DISCOVERING EXPERT'S KNOWLEDGE FROM SEQUENCES OF DISCRETE EVENT CLASS OCCURRENCES

Le Goc Marc and Benayadi Nabil
LSIS, Laboratory for Sciences of Information and Systems, UMR CNRS 6168
Aix-Marseille University, 20 Avenue Escadrille Normandie Niemmen, Marseilles cedex 20, France

Keywords: Temporal Knowledge Discovering, Markov Process, Information Thoeory, Knowledge Based Systems.

Abstract: This paper is concerned with the discovery of expert's knowledge from a sequence of alarms provided by a knowledge based system monitoring a dynamic process. The discovering process is based on the principles and the tools of the Stochastic Approach framework where a sequence is represented with a Markov chain from which binary relations between discrete event classes can be find and represented as abstract chronicle models. The problem with this approach is to reduce the search space as close as possible to the relations between the process variables. To this aim, we propose an adaptation of the J-Measure to the Stochastic Approach framework, the BJ-Measure, to build an entropic based heuristic that help in finding abstract chronicle models revealing strong relations between the process variables. The result of the application of this approach to a real world system, the Sachem system that controls the blast furnace of the Arcelor-Mittal Steel group, is provided in the paper, showing how the combination of the Stochastic Approach and the Information Theory allows finding the a priori expert's knowledge between blast furnace variables from a sequence of alarms.

1 INTRODUCTION

In supervised and monitored processes like production or manufacturing processes, telecommunication networks or web servers, a very large amount of timed messages (alarms or simple records) are generated and collected in databases. There is an increasing interest in mining these messages to discover the underlying relations between the variables that govern the dynamic of the process and to improve its management. This problem is still an open problem (cf. problems 2 and 3 formulated in (Mannila, 2002)) and one of the difficulties comes from the combination of logical relations and timed constraints (Cauvin et al., 1998; Hanks and Dermott, 1994).

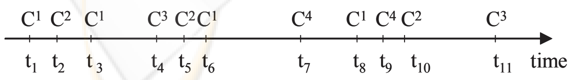

Figure 1: Example of sequence.

This paper addresses this problem in the framework of the Stochastic Approach (Le Goc et al., 2005): discovering such relations from a set of sequences of timed messages provided by a knowledge based system that monitors a dynamic process. In this framework, the messages are timed with a continuous time structure clock and are considered as occurrences (t_k, C^i) of discrete event classes C^i like in the figure 1. A class is then a type of message (alarms, simple records, url of a web site page, ...).

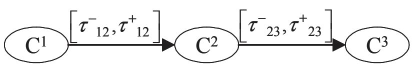

Figure 2: Example of Abstract Chronicle Model.

The Stochastic Approach considers a sequence $\omega = (o_k :: C^i), k \in K = \{0, \ldots, m-1\}$ of m occurrences o_k of discrete event classes C^i as the observable effects of a series of transition state in a timed stochastic automata (i.e. a Markov process). A set of timed binary relations between discrete event classes can be deduced from this automata and represented with abstract chronicle models as in figure 2. Such an abstract chronicle model will come into effect when it allows to predict most of the occurrences of a given class. For example, if $[t_2 - t_1]$ and $[t_{10} - t_8]$ $\in [\tau^-_{12}, \tau^+_{12}]$, the abstract chronicle model of figure 2 can be used to predict two occurrences of the C^3 class in the sequence of figure 1. If $[t_4 - t_2]$ and $[t_{11} - t_{10}]$ $\in [\tau^-_{23}, \tau^+_{23}]$, then the prediction is successful.

Goc Marc L. and Nabil B. (2008).
DISCOVERING EXPERT'S KNOWLEDGE FROM SEQUENCES OF DISCRETE EVENT CLASS OCCURRENCES.
In *Proceedings of the Tenth International Conference on Enterprise Information Systems - AIDSS*, pages 253-260
DOI: 10.5220/0001695702530260

One of the problem with the Stochastic Approach is the size of timed stochastic automata: the search space of timed binary relations evolve exponentially with the number of discrete event classes. There is then a need for reducing the search space as closed as possible to the potentially useful timed binary relations. To this aim, this paper proposes to use an adaptation of the J-measure, called the BJ-measure, to evaluate the interestingness of hypothesis of timed sequential binary relations.

The next section presents briefly the main works that are related with the problem of mining a timed data set. Section 3 introduces the basis of the Stochastic Approach. Section 4 defines the BJ-measure used to build an heuristic to prune a tree of abstract chronicle models. Section 5 presents the abstract chronicle model discovered in 2007 with this heuristic and shows that this result is similar to the *a priori* causal knowledge an expert group formulated in 1995 about a very complex real world process: a Sachem monitored blast furnace of the Arcelor-Mittal steel group.

2 RELATED WORKS

Data Mining was developed at the confluence of research in Artificial Intelligence (Machine Learning), Statistics and database systems (refer to (Roddick and Spiliopoulou, 2002) for a complete survey of the main paradigms and methods for mining a sequence).

