems of the Timed Data Mining techniques (Mannila, 2002).

The next section presents the main works related to the problem of discovering a predictive model of alarms. Section 3 introduces the basis of the Stochastic Approach framework we propose to tackle this problem and defines the concept of Average Time Sequence. The use of such a model is illustrated with a real world industrial process in section 4. The paper concludes on the operational aspects of the proposed approach.

Bouché P., Le Goc M. and Coinu J. (2008).
A GLOBAL MODEL OF SEQUENCES OF DISCRETE EVENT CLASS OCCURRENCES.
In *Proceedings of the Tenth International Conference on Enterprise Information Systems - AIDSS*, pages 173-180
DOI: 10.5220/0001680701730180

2 RELATED WORKS

The problem of discovering a signature from a sequence of alarms can be formulated in the following way: given a sequence ω, what is the abstract chronicle model that allows predicting the occurrences of a given discrete event class?

This problem has been for example tackle in the context of a telecommunication network (Cordier and Dousson, 2000; Dousson and Vu Duong, 1999). This approach is based on a frequency analysis of alarm logs in order to discover some frequent "patterns" of alarms that are represented under the form of "chronicles". This constitutes an application of Frequency Approach of the Data Mining domain to the content of timed data bases.

The Data Mining domain aims at defining tools and methods to discover knowledge from large data sets. The basic principle consists in identifying a minimal set of relations that characterize a data set. The different approaches are based on the "Apriori" algorithm. For example, (Agrawal and Al, 1993) propose a method to mine association rules from a large sequence of purchasing transactions carried out by a customer. A transaction is characterized by a set of bought and buys item times. The problem consists in finding a sequence of items called a pattern that is frequently observed in the transaction sequences. To this aim, the "Apriori" algorithm computes the support of a pattern as the number of times the pattern is observed in a given data base. Only patterns with a support greater than a minimal threshold are retained. This explains why this approach is called the Frequential Approach. This approach has been extended to sequential pattern through a set of algorithms like AprioriAll, AprioriSome and DynamicSome (Agrawal and Al, 1993).

(Manilla and al, 1997) propose another approach to discover temporal patterns, called "episodes", in a discrete events sequence corresponding to the alarms of a telecommunication network (Hatonen and Al, 1996). An episode is a collection of events that appear relatively close to each other in a partial order. The discovering process of temporal patterns is based on the frequency of an episode α in a sequence s, which is the fraction of the number of temporal windows in which the episode α occurs over the total number of temporal windows contained in the sequence s. The episode α having a frequency over a minimal threshold is then considered as a temporal pattern (Winepi and Minepi algorithms).

On another hand, in the Temporal Logic domain, Ghallab proposes the notion of chronicle model to represent a set of timed binary relations between events (Ghallab, 1996). A chronicle model is a kind of temporal pattern specification where nodes are events and links are timed binary constraints represented with [min, max] intervals. A chronicle model is a richer representation of temporal knowledge than an episode because it allows the adding of timed binary constraints between alarms. Ghallab's method for discovering chronicle models consists in splitting a set of event sequences in examples and counter examples and to order the sequences with the time of the events. When forgetting the times, the method determines the longest patterns that are common to the examples and that are not included in the counter examples. The timed constraints between events are then added by experts or computed with an ad'hoc algorithm.

Ghallab's method is not general because it supposes to be able to define what an example and a counter example are. With the Face algorithm, Dousson and Vu Duong (Dousson and Vu Duong, 1999) adapt the notion of chronicle models to the "Apriori" algorithm to discover recurrent chronicle models from a log of events but do not propose a sound method to evaluate the timed constraints.

According to (Manilla, 2002), the problem of discovering timed relations from a set of timed data is still an open problem. One of the reasons is the combination of logical relations and temporal constraints (Cauvin et Al, 1998; Hanks and McDermott, 1994). In particular, the relations provided by the Frequency Approach are local models of the studied sequences that are difficult to generalize.

The Stochastic Approach has been developed to tackle these difficulties and propose to discover abstract chronicle models from a sequence of alarms considered as occurrences of discrete event classes (Le Goc, 2004), (Bouché and Le Goc, 2004). This approach is based on the representation of a sequence of discrete alarms generated by a couple (Process, KBS) under the dual form of a homogeneous Markov chain and its superposition of Poisson processes. A set of tools have then been designed according to the Stochastic Approach and implemented in a Java environment called the "ELP Lab" (Le Goc and Al, 2006).

