Extracting Uncertain Temporal Relations from Mined Frequent Sequences

Francisco Guil
Dept. LyC
Universidad de Almería
04120 - Almería, Spain
fguil@ual.es

Roque Marín
Dept. IIC
Universidad de Murcia
30100 - Murcia, Spain
roque@dif.um.es

Abstract

In this work we address an approach for solving the problem of building a temporal constraint network from the set of frequent sequences obtained after a temporal data mining process. In particular, the temporal data mining algorithm used is TSET [7], an algorithm based on the inter-transactional framework that uses a unique tree-based structure to discover frequent sequences from datasets. The model of temporal network is the proposed by Hadjali, Dubois and Prade [8] where each constraint is formed by three possibility values expressing the relative plausibility of each basic relations between two point-based events, that is, "before", "at the same time" and "after". We propose the use of the Shafer Theory for computing the possibility values of the temporal relations involved in the network from the calculated probability masses of the sequences. The final goal is to obtain a more understandable and useful sort of knowledge from a huge volume of temporal associations resulting after the data mining process.

1 Introduction

Data mining is an essential step in the process of knowledge discovery in databases that consists of applying data analysis and discovery algorithms that produce a particular enumeration of structures over the data [6]. There are two types of structures: models and patterns. So, we can talk about local and global methods in data mining. In the case of local methods, the simplest case of pattern discovery is finding *association rules* [1]. The initial motivation for association rules was to aid in the analysis of large transactional databases. The discovery of association rules can potentially aid decision making within organizations.

Since the problem of mining association rules was introduced by *Agrawal* in [1], a large amount of work has been done in several directions, including improvement of the *Apriori* algorithm, mining generalized, multi-level or quantitative association rules, mining weighted association rules, fuzzy association rules mining, constraint-based rule mining, efficient long patterns mining, etc. We want to point out the work in which a new type of association rules was introduced, the *inter-transaction association rules* [9]. Temporal data mining can be viewed as an extension of the aforementioned work.

Temporal data mining can be defined as the activity of looking for interesting correlations (or patterns) in large sets of temporal data accumulated for other purposes. It has the capability of mining activity, inferring associations of contextual and temporal proximity, that could also indicate a cause-effect association. This important kind of knowledge can be overlooked when the temporal component is ignored or treated as a simple numeric attribute [11]. In non-temporal data mining techniques, there are usually two different tasks: the description of the characteristics of the database (or analysis of the data), and the prediction of the evolution of the population. However, in temporal data mining this distinction is less appropriate, because the evolution of the population is already incorporated in the temporal properties of the data being analyzed.

In [7] we presented an algorithm, named $TSET$, based on the inter-transactional framework for mining frequent sequences from several kind of datasets. The improvement of the proposed solution was the use of a unique structure to store all frequent sequences. The data structure used is the well-known set-enumeration tree, commonly used in the data mining area, in which the temporal semantic is incorporated. The result is a set of frequent sequences describing partially the dataset. This set forms a potential base of temporal information which, after the study and interpretation by human experts, can be very useful to obtain valuable knowledge. However, the overwhelming number of discovered frequent sequences may make such a task absolutely impossible in practice. This problem can be viewed as a second-order data mining problem which consists on the necessity of obtaining a more understandable and useful sort of knowledge from a huge volume of temporal associations

Proceedings of the Thirteenth International Symposium on Temporal Representation and Reasoning (TIME'06)
1530-1311/06 $20.00 © 2006 IEEE

resulting after the data mining process.

In this work, we propose the building of an especial model of temporal networks formed by a set of uncertain relations amongst temporal points. The temporal model, proposed by HadjAli, Dubois and Prade in [8], is based on the Possibility Theory as expressive tool for the representation and management of uncertainty in point-based temporal relations. The uncertainty is represented by a vector describing three possibility values, expressing the relative plausibility of the three basic relations between two temporal points, that is, ”before”, ”at the same time” and ”after”. Also, the authors define the basic operations (inversion, composition, combination and negation) that enable us to infer new temporal information and to propagate uncertainty in a possibilistic way.

After obtaining the sequences base, which are characterized by a basic probability assignment function obtained from the frequencies of the sequences, we propose a Shafer Theory-based technique for calculating the possibility degrees of the basic temporal relations involved in the temporal relations between temporal points.