The approaches for looking for a minimum set of association rules of the form $y_1, y_2, ..., y_n \rightarrow x$ that characterizes the relations between the data contained in a database are mainly based on the Apriori algorithm (Agrawal et al., 1993). This algorithm computes the number of time a *pattern* $(y_1, y_2, ..., y_n, x)$ is observed in a data set. This number is called the support. When the support of a pattern is greater than a minimum threshold, the pattern is considered as a potential association rule. This characterizes the "Frequency Approach". When data are timed, the data set is ordered and is called a sequence. The adaptation of the Frequency Approach to mine sequences takes into account the order of a *sequential* pattern $(y_1, y_2, ..., y_n, x)$ and leads to a division of the ordered data set into a set of sequences (Agrawal and Srikant, 1995). The support is then the number of sequences containing a sequential pattern and is used in algorithms like AprioriAll, AprioriSome or DynamicSome. The adaptation proposed in (Mannila et al., 1997; Hatonen et al., 1996a; Hatonen et al., 1996b) get round the problem of the arbitrariness of the division of an ordered data set through a systematic division into sequences having the same temporal length (Winepi and Minepi algorithms). In the Temporal Reasoning domain, (Ghallab, 1996) proposes the notion of chronicle to represent a set of timed binary relations between events. A chronicle is a kind of temporal pattern specification where nodes are events and links are timed binary constraints represented with [min, max] intervals. Gallab's method for discovering chronicle models splits a set of sequences in examples and counter examples and look for the longest patterns that are common to the examples but not included in the counter examples. With the FACE algorithm, Dousson and Duong (Dousson and Duong, 1999) adapt the Frequency Approach of (Agrawal and Srikant, 1995) to discover frequent chronicles, but do not propose a sound method to evaluate the timed constraints.

The Frequency Approach generates a large amount of relations (Roddick and Spiliopoulou, 2002) and fails at providing a global description of a given sequence (Mannila, 2002). So measures have been defined to evaluate the interest of the discovered relations (Liu et al., 2000; Padmanabhan and Tuzhilin, 1999; Tan et al., 2004; Vaillant et al., 2004; Huynh et al., 2005; Bayardo and Agrawal, 1999; Hilderman and Hamilton, 2001; Jaroszewicz and Simovici, 2001; Theil, 1970). In the Timed Data Mining domain, the J-measure is used to evaluate the "informativeness" of a rule (Smyth and Goodman, 1992). Let X and Y be two random variables taking a value in the respective sets $X = \{x_i\}_{i=0,1,...,n}$ and $Y = \{y_j\}_{j=0,1,...,m}$. The J-measure is the amount of mutual information shared between the variable X and the value y_j (Smyth and Goodman, 1992). When denoting $p(X = x_i) \equiv p(i)$ and $p(Y = y_i) \equiv p(j)$, the J-measure is given by the equation 1.

$$\begin{aligned} J(X, Y = y_j) &= p(j) \times \sum_i p(i|j) \times \log(\frac{p(i|j)}{p(i)}) \\ &\equiv p(j) \times j(X, Y = y_j) \qquad (1) \end{aligned}$$

The J-measure compares the posterior probability of each rule consequent given the antecedent with the prior probability of the consequent ($j(X, Y = y_j)$), as done with the cross-entropy measure, but also takes the prior probability $p(j)$ of the antecedent into account (Roddick and Spiliopoulou, 2002), (Shore and Johnson, 1980). The J-measure is unique, never negative and null at the independence point (Blachman, 1968). These properties explains its usage to mine sequences. The framework of the Stochastic Approach (Le Goc et al., 2005; Bouché et al., 2005) being closed to the Shannon's Information Theory framework (Shannon and Weaver, 1949), we propose then to combine them to define an entropic based heuristic for finding strong relations between variables from a set of sequences.

3 THE STOCHASTIC APPROACH

A discrete event e_i is a pair (x, δ_i), where $x \in X$ is the name of a discrete variable and $\delta_i \in \Delta^x$ is a constant. A discrete event occurrence $o_k \in O$ is a tuple (t_k, x, δ_i), where $t_k \in \Gamma \subseteq \Re$ is the time of the assignation of δ_i to x, so that $o_k \equiv (t_k, x, \delta_i)$ corresponds to the assignation: $x(t_k) = \delta_i$ (equation 2). The occurrences are timed with a continuous clock structure: $\forall t_{k-2}, t_{k-1}, t_k \in \Gamma, t_{k-2} - t_{k-1} \neq t_{k-1} - t_k$.

$$\begin{array}{l} \forall t_k \in \Re, \forall \delta_i \in \Delta^x, \exists t < t_k, \\ x(t) \neq \delta_i \wedge x(t_k) = \delta_i \Rightarrow o_k \equiv (t_k, x, \delta_i) \end{array} \quad (2)$$

A discrete event class $C^i = \{e_i\}$ is an arbitrary set of discrete events $e_i \equiv (x, \delta_i)$. An occurrence $o_k \equiv (t_k, x, \delta_i)$ of a discrete event class $C^i = \{(x, \delta_i)\}$ is denoted either $o_k :: C^i$ or $o_k \equiv (t_k, C^i)$. A sequence $\omega = \{o_k :: C^i\}_{k=0,\dots m-1}$ of discrete event class occurrences is an ordered set of m occurrences $o_k :: C^i$ of the set $C_\omega = \{C^i\}$ of the discrete event classes having at least one occurrence in ω.