3 BASIS OF THE STOCHASTIC APPROACH

A sequence $\omega=\{o_k\}_{k=0,\ldots,m-1}$ is an ordered set of m occurrences $o_k\equiv(t_k, x, i)$ of discrete event $e_k\equiv(x, i)$, where $x\in X$ is the name of a discrete variable, $i\in I_x\subseteq\aleph$ is a discrete value of x and $t_k\in\Gamma=\{t_i\}$, $t_i\in\Re$, is the time of the assignation of the discrete value i to the variable x so that: $o_k\equiv(t_k, x, i) \Leftrightarrow x(t_k)=i$. The occurrences are timed with a continuous clock structure (i.e. $t_{k-2}-t_{k-1} \neq t_{k-1}-t_k$):

$$\begin{aligned}&\forall t_k \in \Re, \forall i \in \aleph, \exists t_{k-1} < t_k, \\ &x(t_{k-1}) \neq i \wedge x(t_k) = i \Rightarrow o_k \equiv (t_k, x, i)\end{aligned} \tag{1}$$

A couple (o_k, o_n) of two successive occurrences related to a variable x describes the modification of the values of the variable x over the interval $[t_k, t_n[$:

$$\begin{aligned}&\forall o_k \equiv (t_k, x, i), o_n \equiv (t_n, x, j), \\ &(o_k, o_n) \Rightarrow \forall t \in [t_k, t_n[, x(t) = i \wedge x(t_n) = j\end{aligned} \tag{2}$$

As a consequence, a sequence $\omega=\{o_k\}$ of discrete event occurrences $o_k\equiv(t_k, x, i)$ concerned with a variable x describes the temporal evolution of a discrete function $x(t)$ defined on $\aleph$.

$$\begin{aligned}&R(C^i, C^o, [\tau^-, \tau^+]) \Leftrightarrow \exists o_n, o_k \in \omega, \\ &(o_n :: C^o) \wedge (o_k :: C^i) \wedge (d(o_n) - d(o_k) \in [\tau^-, \tau^+]) \\ &\text{where } \forall o_k \equiv (t_k, x, i) \in \omega, d(o_k) = t_k\end{aligned} \tag{3}$$

A discrete event class is a set $C^j=\{e_i\}$ of discrete events $e_i\equiv(x, i)$. The notation "e_i::C^j" (resp. "o_k::C^j" or "C^j_k") denotes that the discrete event e_i (resp. the occurrence $o_k\equiv C^j_k$) belongs to the class C^j. A timed binary relation $R(C^i, C^o, [\tau^-, \tau^+])$ describes an oriented relation between two discrete event classes that is timed constrained. "$[\tau^-, \tau^+]$" is the time interval for observing an occurrence of the output class C^o after the occurrence of the input class C^i (equation (3)).

3.1 Abstract Chronicle Model

In this context, an abstract chronicle model is a set of binary relations with timed constraints between classes discrete events. Such a model is called an "ELP" model (ELP is the acronym of Event Language of Processing, (Le Goc and Al, 2006)). For example, the ELP model $M_{123}= \{R_{12}(C^1, C^2, [\tau_{12}^-, \tau_{12}^+]), R_{23}(C^2, C^3, [\tau_{23}^-, \tau_{23}^+])\}$ of Figure. 1 is made of two binary relations between three discrete event classes. A sequence ω satisfies the M_{123} ELP model when:

$$\begin{aligned}&\exists o_k, o_n, o_m \in \omega, (o_k :: C^1) \wedge (o_n :: C^2) \wedge (o_m :: C^3) \\ &\wedge (d(o_n) - d(o_k) \in [\tau_{12}^-, \tau_{12}^+]) \wedge (d(o_m) - d(o_n) \in [\tau_{23}^-, \tau_{23}^+])\end{aligned} \tag{4}$$

ELP models can be used to predict the occurrences of discrete event classes (like C^3 in the ELP model M_{123}) in an unknown sequence ω'.