The remainder of the paper is organized as follows. Section 2 describes briefly the $TSET$ algorithm and gives a formal description of the problem of mining frequent sequences from datasets. Section 3 introduces the formulae for obtaining the sequences base from the set of mined frequent sequences. Section 4 describes briefly the representation aspects of the possibilistic temporal model. In Section 5 we describe the approach for obtaining the uncertain vectors associated with the basic temporal relations from the sequences base. Section 6 presents an example to illustrate the proposed approach. Conclusions and future work are finally drawn in Section 7.

2 The $TSET$ algorithm

$TSET$ is an algorithm designed for mining frequent sequences (or frequent temporal pattern) from large relational datasets. It is based on the 1-dimensional inter-transactional framework [9], and therefore, the aim is to find associations of events amongst different records (or transactions), and not only the associations of events within records. The main improvement of $TSET$ is that it uses a unique tree-based structure to store all frequent sequences. The data structure used is the well known set-enumeration tree [2, 3], in which the temporal semantic is incorporated.

The algorithm follows the same basic principles as most apriori-based algorithms [1]. Frequent sequence mining is an iterative process, and the focus is on a *level-wise* pattern generation. Firstly, all frequent 1-sequences (frequent events) are found, these are used to generate frequent 2-sequences, then 3-sequences are found using frequent 2-sequences, and so on. In other words, (k+1)-sequences are generated only after all k-sequences have been generated. On each cycle, the *downward closure* property is used to prune the search space. This property, also called anti-monotonicity property, indicates that if a sequence is infrequent, then all super-sequence must also be infrequent.

In the sequel, we will introduce the terminologies and the definitions necessary to establish the problem of mining frequent sequences from large datasets.

2.1 Concepts and terminologies

Definition 1 *A* dataset D *is an ordered sequence of records* $D[0]$, $D[1]$,...,$D[r-1]$ *where each* $D[i]$ *can have c columns or attributes,* $A[0], ..., A[c-1]$. *The 0-attribute will be the dimensional attribute, the temporal data associated with the record, expressed in temporal units. The rest of attributes can be quantitative or categorical.*

We assume that the domain of each attribute is a finite subset of non-negative integers, and we also assume that the structure of time is discrete and linear. Due to every event registered has its absolute date identified, we represent the time for events with an absolute dating system [10].

In order to simplify the calculations, we transform the original dataset subtracting the date of each record from the date of the first record, i.e. the time origin.

Definition 2 *An* event e *is a 3-tuple* $(A[i], v, t)$, *where* $0 < i < c$, $v \in dom\{A[i]\}$, *and* $t \in dom\{A[0]\}$, *that is,* $t \in \mathbb{N}$. *Events are ”things that happen”, and they usually represent the dynamic aspect of the world [10].*

In our case, an event is related to the fact that a value v *is assigned to a certain attribute* $A[i]$ *with the occurrence time* t. *The set of all distinct pairs (A[i],v) can be also called* event types. *We will use the notation* $e.a$, $e.v$, *and* $e.t$ *to set and get the attribute, value, and time variables related to the event* e, *and* $e.type$ *to get the event type associated with it.*

Definition 3 *Given two events* e_1 *and* e_2, *we define the* $\leq$ *relation as follows:*

1. $e_1 = e_2$ *iff* $(e_1.t = e_2.t) \wedge (e_1.a = e_2.a) \wedge (e_1.v = e_2.v)$
2. $e_1 < e_2$ *iff* $(e_1.t < e_2.t) \vee ((e_1.t = e_2.t) \wedge (e_1.a < e_2.a))$
3. $e_1 \leq e_2$ *iff* $(e_1 < e_2) \vee (e1 = e2)$

We assume that a lexicographic ordering exists among the pairs (attribute, value), the events types, in the dataset.

Definition 4 *A* sequence (or event sequence) *is an ordered set of events* $S = \{e_0, e_1, ..., e_{k-1}\}$, *where for all* $i < j$, $e_i < e_j$.