A timed binary relation $R(C^i, C^o, [\tau^-, \tau^+])$ is a sequential relation between two classes that is timed constrained. "$[\tau^-, \tau^+]$" is the time interval for observing an occurrence of the output class $o_n :: C^o$ after the occurrence of the input class $o_k :: C^i$. Equation 3 defines a relation observed in a sequence ω where d is a function returning the occurrence time $(\forall o_k \equiv (t_k, C^i), d(o_k) = t_k)$.

$$\begin{array}{l} R(C^i, C^o, [\tau^-, \tau^+]) \Leftrightarrow \exists o_n, o_k \in \omega, \\ (o_n :: C^o) \wedge (o_k :: C^i) \\ \wedge (d(o_n) - d(o_k) \in [\tau^-, \tau^+]) \end{array} \quad (3)$$

An abstract chronicle model is a set $M = \{ R_{ij}(C^i, C^j, [\tau^-_{ij}, \tau^+_{ij}]) \}$ of timed sequential binary relations. For example, the abstract chronicle model $M_{123} = \{ R_{12}(C^1, C^2, [\tau^-_{12}, \tau^+_{12}]), R_{23}(C^2, C^3, [\tau^-_{23}, \tau^+_{23}]) \}$ defines two relations between three classes. This model is represented with the ELP knowledge representation language (Le Goc et al., 2005) in Figure 2. A sequence ω satisfies the M_{123} model when:

$$\begin{array}{l} \exists o_k, o_n, o_m \in \omega, \\ (o_k :: C^1) \wedge (o_n :: C^2) \wedge (o_m :: C^3) \\ \wedge (d(o_n) - d(o_k) \in [\tau^-_{12}, \tau^+_{12}]) \\ \wedge (d(o_m) - d(o_n) \in [\tau^-_{23}, \tau^+_{23}]) \end{array} \quad (4)$$

A path of an Elp Model is a series $M = \{ R(C^i, C^{i+1}, [\tau^-_i, \tau^+_i]) \}$, $i = 0 \dots n$, of n timed timed binary relations (M_{123} for example).

An instance ω_m of an Elp Model M is a sequence containing occurrences of classes which are consistent with the logical and the timed constraints of M. For example, if $[\tau^-_{12}, \tau^+_{12}] = [0, 5]$ and $[\tau^-_{23}, \tau^+_{23}] = [3, 8]$, the sequence $\{ (1, C^1), (3, C^4), (4, C^2), (8, C^1), (10, C^3) \}$ is an instance of the Elp Model M_{123} (Figure 2) because the occurrences $(1, C^1)$, $(4, C^2)$, and $(10, C^3)$ satisfy the logical and the timed constraints of M.

Given a sequence ω, the anticipation rate (AR) of a path M is the ratio between the number $i(M)$ of instances of M and the number $i(M')$ of instances where M' is the path M minus the last timed binary relation (i.e. $R(C^{n-1}, C^n, [\tau^-_{n-1}, \tau^+_{n-1}])$). For example, if $i(M_{123}) = 10$, and $i(M_{12}) = 15$ in a given a sequence ω, $M_{12} = M_{123} - R_{23}(C^2, C^3, [\tau^-_{23}, \tau^+_{23}])$, then the anticipation ratio $AR(M_{123})$ is equal to $10/15 = 66\%$ in ω. The cover rate CR of a path M is the ratio between the number of instances $i(M)$ and the number of occurrences of the final (output) class of M in ω. For example, if $i(M_{123}) = 10$ and the number of occurrences $o :: C^3$ in ω is 20, then $CR(M_{123}) = 10/20 = 50\%$. The Cover Rate is a kind of timed version of the support notion of the Frequency Approach. When a path has an anticipation rate and a cover rate over significant thresholds, like M_{123} in the example, it can be transformed in diagnosis rules (equation 5) (Le Goc et al., 2005). A set of paths that can be transformed in diagnosis rules is called a *signature*. So, the aim of the Stochastic Approach is to help in discovering signatures.

$$\begin{array}{l} \forall o_k, o_n \in \omega', \\ (o_k :: C^1) \wedge (o_n :: C^2) \wedge (d(o_n) - d(o_k) \in [\tau^-_{12}, \tau^+_{12}]) \\ \Rightarrow \exists o_m \in \omega', (o_m :: C^3) \wedge (d(o_m) - d(o_n) \in [\tau^-_{23}, \tau^+_{23}]) \end{array} \quad (5)$$

When the discrete event class occurrences are independent, an ω sequence can be represented with a homogeneous Markov chain $X = (X(t_k); k \in K)$. To this aim, the set of discrete event classes $C_\omega = \{C^i\}_{i=0\dots n-1}$ of ω is confused with the state space $Q = \{i\}_{i=0\dots n-1}$ of the Markov chain X. A binary subsequence $\omega' = (o_{k-1} :: C^i, o_k :: C^j) \subseteq \omega$ corresponds then to a state transition in $X : X(d(o_{k-1})) =$