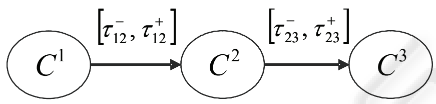

Figure 1: ELP representation of the M_{123} model

To this aim, rules of the equation (5) form can be used in a diagnosis task. When such a rule predicts an occurrence of a discrete event class with a minimal confidence, the corresponding ELP model is called a "signature" (Le Goc and Al, 2006).

$$\begin{aligned}&\forall \omega', \forall o_k, o_n \in \omega', \\ &(o_k :: C^1) \wedge (o_n :: C^2) \wedge (d(o_n) - d(o_k) \in [\tau_{12}^-, \tau_{12}^+]) \\ &\Rightarrow \exists o_m \in \omega', (o_m :: C^3) \wedge (d(o_m) - d(o_n) \in [\tau_{23}^-, \tau_{23}^+])\end{aligned} \tag{5}$$

To measure the confidence of such rules, we define the anticipating ratio of an abstract chronicle model as the number of sub sequences of a sequence ω that matches the complete abstract chronicle model, divided by the number of the sub sequences that matches the abstract chronicle model but without the final binary relation (the class C^3 in Figure 1). An abstract chronicle model is a signature when its anticipating ratio is equal to or greater than 50%.

3.2 Stochastic Representation

When the discrete event classes are independent and the distribution of the inter-occurrence times of a discrete event class complies with a Poisson law of the form $f(t)=1-e^{-\lambda t}$ (λ is the average number of occurrence in a unit of time and is called the Poisson rate(Cassandras and Lafortune, 2001)), the couple made with the process and its monitoring KBS can be considered as a stochastic discrete event generator (Le goc et Al, 2006). Consequently, a sequence of discrete event classes provided by such a generator can be represented under the dual form of a homogeneous Markov chain and its associated superposition of Poisson processes: a chronicle model is then connected with a specific path in the state space of the Markov Chain, and the timed relations will be provided by the corresponding superposition of Poisson processes.

To represent a sequence $\omega=(C^i_k)_{k\in K=\{0,\ldots,m\}}$ as a Markov chain $X=(X(t_k);\ k\in K)$, the set of discrete event classes $C^\omega=\{C^i\}_{i=0\ldots n-1}$ in ω is confused with the state space $Q=\{i\}_{i=0\ldots n-1}$ of X. A binary sub sequence $\omega'=(C^i_{k-1},\ C^j_k)$ of ω corresponds then to a state transition in X: $X(d(C^i_{k-1}))=i\rightarrow X(d(C^j_k))=j$, where d is the function providing the time of a class occurrence. A simple depth-first backward search algorithm (i.e. from an output class to input classes) is used to generate the tree of the most probable paths that lead to an output class. (Le Goc and Al, 2006)

This tree, along with the associated matrix, is a first representation of the sequence of alarms. This result is interesting because, whatever the length of the sequence of alarms, it is entirely contained in a finite matrix. The tree of sequential relations can then be used to produce a functional model of the couple (process, KBS) (Bouché and Al, 2006), or to find signatures of the form of the equation (5).

To constitute a timed binary relation of the form $R(C^i,\ C^j,\ [\tau^-,\ \tau^+])$, the timed constraint $[\tau^-,\ \tau^+]$ is simply added to the sequential relation $R_s(C^i,\ C^o)$. Such a timed constraint is related with the average delay $D_{i\text{-}j}=E[d(C^j_k)\text{-}d(C^i_{k-1})]$ between two successive occurrences $o_{k-1}::C^i$ and $o_k::C^j$ in a specific $\omega_s^{i\text{-}j}$ sequence that contains only the occurrences of the two classes C^i and C^j of the sequence ω_s. The average delay D_{ij} between the occurrences of two classes C^i and C^j of ω is evaluated from two types of Poisson processes:

- A Poisson process $(N_{i\text{-}j}(t\text{-}t_{min});\ t\in T)$ that counts the number of sub sequences $\omega'=(C^i_{k-1},\ C^j_k)$ in each $\omega_s^{i\text{-}j}$.
- A compound Poisson process $(N^D_{i\text{-}j}(t\text{-}t_{min});\ t\in T)$ associated to each Poisson process $(N_{i\text{-}j}(t\text{-}t_{min});\ t\in T)$