Proceedings of the Thirteenth International Symposium on Temporal Representation and Reasoning (TIME'06)
1530-1311/06 $20.00 © 2006 IEEE

Obviously, $|S| = k$. Note that different events with the same temporal unit can belong to the same sequence. Furthermore, the same events with different temporal unit associated can belong to the same sequence. Nevertheless, in any case will exist two or more pairs (attribute, value) associated to the same temporal unit. So, an attribute can not take two different values in the same instant.

Definition 5 *Let U_{tmin} be the minimal dimensional value associated to the sequence S. In other words, $U_{tmin} = min\{e_i.t\}$, for $e_i \in S$. If $U_{tmin} = 0$, we say that S is a* normalized sequence. *Note that any non-normalized sequence can be transformed into a normalized one through a normalization function.*

Let U_{tmax} be the maximal dimensional value associated to the sequence S. This value indicates the maximum distance amongst the events belonging to the normalized sequence S. In other words, $U_{tmax} = e_k.t$, where $|S| = k$. From both, confidence and complexity points of view [9], this value will be always less than or equal to a user-defined parameter called $maxspan$, denoted by ω.

Definition 6 *The* support *(frequency) of a sequence is defined as:*

$$support(S) = \frac{f_r(S)}{|D|},$$

where $f_r(S)$ denotes the number of occurrences of the sequence S in the dataset, and $|D|$ is the number of records in the dataset D, in other words, r.

Definition 7 *A* frequent sequence *is a normalized sequence whose support is greater than or equal to a user-specified threshold called minimum support. We denote this user-defined parameter as $minsup$, or simply σ.*

Given a dataset D and the user-defined parameters ω and σ, the goal of sequence mining is to determine in the dataset the set $\mathcal{S}_f^{D,\sigma,\omega}$, formed by all the frequent sequences whose support are greater than or equal to σ, that is,

$$\mathcal{S}_f^{D,\sigma,\omega} = \{S_i | support(S_i) \geq \sigma\}.$$

This set, formed by a large number of time-stamped sequences, is the goal of the temporal data mining algorithm and the input of the method proposed in this paper for obtaining a temporal constraint network. Basically, the idea is to obtain a global temporal model of the dataset from the discovered set of local temporal patterns. In order to do this, we firstly need to calculate the basic probability assignment using the frequencies of the patterns, obtaining the *sequences base*, whose structure is presented in the next section.

3 The Sequences Base

In order to obtain the basic probability assignment function (bpa function) from the frequencies of the sequences, it is necessary to calculate the number of all possible sequences belonging to the dataset. Note that the set of frequent sequences ($\mathcal{S}_f^{D,\sigma,\omega}$) only gives a partial description of the dataset since it is formed by the sequences whose support exceed a user-given threshold σ. Therefore, we should determine the number of the infrequent sequences presented with the purpose of obtaining a complete description of the dataset. So, we should calculate the number of sequences belonging to the set of infrequent sequences, that we denote as $\mathcal{S}_{\neg f}^{D,\sigma,\omega}$, where

$$\mathcal{S}_{\neg f}^{D,\sigma,\omega} = \{S_i | support(S_i) < \sigma\}$$

and therefore, as

$$\mathcal{S}^{D,\sigma,\omega} = \mathcal{S}_{\neg f}^{D,\sigma,\omega} \cup \mathcal{S}_f^{D,\sigma,\omega}$$

it is possible to calculate the number of all possible sequences presented in the dataset just adding up the frequencies of the sequences in the global set $\mathcal{S}^{D,\sigma,\omega}$. Let

$$F_r(\mathcal{S}) = \sum_i f_r(S_i), \qquad S_i \in \mathcal{S}$$

be the function that calculates the total number of sequences presented in the set $\mathcal{S}$. Obviously,

$$F_r(\mathcal{S}^{D,\sigma,\omega}) = F_r(\mathcal{S}_{\neg f}^{D,\sigma,\omega}) + F_r(\mathcal{S}_f^{D,\sigma,\omega})$$

However, $\mathcal{S}_{\neg f}^{D,\sigma,\omega}$ is an incomplete set formed by the shortest mined infrequent sequences (recall the downward closure property), and therefore, it is impossible to compute $F_r(S_{\neg f}^{D,\sigma,\omega})$. However, as we can see in the sequel, there is a theoretical way to obtain the value of $F_r(\mathcal{S}^{D,\sigma,\omega})$. After obtaining this value, it is ease to calculate $F_r(\mathcal{S}_{\neg f}^{D,\sigma,\omega})$ via the next formula:

$$F_r(\mathcal{S}_{\neg f}^{D,\sigma,\omega}) = F_r(\mathcal{S}^{D,\sigma,\omega}) - F_r(\mathcal{S}_f^{D,\sigma,\omega})$$

3.1 Preliminary results

This subsection contains preliminary results that are needed in the construction of the sequence base.