$i \to X(d(o_k)) = j$. X being homogeneous, the transition probability from a state i to a state j is a constant $P[j|i] \equiv P\left[C^j|C^i\right] \equiv p_{ij}$:

$$\begin{array}{l} \forall k \in K, p_{ij} = P\left[X(t_k) = j | X(t_{k-1}) = i\right] \\ p_{ij} = P\left[(o_{k-1} :: C^i, o_k :: C^j) \subseteq \omega | o_{k-1} :: C^i\right] \end{array} \quad (6)$$

A timed sequential binary relation $R(C^i, C^j, [\tau^-, \tau^+])$ is made with a sequential relation $R_s(C^i, C^j)$ deduced from the transition probability matrix $P = [p_{ij}]$, the timed constraints $[\tau^-, \tau^+]$ being computed from the corresponding Poisson processes superposition (Le Goc, 2006). To this aim, the P matrix is weighted in a $B = [b_{ij}]$ matrix with the probability of the sub-sequence $(o_{k-1} :: C^i, o_k :: C^j)$ in ω:

$$b_{ij} = p_{ij} \times P[(o_{k-1} :: C^i, o_k :: C^j) \subseteq \omega] \quad (7)$$

Given a set $\Omega = \{\omega_i\}$ of sequences, the role of the *BJT* algorithm (Backward Jump with Timed constraints (Le Goc et al., 2005; Bouché et al., 2005)) is to compute the B matrix and the Poisson process superposition associated with Ω. Given a maximum depth and a maximum width, the *BJT4T* algorithm (BJT for Tree) uses these representations to constitute a tree $M = \{R_{ij}(C^i, C^j, [\tau^-, \tau^+])\}$ of the most probable timed binary relations leading to a specific output class C^k. The algorithm *BJT4S* (BJT for Signatures) looks for the anticipating and the cover rates of each paths of such a tree.

The problem with this method is that the number of paths contained in $M = \{R_{ij}(C^i, C^j, [\tau^-, \tau^+])\}$ is exponential with its width. So there is a need to define a punning method to keep only sub branches containing strong relations between the classes.

4 THE BJ-MEASURE

According to the memoryless property of a Markov chain, the sequential relation $R_s(C^i, C^j) \equiv C^i \mapsto C^o$ of a timed binary relation $R(C^i, C^o, [\tau^-, \tau^+])$ between the classes C^i and C^o can be view like one of the four relations linking the values of two random binary variables $Y = \{C^i, \neg C^i\}$ and $X = \{C^o, \neg C^o\}$ connected through a discrete memoryless channel ((Shannon and Weaver, 1949), Figure 3). This means that $\neg C^i \equiv C_\omega - \{C^i\}$ and $\neg C^o \equiv C_\omega - \{C^o\}$, so that $p(C^o|C^i) + p(\neg C^o|C^i) = 1$. The basic definition of Shannon's condition information entropy is then directly applied (equation 8) and an oriented J-measure can be defined on a relation $C^i \mapsto C^o$.

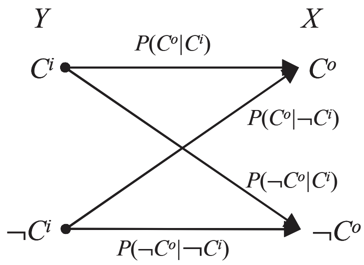

Figure 3: Memoryless Channel.

$$\begin{array}{rcl} H(\{C^o, \neg C^o\}|C^i) & = & \\ & - & p(C^o|C^i) \times \log(p(C^o|C^i)) \\ & - & p(\neg C^o|C^i) \times \log(p(\neg C^o|C^i)) \\ & = & H(C^o|C^i) + H(\neg C^o|C^i) \end{array}$$

Definition 1. *The occurrences of a class C^i bring information about to the occurrences of a class C^o if and only if $p(C^o|C^i) > p(C^o)$.*

When $p(C^o|C^i) = p(C^o)$, the occurrences of the classes C^i and C^o are independent.

Definition 2. *Considering a sequential binary relation $C^i \mapsto C^o$ such that $p(C^o|C^i) > p(C^o)$, the BJ-measure $BJM(C^i \mapsto C^o)$ of $C^i \mapsto C^o$ is the cross-entropy between the occurrences of the C^i class and the occurrences of the set of classes $\{C^o, \neg C^o\}$ given by the equation 8.*

$$\begin{array}{l} BJM(C^i \mapsto C^o) = p(C^o|C^i) \times \log(\frac{p(C^o|C^i)}{p(C^o)}) \\ + \frac{(1-p(C^o|C^i))}{N(C_\omega)-1} \times \log(\frac{(1-p(C^o|C^i))}{1-p(C^o)}) \end{array} \quad (8)$$

where $n(C_\omega)$ is the number of discrete event classes with at least one occurrence in a sequence ω and $1 - p(C^o|C^i) = p(\neg C^o|C^i)$. The BJ-measure $BJM(C^i \mapsto C^o)$ (cf. Figure 4) is then an adaptation of the j function of the equation 1 to the P matrix of transition probabilities of the Markov chain X corresponding to a set of sequences $\Omega = \{\omega_i\}$.