The average delay D_{ij} is then given by (Le Goc and Al, 2006):

$$D_{ij} = E\left[d\left(C^j_k\right) - d\left(C^i_{k-1}\right)\right] = \frac{1}{\lambda_{i-j}} = \frac{N^D_{i-j}\left(t_{\max} - t_{\min}\right)}{N_{i-j}\left(t_{\max} - t_{\min}\right)} \qquad (6)$$

In our applications, the timed constraints are often intervals of the form $[0,\ 2/\lambda_{i\text{-}j}]$ because experts generally agree with this choice, which takes into account 60% of the occurrences. The role of the "BJT4T" algorithm (Backward Jump with Timed constraints for Trees) is to compute the set of the most probable timed binary relations $R(C^i,\ C^j,\ [\tau^-,\ \tau^+])$ in a set Ω of sequences ω_i that leads to a specific discrete event class C^j. The "BJT4S" algorithm evaluates the anticipating ratio of each branch of the tree: the signatures are the branches of the tree having an anticipating ratio greater that an arbitrary threshold.

3.3 Average-Time Sequence

A signature subsumes a particular sequence called the "Average-Time Sequence" (A-TS).

The average-time sequence ω_s of a signature S containing k classes C^i is made with the occurrences of the only k classes of S. For each C^i class, the number of occurrences is generated and temporally distributed according to the Poisson rates λ_i of the C^i class. The A-TS ω_s is then the result of the ordering of the occurrences of all the classes according to their time.

The period of ω_s is computed when finding the real number T_S so that:

$$\exists T_S \in \Re,\ \forall i \in \aleph,\ \exists m_i \in \aleph,\ \lambda_i * T_S = m_i \qquad (7)$$

The natural number m_i is the number of occurrences of the class C^i during the period T_S. This means that, C^i_k being the k^{th} occurrence of ω_s, the occurrence of time $d(C^i_k)+(j*T_S)$ is also an occurrence of the C^i class:

$$\forall j \in \aleph, C^i_k \in \omega_S \Rightarrow \exists C^i_m \in \omega_S,\ d(C^i_m)=d(C^i_k)+(j*T_S) \quad (8) \qquad (8)$$

An average-time sequence of a signature S is made with the following method. For each discrete event classes C^i of S, a standard Poisson number generator is used to produce m_i natural numbers according to the Poisson rate λ_i of C^i. To each of the m_i natural numbers corresponds a particular inter-occurrence time. This time is provided when superposing the natural number distribution with the corresponding time distribution of the occurrences of C^i (figure 2).

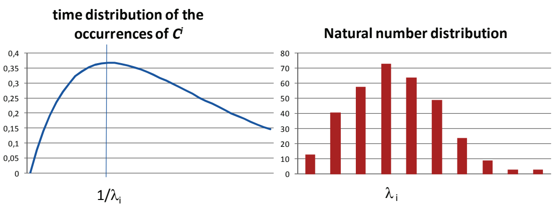

Figure 2: Times and Numbers Distributions.

The maximum of the two distributions match together: the most frequent natural number λ_i corresponds to the most frequent inter-occurrence time $1/\lambda_i$. This means that the inter occurrence time corresponding to the number λ_i is $1/\lambda_i$. So the inter occurrence time corresponding to the number 1 is $1/\lambda^2_i$. This lead to the equation 9 providing the inter-occurrence time for any number n>0:

$$\forall n \in \aleph, n > 0, n \Leftrightarrow \frac{n}{\lambda_i^2} \qquad (9)$$

When n=0, equation 9 leads to an inter-occurrence time equal to 0 that means simultaneous occurrences. To avoid this problem, an arbitrary constant τ_i is associated with the number 0. This constant corresponds to a shift of the time of the occurrence series. In practice, we define the values of the constants τ_i from the value of the Poisson rate λ_i:

- When $\lambda_i \geq 1$, then $\tau_i = 1/2\lambda^2_i$, (i.e. the half of the inter occurrence time for n=1)
- When $\lambda_i<1$, $\tau_i = 1/2\lambda_i$. (i.e. the half of $1/\lambda_i$)

This leads to the equation 10 that provides the inter occurrence time corresponding to each natural number of a series generated with a standard Poisson number generator parameterized λ_i.

$$\begin{aligned} &\forall n \in \aleph, n \Leftrightarrow \tau_i + \frac{n}{\lambda_i^2} \\ &\lambda_i \geq 1 \Rightarrow \tau_i = \frac{1}{2 \cdot \lambda_i^2} \\ &\lambda_i < 1 \Rightarrow \tau_i = \frac{1}{2 \cdot \lambda_i} \end{aligned} \qquad (10)$$

The occurrence series of a C^i class is made when substituting each n of the natural number series with an occurrence of the time: the time of the preceding occurrence plus the inter-occurrence time given by the equation 10 (cf. Table 1 for an example with λ_i=0,58). An instance of the A-TS ω_s of a signature S is then the superposition of the occurrences series of each class C^i of the signature S.