Based in the sort of dataset, which can be viewed as a $r \times c$ data matrix, and the combinatorial nature of the problem of mining frequent sequences, we propose a formulae for calculating the number of all possible sequences belonging to the dataset in absence of the minimum support constraint, that is, for calculating $F_r(\mathcal{S}^{D,\sigma,\omega})$.

Proceedings of the Thirteenth International Symposium on Temporal Representation and Reasoning (TIME'06)
1530-1311/06 $20.00 © 2006 IEEE

As

$$\mathcal{S}^{D,\sigma,\omega} = \bigcup_{k} \mathcal{S}_k^{D,\sigma,\omega} \qquad 1 \le k \le k_{max},$$

where $\mathcal{S}_k^{D,\sigma,\omega}$ is the set formed by the k-sequences (the sequences of length k), the idea is first to obtain $F_r(\mathcal{S}_k^{D,\sigma,\omega})$, $\forall k$, and finally,

$$F_r(\mathcal{S}^{D,\sigma,\omega}) = \sum_{k=1}^{k_{max}} F_r(\mathcal{S}_k^{D,\sigma,\omega}).$$

The following proposition introduces the formula for calculating k_{max}.

Proposition 1 *Given a $r \times c$ dataset D, and a value for ω, the maximal length of a sequence in the dataset, k_{max}, is:*

$$k_{max} = \begin{cases} (\omega+1)*c & \text{if} \quad \omega < r, \\ c*r & \text{otherwise.} \end{cases}$$

For each $k > 1$, $TSET$ generates the candidate set from the set of frequent $(k-1)$-sequences, access the dataset for counting the occurrences of each candidate sequence using a sliding windows-based approach, and then determines the frequent k-sequences, pruning off those sequences whose support is lower than $minsup$. In Figure 1, the movement of the sliding window is illustrated in order to see the influence of ω in the structure of the event sets studied in the counting phase for a generic 4×2-dataset. The size of the event

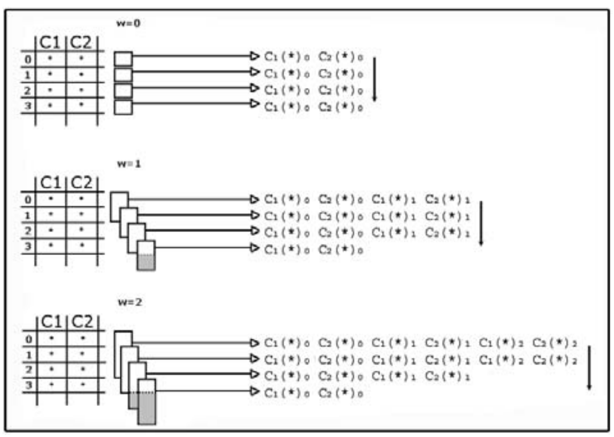

Figure 1. Influence of ω in the event sets formed during the counting phase

set $formedineachstepofthecountingmethod$ can be obtained via the formula expressed by the following proposition.

Proposition 2 *Given a $r \times c$ dataset D, and a sliding window of size ω, the number of studied events in each step, N_i^ω, where $0 \le i < r$:*

$$N_i^\omega = \begin{cases} (\omega+1)*c & \text{if} \quad i + (\omega+1) \le r, \\ (r-i)*c & \text{otherwise.} \end{cases}$$

In order to obtain the total number of possible ω-constrained k-sequences belonging to the dataset D, for every sliding windows, that is, for every record of the dataset, it is necessary to find all possible *normalized k-sequences* from the associated event set of size N_i^ω. In other words, it is necessary to sum up all possible combinations of k events minus the possible combinations of k events which form non-normalized sequences, that can be taken from N_i^ω, where $0 \le i < r$.