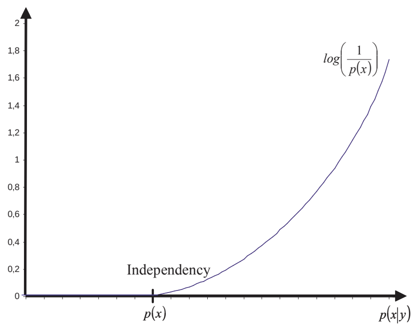

Figure 4: BJ-Measure.

To defines the BJ-measure of a path $M = \{C^i \mapsto C^{i+1}\}_{i=0\dots n-1}$, let us consider that a sequence $\omega_i = \{o_k\}$ of n occurrences is also a sequence of n-1 couples: $\omega_i = \{\ (o_0,o_1),\ (o_1,o_2),\ \dots,\ (o_{n-3},\ o_{n-2}),\ (o_{n-2},\ o_{n-1})\ \}$. The probability of a couple (o_{k-1},o_k) to be an occurrence of the couple (C^i,C^o) is $p(o_{k-1} :: C^i, o_k :: C^o) \equiv p(i,o)$. This means that ω_i contains $p(i,o)\times n$ occurrences of a couple (C^i,C^o). The probability of ω_i will be roughly:

$$p(\omega_i) = \prod_{i=0\dots n-1}(p(i,i+1)^{p(i,i+1)\times n}) \qquad (9)$$

Denoting H the entropy function, we have then:

$$\begin{aligned}&\log(p(\omega_i))\\ &= n\times\sum_{i=0,n-1}(p(i,i+1)\times\log(p(i,i+1)))\\ &\equiv n\times\sum_{i=0,n-1}H((i,i+1))\\ &= n\times\sum_{i=0\dots n-1}(H(i)+H(i+1|i))\end{aligned} \qquad (10)$$

This means that the probability of a sequence is linked with the size of the sequence and the conditional entropy of the successive occurrences. The term $n\times\sum_{i=0\dots n-1}(H(i+1|i))$ is concerned with the P matrix of the Markov chain of ω_i, that is to say the probability of the series of relations $M = \{(C^i \mapsto C^{i+1})\}$, $i = 0\dots n-1$. This leads to the following definitions:

Definition 3. *The BJ-measure of a path $M = \{C^i \mapsto C^{i+1}\}_{i=0\dots n-1}$ exists if and only if, $\forall(C^i \mapsto C^{i+1}) \subseteq M$, $p(C^{i+1}|C^i) > p(C^{i+1})$.*

Definition 4. *When it exists, the BJ-measure of a path $M = \{C^i \mapsto C^{i+1}\}_{i=0\dots n-1}$ is the product of the number of binary relations it contains with the sum of the BJ-measure of each binary sequential relation $C^i \mapsto C^{i+1}$ of M.*

$$\begin{aligned}BJM(M) &= BJM(\{C^i \mapsto C^{i+1}\}_{i=0\dots n-1})\\ &= n\times\sum_{i=0,\dots,n-1}BJM(C^i \mapsto C^{i+1})\end{aligned} \qquad (11)$$

The quantity $BJM(M)$ can then be interpreted as an estimation of the quantity of information that flows through the path M. The probability of an n-ary relation $M = \{C^i \mapsto C^{i+1}\}_{i=0\dots n-1}$ is the probability of the path $(i, i+1, \dots, n)$ in the Markov chain X corresponding to ω_i and is given by the Chapmann-Kolmogorov equation:

$$p(M) = \prod_{i=0,n-1}p(C^{i+1}|C^i) \qquad (12)$$

By definition, the quantity $P(M)$ decreases exponentially with the number n of relations in M when the quantity $BJM(M)$ increases monotonically. The idea is then to combine these quantities in order to find a good tradeoff between the probability of a path M and the quantity of information flowing through it (cf. (Smyth and Goodman, 1992) for a similar principle with the J-measure). Let us call $L(M) = P(M)\times BJM(M)$ this quantity (Benayadi and Le Goc, 2007). The following demonstration shows that $L(M)$ is always limited by a convex function of the form $\phi\times\log(\frac{1}{\phi})$ (Figure 5).

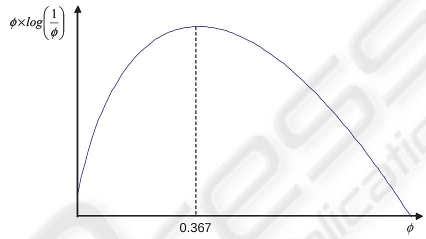

Figure 5: $\phi\times\log(\frac{1}{\phi})$ curve.