Table 1: Example of Occurrence Time Computation.

natural number	inter-occurences time	occurrences dates
0	0,29726516	0,29726516
3	9,21521999	9,51248515
1	3,26991677	12,78240192
0	0,29726516	13,07966708
0	0,29726516	13,37693224

Using this method, an instance of the average time sequence ω_s of a signature S can be used to generate sequences whose stochastic and timed properties are as close as necessary of the initial set of sequences. The average time sequence ω_s constitutes then a global model of a given set of sequences according to a signature S.

The next section illustrates the interest of this model when the process is a lime kiln production unit supervised with an industrial automaton.

4 APPLICATION

The application is a lime kiln unit used to produce quicklime by the calcination of limestone (calcium carbonate). The main inputs of this process are stones and energy flows. The main output is the evacuated flow of quicklime. The supervision system monitors 9 components of the Lime Kiln process (Figure 2) and detects 174 types of alarms.
The diagram of Figure 2 are the relation of each of the 174 types of alarms with one of the 9 component of the lime kiln production unit are the only elements provided to analyze the given sequence. In other words, there is no *a priori* knowledge available about the behavior and the functions of lime kiln production unit.

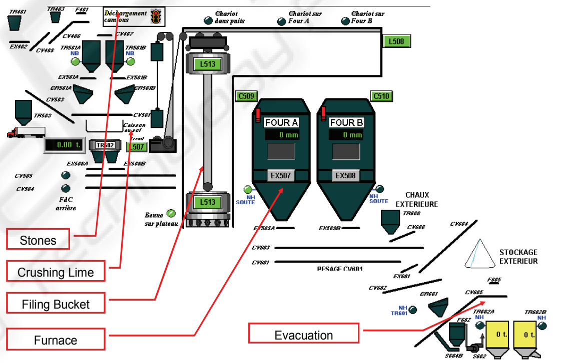

Figure 3: Structural Model of a Lime Kiln Process.

4.1 Stochastic Representation

To apply the Stochastic Approach, the 174 types of alarms are considered as 174 classes of discrete events and 9 variables are associated with one of the 9 components. A class is then constructed with a variable and a set of 19 possible values in the average. Alarms are designated with natural numbers in the interval [2000, 2173].

The two conditions of the Stochastic Approach must be verified: the independence of the classes and the distribution of the occurrences according to the Poisson law. The first condition is guaranteed by the supervision system: an alarm occurrence does not depend on a preceding occurrence of alarm. In that case, the second condition is often verified (see (Lang and Al, 1999) for a more general discussion about these conditions). Figure 3 shows the counting processes of the occurrences of some of the classes: the second condition is verified at least with visualization: there is no anomaly in the global growth of each curves. The Markov chain of the Stochastic representation contains then 174 states and 30276 potential transitions.

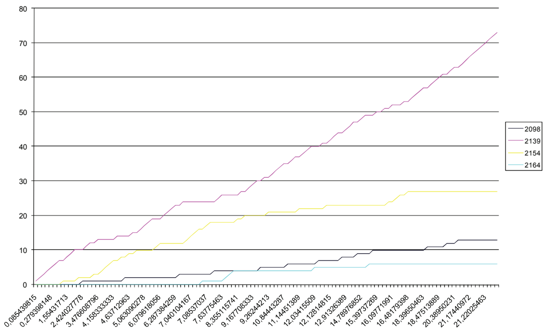

Figure 4: Part of Poisson Processes of ω.

The analyzed sequence ω contains 2852 occurrences and covers around 22 days. During this period, the global occurrences counting process of ω behaved like a Poisson process with a rate λ equal to 5,4 occurrences per hour, that is to say one occurrence every 11'.