Proposition 3 *Given a $r \times c$ dataset D, and a value for ω, the total number of possible normalized k-sequences is:*

$$F_r(S_k^{D,\omega}) = \sum_{i=0}^{r-1} \left[\binom{N_i^\omega}{k} - \mathcal{I}_i \right],$$

where $\mathcal{I}_i$ is the number of non-normalized sequences associated with the sliding window i:

$$\mathcal{I}_i = \begin{cases} \binom{N_i^\omega - c}{k} & \text{if } N_i^\omega - c \ge k, \\ 0 & \text{otherwise.} \end{cases}$$

3.2 Obtaining the Sequences Base

As we have pointed out in the previous section,

$$\mathcal{S}^{D,\sigma,\omega} = \mathcal{S}_{\neg f}^{D,\sigma,\omega} \cup \mathcal{S}_f^{D,\sigma,\omega}.$$

This set, the global set of ω-constrained sequences belonging to the dataset D, is an incomplete set since we only know the structure of the frequent sequences. For easy understanding, let us denote the set of non-frequent sequences, $\mathcal{S}_{\neg f}^{D,\sigma,\omega}$, simply by $\mathcal{S}_\emptyset = \{S_0\}$. So, we redefine $\mathcal{S}^{D,\sigma,\omega}$ in the next definition.

Definition 8 *The set of sequences describing the global dataset is defined as:*

$$\mathcal{S}^{D,\sigma,\omega} = \{S_i | i \ge 0\},$$

where

$$f_r(S_0) = F_r(\mathcal{S}_\emptyset) = F_r(\mathcal{S}_{\neg f}^{D,\sigma,\omega}) = F_r(\mathcal{S}^{D,\sigma,\omega}) - F_r(\mathcal{S}_f^{D,\sigma,\omega}),$$

and for $i \ge 1$, $f_r(S_i)$ is the support value of the mined frequent sequences. It is easy to see that

$$F_r(\mathcal{S}^{D,\sigma,\omega}) = \sum_i f_r(S_i), \quad \text{where} \quad S_i \in \mathcal{S}^{D,\sigma,\omega}$$

Proceedings of the Thirteenth International Symposium on Temporal Representation and Reasoning (TIME'06)
1530-1311/06 $20.00 © 2006 IEEE

Definition 9 *The Sequences Base, $\mathcal{BS}^{\mathcal{D},\sigma,\omega}$, is defined as the set formed by the global sequences and the probability masses associated with them.*

$$\mathcal{BS}^{\mathcal{D},\sigma,\omega} = \{(S_i, m_i)\},$$

where

$$S_i \in \mathcal{S}^{D,\sigma,\omega} \qquad \text{and} \qquad m_i = m(S_i) = \frac{f_r(S_i)}{F_r(\mathcal{S}^{D,\sigma,\omega})}$$

4 Representation of Uncertain Temporal Relations

In literature can be found a large amount of work trying to handle uncertainty in temporal reasoning. However, very few work deal with time points as ontological primitives for expressing temporal elements. Basically, two temporal point-based approaches have been recently proposed for representing and managing uncertain relations between events, the probabilistic model done by Ryabov and Puuronen [12], and the possibilistic model proposed by HadjAli, Dubois, and Prade [8]. In this paper, the authors argued the main differences between these two approaches. Mainly, there are two main differences. First, the possibilistic modeling can be purely qualitative, avoiding the necessity of quantifying uncertainty if information is poor. Second, their proposal is capable of modeling ignorance in a non-biased way. In our case, the selection of the possibilistic model is reinforced by the fact that we need a model which make the fusion of mined and expert knowledge easier [5].

The selected model is based on possibility theory [4] for the representation and management of uncertainty in temporal relations between two point-based events. Uncertainty is represented as a vector involving three *possibility values* expressing the relative plausibility of the three basic relations (" $<$ "," $=$ ", and " $>$ ") that can hold between these points. Also, they describe the inference rules $that form the basis of the reasoning method$ defining a set of operations: inversion, composition, combination, and negation, the operations that govern the uncertainty propagation in the inference process. The authors show that the whole reasoning process can actually be handled in possibilistic logic.