Demonstration 1. *According to definition 2:*

- $\forall(C^i \mapsto C^o) \subseteq M$, $BJM(C^i \mapsto C^o) = \alpha+\beta$,
- $\alpha \equiv p(C^o|C^i)\times\log(\frac{p(C^o|C^i)}{p(C^o)})$, $0 < \alpha \leq \log(\frac{1}{p(C^o)})$
- $\beta \equiv \frac{(1-p(C^o|C^i))}{N(C)-1}\times\log(\frac{(1-p(C^o|C^i))}{1-p(C^o)})$, $\beta < 0$.
- *So*, $\forall(C^i \mapsto C^{i+1}) \subseteq M$, $BJM(C^i \mapsto C^{i+1}) \leq \alpha < \log(\frac{1}{p(C^{i+1})})$.
- *Consequently:*

$$\begin{aligned}BJM(M) &\leq n\times\sum_{i=0,n-1}\log(\frac{1}{p(C^{i+1})})\\ BJM(M) &\leq n\times\log(\frac{1}{\prod_{i=0,n-1}(p(C^{i+1}))})\end{aligned} \qquad (13)$$

- *Rewriting $p(C^{i+1}|C^i) \equiv k_i\times p(C^o)$, $k_i > 1$ being a constant, the probability of the series of binary relation of M becomes:*

$$\begin{aligned}p(M) &= \prod_{i=0,n-1}p(C^{i+1}|C^i)\\ &\equiv \prod_{i=0,n-1}(k_{i+1})\times\prod_{i=0,n-1}(p(C^{i+1}))\\ &\equiv K\times\prod_{i=0,n-1}(p(C^{i+1}))\end{aligned} \qquad (14)$$

- *Denoting $\phi \equiv \prod_{i=0,\dots,n-1}p(C^{i+1})$ and $0 < \phi \leq 1$, the function $L(M) = p(M)\times BJM(M)$ is bounded:*

$$L(M) = K\times\phi\times\log(\frac{1}{\phi}) \equiv K\times f(\phi) \qquad (15)$$

- *$f(\phi)$ is a convex function that has one maximum:*

$$\begin{aligned}\frac{\partial f}{\partial\phi} = 0 &\Leftrightarrow \log(\frac{1}{\phi}) - 1 = 0\\ &\Leftrightarrow \phi = \frac{1}{\exp(1)} = 0.367\end{aligned} \qquad (16)$$

This leads to the following heuristic for pruning a path $M = \{C^i \mapsto C^{i+1}\}_{i=0...n-1}$ when there is a sub-path M^1 of M that maximize the quantity L:

Definition 5. *Given a path* $M = \{C^i \mapsto C^{i+1}\}_{i=0...n-1}$, *M can be decomposed into tree series* $M^1 = \{C^i \mapsto C^{i+1}\}_{i=0...n-k-1}$, $M^2 = \{C^i \mapsto C^{i+1}\}_{i=0...n-k}$ *and* $M^3 = \{C^i \mapsto C^{i+1}\}_{i=0...n-k+1}$, $k \geq 1$, *so that:* $L(M^1) \leq L(M^2) > L(M^3)$.

The function $L(M)$ is an heuristic because there is no guarantee that the first maximum is the global maximum. This heuristic has been implemented in the algorithm *BJT4P* meaning BJT for Pruning, and the *BJT4T* algorithm has been modified to take into account the condition of definition 1 so that the resulting trees contains branches with a no null BJ-measure. Next section shows that this simple heuristic provides operational results when used in a very complex real world dynamic process: a Sachem monitored blast furnace.

5 APPLICATION

Sachem is the name of the very large scale knowledge-based system the Arcelor-Mittal Steel group has developed at the end the 20th century to help the operators to monitor, diagnose and control the blast furnace, a very complex production process (Le Goc, 2004; Le Goc et al., 2005)).

With a Sachem system, the blast furnace behavior is described with a flow of occurrences of phenomenon classes (i.e. a series of timed instances). A phenomenon corresponds to the logical description of a type of physical or chemical transformation that can occur in a blast furnace. A phenomenon is represented with a class characterized by a name, an identifier, a set of attributes and two times: the start time and the end time of the phenomenon instance. A phenomenon occurrence is then an instance of such a class where the attributes and the times are valuated by the perception function of a Sachem system. A phenomenon occurrence is created when a behavior corresponding to a particular phenomenon is recognized by Sachem (cf. (Le Goc, 2004) for examples). Sachem describes then the current behavior of a blast furnace with a series of phenomenon occurrences. According to the Stochastic Approach framework, when ordered by their start time, such a series is a sequence of discrete event class occurrences where the classes are the observed phenomena. The application presented in this paper is concerned with the *omega* variable that reveals the right or the wrong usage of the gas inside the blast furnace burden: any distance of the *omega* variable from its ideal value means that the gas is not well used. This is a consequence of a wrong management of the whole blast furnace. The *omega* variable is a very abstract variable corresponding roughly to the ratio of the number of carbon atoms used to produce a ton of hot metal (the main blast furnace output) with the number of iron (f_e) atoms it contains (the studied blast furnace produces 6,000 tons of hot metal per day). When the *omega* variable is equal to the ideal value, the blast furnace is perfectly adjusted: the right quantity of carbon atoms is provided to the blast furnace to produce the required hot metal quantity and every carbon atoms are used to only produce the f_e atoms of the hot metal (no loss of energy). The values of the *omega* variable over time are provided by a mathematical model (the *MMHF* model) which is a set of 17 differential equations linking together 53 high level variables synthesizing the whole the blast furnace behavior. This model aims at equilibrating together a material balance and an energetic balance that defines the function point of the blast furnace.