4.2 Signatures of the 2139 Class

The alarms corresponding to the 2139 class show a problematic level of material on the evacuation of the quicklime. This type of alarms is on of the most problematic to manage the lime kiln production unit.

The BJT4T algorithm is used to build the tree of the most probable sequential relations leading to the 2139 class. The algorithm is parameterized so that the tree has a depth of 4 levels, each node having a maximum of 4 children's. The BJT4S extract from this tree the 2139 class signature of Figure 4: this branch is the only branch of the 2139 class tree having an anticipating ratio greater than 50%.

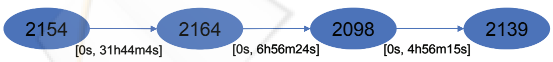

Figure 5: 2139 Class Signature.

The anticipating ratio of the 2139 class signature of Figure 4 is 150%: 3 sub sequences of ω satisfy the constraints of the complete 2139 class signature and 2 sub sequences satisfy the constraints of the signature without the final link (i.e. 2098→2139). In other words, two occurrences of the 2139 class, 2139_{648} and 2139_{669}, satisfy the timed constraint of the 2098→2139 relation while the corresponding 2098_{640} occurrence belongs to only one triplet of occurrences (2154_{499}, 2164_{625}, 2098_{640}) that satisfy the 2139 class signature without the final link.

This 2139 class signature means that there is a strong probability that a problem with the evacuation (2139) can occur when there is a problem on the filling bucket (2154) which is correlated with a problem on the furnace B (2164, 2098).

4.3 Average-Time Sequence

The 2139 class signature is made with 4 classes, the Poisson rates of which are given in the table of figure 6.

Table 2: Poisson rates of the 2139 class signature.

	2098	2139	2154	2164
Lambda	0,58	3,28	1,03	0,4

Using equation (13), the period of the associated A-TS ω_s is Ts=100 days long and contains 528 occurrences (58 occurrences of the 2098 class, 328 occurrences of the 2139 class, etc). The time of these occurrences is given by the equation (14) for each of these classes. The beginning of ω_s is the following: {(0,3; 2139); (0,6; 2139); (0,9; 2139); (0,97; 2154); (1,2; 2139); (1,5; 2139); (1,72; 2098); (1,8; 2139); (1,94; 2154); (2,1; 2139); (2,4; 2139); (2,5; 2164); (2,7; 2139);(2,91; 2154); (3; 2139); (3,3; 2139); (3,44; 2098); (3,6; 2139); (3,88; 2154); (3,9; 2139); (4,2; 2139); (4,5; 2139); (4,8; 2139); (4,85; 2154); … }

Using the properties of the exponential distribution, ω_s can be used to produce a new sequence ω_s' the stochastic properties of which are as close as desired to the filtered sequence $\omega_{2139} \subset \omega$ containing the only occurrences of the 4 classes of the 2139 class signature.

To this aim, a Poisson number generator using the Poisson rates of Table 1 allows to define the time of each occurrences of the ω_s' sequence so that the inter-occurrence time is not a constant but follows the exponential law $\lambda t e^{-\lambda t}$. ω_s' is then a particular realization of the A-TS ω_s. Given such a sequence, the BJT4T algorithm will produce the tree of Figure 5 for the 2139 class.

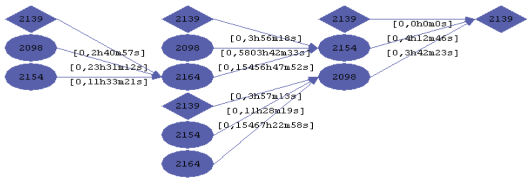

Figure 6: 2139 Class Tree according to ω_s'.

This tree can be compared with the 2139 class tree of the filtered sequence ω_{2139} Figure 6.

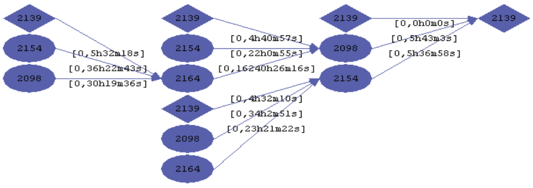

Figure 7: 2139 Class Tree according to ω_{2139}.

Figures 5 and 6 differ only with the position of the 2098 and 2154 classes. This difference comes from the fact that the 2154 class has a non-homogenous behavior in ω_{2139} (and consequently in ω): during the seven first days, the Poisson rate of the 2154 class is three times greater than during the 13.9 last days. The Poisson rate of the 2154 class of the average time sequence ω_s (and consequently ω_s') is closer to the Poisson rate of the 13.9 last days. This means that the 2154 class defines two periods where its Poisson differs but are constant. This leads to cut up ω_{2139} in two periods.

It is to note that only the 2154 class Poisson rate differs from the first period to the second period; the Poisson rates of the other classes are not significantly different.

Table 3: Poisson rates of the second part of ω_{2139}.

	Second Period 13,9 days			
	2098	2139	2154	2164
Lambda	0,64	3,37	0,64	0,35

Containing only 48 occurrences, the first period of the sequence ω is too short to provide a significant tree, so no studies can be done.

Using the same method, a new realization ω_s" of the A-TS is made with the Poisson rates of the second part of the ω_{2139} sequence (Table 2). The BJT4T algorithm produces the 2139 class tree of Figure 7 with ω_s":

The tree of Figure 7 is now very similar to the tree of any realization of the A-TS ω_s (Figure 5): the only difference is the relative position of the leaves corresponding to the 2154 and the 2098 classes (at the left side of the trees). We supposes that the cause of this difference is the too short length of the second part of the ω_{2139} sequence (14 days) to be representative to the couple made with the process and its monitoring system.

Nevertheless, we can consider that, for a given class C^i, the stochastic properties of the occurrences of the class contained in any realization of the corresponding A-TS ω_s are very close to those of the filtered sequence $\omega_{C_i} \subset \omega$, the temporal properties of the occurrences being the same. This shows that, according to a signature, the corresponding Average-Time Sequence is a global model of a sequence. This result is true with any signature.

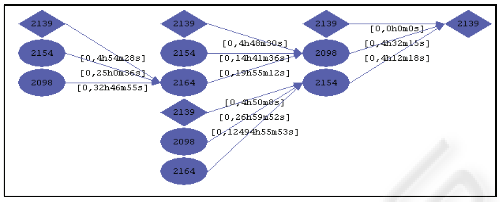

Figure 8: 2139 Class Tree according to ω_s".

5 CONCLUSIONS

This paper presents the Average-Time Sequence model of a log of alarms and shows that this model is a global model of the relations between the alarms.

The modeling process is based on the Stochastic Approach for discovering temporal knowledge from a set of sequences of discrete event class occurrences. The Stochastic Approach represents such a set in the dual forms of a homogeneous Markov chain and a superposition of Poisson processes. The advantage of this approach is that the timed binary constraints are provided by the Poisson process theory and are coherent with the probability of the binary sequential relation between two classes. The discovered knowledge is represented as abstract chronicle models made with a set of binary relations between discrete event classes that are timed constraints.

The paper shows that an abstract chronicle model usable to predict the occurrences of a discrete event class subsumes a global model of the sequence, the average time sequence. Such a model can be used to produce a sequence the stochastic and timed properties of which are as closed as desired of those of the given set of sequences.

The Stochastic Approach has been used to study the alarms or the messages generated by a wide variety of monitored process like the blast furnace and the Sachem monitoring system (Le Goc, 2004), a galvanization bath and the Apache monitoring system (Le Goc and Al, 2006) or the wafer manufacturing production tools and its the supervision system of the STMicroelectronics company (Benayadi and Al, 2006). The application described in this paper shows that the Stochastic

Approach can also be applied to analyze the alarms generated by an industrial automaton supervising a production process.

Currently, we are working at introducing an entropic criterion in the Stochastic Approach to prune the trees produced with the BJT4T algorithm (Benayadi and Le Goc, 2007) and at defining a cognitive approach of modeling dynamic systems that is compatible with the Stochastic Approach of modeling (Masse and Le Goc, 2007).

REFERENCES

R. Agrawal, T. Imielinski and A. Swami (1993). Mining Association Rules between sets of Items in Large Databases. *Proceeding of the 1993 ACM SIGMOD International Conference on Management of Data*, pages 207-216.

R. Agrawal and R. Srikant (1995). Mining Sequential Patterns. *Proceedings of the International Conference on Data Engineering (ICDE'95)*, Taipei, Taiwan.