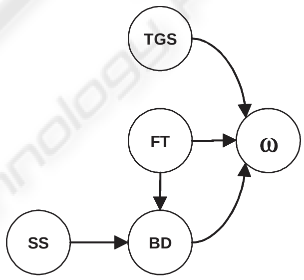

Figure 6: Expert's knowledge (1995).

In 1995, during the Sachem system design phase, the Arcelor Mittal Steel group of experts defines in a knowledge model the variables the modifications of which cause the main modifications of the *omega* variable (figure 6). These variables are the top gas speed (TGS), the flam temperature (FT), the burden permeability (*BD*) and the the size of the sinter (*SS*) but through the burden permeability. Sachem monitors the evolutions of these variables through a specific set of phenomenon classes, but without any knowledge about the causal relation between the variables TGS, FT, *BD*, *SS* and *omega*. Sachem marks the observed modifications of the *omega* variable with occurrences of the 1463 class corresponding to a gradient of the form: $omega(t) = \alpha \cdot t + \beta$, $\alpha \geq \alpha_{min}$. 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). The associated Markov chain is made of $45 \times 45 = 2025$ states. For the 1463 class linked to the *omega* variable, the *BJT4T* algorithm is parameterized to produce a tree of 5 classes depth and 20 classes width, that is to say $20^5 = 3,200,000$ nodes. Such a tree is very difficult to analyze handily. The BJT4P algorithm implementing the $L(M)$ heuristic described in the preceding section produces a pruned tree containing 195 nodes, that is to say a reduction factor of more than 16,000. Parameterized with an anticipating ratio of 50%, the *BJT4S* algorithm produces the signatures of Figure 7, where AR and CR mean respectively Anticipating Rate and Cover Rate.

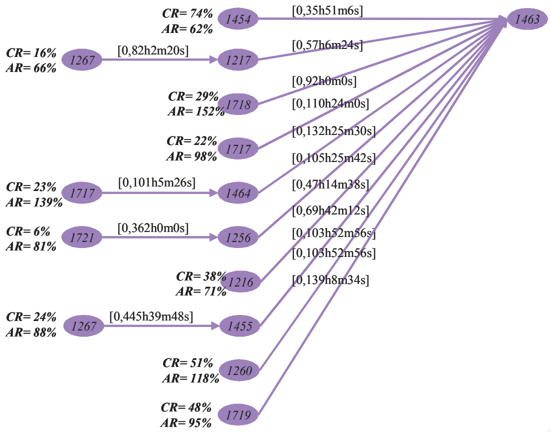

Figure 7: 1463 Class Signatures.

To compare this result with the *a priori* knowledge of the experts in 1995 (figure 6), let us substitutes the class with its associated variable (the *omega* variable with the class 1463 for example) and transform the signatures of Figure 7 in the graph of Figure 8 by merging the nodes having the same variable name. Figure 8 contains then the relations between the variables according to the Stochastic Approach. The graph provided by the Expert’s in 1995 (figure 6) is included in the graph provided by the Stochastic Approach using the BJ-measure (Figure 8). The only difference is the direction of the relation between the variables FT and *BD*. A discussion is then necessary to define this difference because during the development of Sachem, due to the rigor in the dating method of phenomena, Sachem conclusions have lead the experts to inverse their believes about the causality of the relation between some variables. Nevertheless, this result shows that when pruning the branches bringing few information from a class to another, the BJ-measure allows to consider only the branches with a strong potentiality to be a signature: every signatures of Figure 7 have a strong credibility according to the laws governing the underlying process. It is to note that the same result is observed on the Apache system, a clone of Sachem design to monitor and diagnose a galvanization bath.

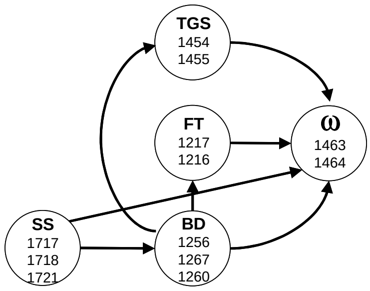

Figure 8: Variable relations (2007).

6 CONCLUSIONS

This paper proposes an adaptation of the J-measure for pruning trees of abstract chronicle models produced according to the Stochastic Approach framework: the BJ-measure. This framework provides a global description of a set of sequences with a probability transition matrix of a Markov chain from which the BJT4T algorithm deduces a tree of the most probable abstract chronicle models. The BJ-measure allows the definition of a heuristic for pruning these trees with the aim of reducing the search space of potential diagnosis rules.

This paper presents the results of the application of this approach to a very complex real world application, an Arcelor Mittal Steel blast furnace monitored with a Sachem knowledge based system. The *a priori* expert’s knowledge about the causal relations between some blast furnace variables as formulated in 1995 has been discovered in 2007 with this approach from a sequence of discrete event class generated by Sachem in 2001.

Our currents works are concerned with the definition of an heuristic for pruning a tree of abstract chronicle models according to its width to improve and generalize the usage of the BJ-measure in the knowledge discovering process of the Stochastic Approach framework.

REFERENCES

Agrawal, R., Imielinski, T., and Swami, A. (1993). Mining association rules between sets of items in large databases. *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, pages 207–216.

Agrawal, R. and Srikant, R. (1995). Mining sequential patterns. *Proceedings of the 11th International Conference on Data Engineering (ICDE95)*, pages 3–14.

Bayardo, R. J. and Agrawal, R. (1999). Mining the most interesting rules. *In Proceedings of ACM KDD1999, ACM Press*, page 145154.

Benayadi, N. and Le Goc, M. (2007). Using an oriented j-measure to prune chronicle models. *To appear in the proceedings of the 18th International Workshop on the Principles of Diagnosis (DX07), Nashville, USA*.

Blachman, N. (1968). The amount of information that y gives about x. *Information Theory, IEEE Transactions*, 14:27–31.

Bouché, P., Le Goc, M., and Giambiasi, N. (2005). Modeling discrete event sequences for discovering diagnosis signatures. *Proceedings of the Summer Computer Simulation Conference (SCSC05) Philadelphia, USA*.

Cauvin, S., Cordier, M.-O., Dousson, C., Laborie, P., Lévy, F., Montmain, J., Montmain, M., Porcheron, M., Servet, I., and Travé, L. (1998). Monitoring and alarm interpretation in in-dustrial environments. *AI Communications ,IOS Press , 1998*, pages 139–173.

Dousson, C. and Duong, T. V. (1999). Discovering chronicles with numerical time constraints from alarm logs for monitoring dynamic systems. *In D. Thomas, editor, Proceedings of the 16th International Joint Conference on Artificial Intelligence (IJCAI-99)*, 1:620–626.

Ghallab, M. (1996). On chronicles: Representation, on-line recognition and learning. *Proc. Principles of Knowledge Representation and Reasoning, Aiello, Doyle and Shapiro (Eds.) Morgan-Kauffman,*, pages 597–606.

Hanks, S. and Dermott, D. M. (1994). Modeling a dynamic and uncertain world i: symbolic and probabilistic reasoning about change. *Artificial Intelligence*, pages 1–55.

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

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

Hilderman, R. and Hamilton, H. (2001). *Knowledge discovery and measures of interest*. Kluwer Academic publishers.

Huynh, X.-H., Guillet, F., and Briand., H. (2005). Arqat: An exploratory analysis tool for interestingness measures. *In Proceedings of the 11th international symposium on Applied Stochastic Models and Data Analysis ASMDA-2005*, page 334344.

Jaroszewicz, S. and Simovici, D. A. (2001). A general measure of rule interestingness. *In Proceedings of PKDD2001, Springer-Verlag*, page 253265.

Le Goc, M. (2004). Sachem, a real time intelligent diagnosis system based on the discrete event paradigm. *Simulation, The Society for Modeling and Simulation International Ed.*, 80(11):591–617.

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.

Le Goc, M., Bouché, P., and Giambiasi, N. (2005). Stochastic modeling of continuous time discrete event sequence for diagnosis. *16th International Workshop on Principles of Diagnosis (DX'05) Pacific Grove, California, USA*.

Liu, B., Hsu, W., Chen, S., and Ma, Y. (2000). Analyzing the subjective interestingness of association rules. *IEEE Intelligent Systems*, page 4755.

Mannila, H. (2002). Local and global methods in data mining: Basic techniques and open problems. *9th International Colloquium on Automata, Languages and Programming, Malaga, Spain,*, 2380:57–68.

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

Padmanabhan, B. and Tuzhilin, A. (1999). Unexpectedness as a measure of interestingness in knowledge discovery. *Decision Support Systems*, pages 303–318.

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

Shannon, C. and Weaver, W. (1949). The mathematical theory of communication. *University of Illinois Press*, 27:379–423.

Shore, J. and Johnson, R. (1980). Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *Information Theory, IEEE Transactions*, 26:26–37.

Smyth, P. and Goodman, R. M. (1992). An information theoretic approach to rule induction from databases. *IEEE Transactions on Knowledge and Data Engineering 4*, page 301316.

Tan, P.-N., Kumar, V., and Srivastava, J. (2004). Selecting the right objective measure for association analysis. *Information Systems*, 29(4):293–313.

Theil, H. (1970). On the estimation of relationships involving qualitative variables. *American Journal of Sociology*, pages 103–154.

Vaillant, B., Lenca, P., and Lallich, S. (2004). A clustering of interestingness measures. *In Proceedings of the 7th International Conference on Discovery Science*, pages 290–297.