N. Benayadi, M. Le Goc, and P. Bouché (2006). Discovering Manufacturing Process from Timed Data: the BJT4R Algorithm. *2nd international workshop on Mining Complex Data (MCD'06) of the 2006 IEEE International Conference on Data Mining (ICDM'06)*, Hong Kong, China.

N. Benayadi and M. Le Goc (2007). Discovering Expert's Knowledge from Sequences of Discrete Event Class Occurrences. *To appear in the proceedings of the 10th International Conference on Enterprise Information Systems (ICEIS'08)*, Barcelona, Spain, 12-16 June 2007.

P. Bouché and Le Goc M (2004). Discovering Operational Signatures with Time Constraints from a Discrete Event Sequence. *Proceedings of the 4th International Conference on Hybrid Intelligent Systems (HIS'04)*, Kitakyushu, Japan.

P. Bouché, M. Le Goc, and N. Giambiasi (2006). Building a Functionnal Model from a Sequence of Alarms: the Example of APACHE. *IFAC Workshop on Automation in Mining, Mineral and Metal Industry (MMM'2006)*, Cracow, Poland.

C. G. Cassandras and Lafortune S. (2001). *Introduction to discrete event systems*. Kluwer Academic Publishers.

S. Cauvin, M-O. Cordier, C. Dousson, P. Laborie, F. Lévy, J. Montmain, M. Porcheron, I. Servet and L. Travé (1998). Monitoring and Alarm Interpretation in Industrial Environments. *AI Communications*, Vol. 11-3-4, p. 139-173, IOS Press.

M.O. Cordier, and C. Dousson (2000). Alarm Driven Monitoring Based on Chronicle. *Proceedings of SafeProcess 2000*, pages 286-291, Budapest, Hungary.

C. Dousson and T. Vu Duong (1999). Discovering Chronicles with Numerical Time Constraints from Alarms Logs for Monitoring Dynamic Systems. *Proceedings of the 13rd International Join Conference on Artificial Intelligence (IJCAI'99)*, pp. 620-626.

K. Hatonen, M. Klemettinen, H. Mannila, P. Ronkainen and H. Toivonen (1996). Knowledge discovery from telecommunication network alarm databases. *Proceedings of the 12th International Conference on Data Engineering (ICDE '96)*. New Orleans, LA, pp. 115–122.

K. Hatonen, M. Klemettinen, H. Mannila, P. Ronkainen and H. Toivonen (1996). TASA: Telecommunication alarm sequence analyzer, or how to enjoy faults in your network. *Proceedings of the 1996 IEEE Network Operations and Management Symposium (NOMS '96)*, Kyoto, Japan, pp. 520–529.

M. Ghallab (1996). On Chronicles: Representation, On-line Recognition and Learning. *Principles of Knowledge Representation and Reasoning*, Aiello, Doyle and Shapiro (Eds.), p. 597-606, Morgan-Kauffman.

S. Hanks and McDermott D (1994). Modelling a dynamic and uncertain world I: symbolic and probabilistic reasoning about change. *Artificial Intelligence*, Vol. n°66, pp 1-55.

M. Lang, T.B.M.J Ouarda and B. Bobée (1999). Towards operational guidelines for over-threshold modeling. *Journal of Hydrology*, Elsevier Edition, Vol. n°225, p. 103-117.

M. Le Goc (2004). SACHEM, a real Time Intelligent Diagnosis System based on the Discrete Event Paradigm. *Simulation*. The Society for Modeling and Simulation International Ed., Vol. 80, n° 11, pp. 591-617.

M. Le Goc, P. Bouché and N. Giambiasi (2006). Temporal Abstraction of Timed Alarm Sequences for Diagnosis. *Proceedings of the International Conference on COGnitive systems with Interactive Sensors (COGIS'06)*, Paris, France.

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

H. Mannila (2002). Local and Global Methods in Data Mining: Basic Techniques and Open Problems. *Proceedings of the 29th International Colloquium on Automata, Languages and Programming*, Vol. n°2380, pages 57-68, Malaga, Spain.

E. Masse and M. Le Goc (2007). Modeling Dynamic Systems from their Behavior for a Multi Model Based Diagnosis Task. *Proceedings of the 18th International Workshop on Principles of Diagnosis (DX'07)*, Nashville, USA, May 29-31 2007.