Three basic relations can hold between two temporal points, "before ($<$)", "at the same time ($=$)", and "after ($>$)". An uncertain relation between temporal points is expressed as any possible disjunction of basic relations:

$\leq$	$\Longleftrightarrow$	$<$ *or* $=$
$\geq$	$\Longleftrightarrow$	$>$ *or* $=$
$\neq$	$\Longleftrightarrow$	$<$ *or* $>$
?	$\Longleftrightarrow$	$<, =,$ *or* $>$

The last case represents *total ignorance*, that is, any of the three basic relations is possible. The representation is extended using the Possibility Theory for modeling the plausibility degree of each basic relation. Given two temporal points, a and b, an uncertain relation r_{ab} between them is represented by a *normalized vector* $\Pi_{ab} = (\Pi^{<}_{ab}, \Pi^{=}_{ab}, \Pi^{>}_{ab})$, such that $max(\Pi^{<}_{ab}, \Pi^{=}_{ab}, \Pi^{>}_{ab}) = 1$, where $\Pi^{<}_{ab}$ (respectively, $\Pi^{=}_{ab}$, $\Pi^{>}_{ab}$) is the possibility of $a < b$ (respectively $a = b$, $a > b$).

From the uncertain vector $(\Pi^{<}_{ab}, \Pi^{=}_{ab}, \Pi^{>}_{ab})$, and using the duality between possibility and necessity, namely

$$N(A) = 1 - \Pi(A^c), \quad \text{where } A^c \text{ is the complement of A}$$

we can derive the possibility and necessity degree of each basic relation and their disjunctions.

As,

$$\Pi^{\leq}_{ab} = max(\Pi^{<}_{ab}, \Pi^{=}_{ab})$$

$$\Pi^{\geq}_{ab} = max(\Pi^{=}_{ab}, \Pi^{>}_{ab})$$

$$\Pi^{\neq}_{ab} = max(\Pi^{<}_{ab}, \Pi^{>}_{ab}),$$

we can obtain the necessity degrees of the basic relations,

$$N^{<}_{ab} = N(a < b) = 1 - \Pi^{\geq}_{ab}$$

$$N^{=}_{ab} = N(a = b) = 1 - \Pi^{\neq}_{ab}$$

$$N^{>}_{ab} = N(a > b) = 1 - \Pi^{\leq}_{ab}.$$

In a similar way, we can also obtain

$$N^{\geq}_{ab} = N(a \geq b) = 1 - \Pi^{<}_{ab}$$

$$N^{\neq}_{ab} = N(a \neq b) = 1 - \Pi^{=}_{ab}$$

$$N^{\leq}_{ab} = N(a \leq b) = 1 - \Pi^{>}_{ab}.$$

Moreover, the authors defined the rules that enable us to infer new temporal information and to propagate uncertainty in a possibilistic way. The reasoning tool relies on four operations expressing:

$inversion$	$\Longleftrightarrow$	$\widetilde{r}_{ab} = r_{ba}$
$composition$	$\Longleftrightarrow$	$r_{ac} = r_{ab} \otimes r_{bc}$
$combination$	$\Longleftrightarrow$	$r_{ab} = r_{1_{ab}} \oplus r_{2_{ab}}$
$negation$	$\Longleftrightarrow$	$\neg$

These rules completes the definition of a model for representing and reasoning with uncertain temporal relations that uses the Possibility Theory as an expressive tool for dealing with uncertainty in temporal reasoning.

Proceedings of the Thirteenth International Symposium on Temporal Representation and Reasoning (TIME'06)
1530-1311/06 $20.00 © 2006 IEEE

5 Extracting Uncertain Temporal Relations

In this section, we propose a technique for extract the uncertain temporal relation between each pair of event types from the sequences base. The uncertain temporal relation is represented by an uncertain vector formed by three possibility values, expressing the plausibility degree for each basic temporal relation. We propose the use of Shafer Theory of Evidence [13] to obtain the plausibility degrees from the probability masses associated with the global set of sequences. The result will be a temporal constraint network, which is a suitable model for representing and reasoning with temporal information where uncertainty is presented.

5.1 Shafer's Theory of Evidence

The Shafer Theory of Evidence, also known as Dempster-Shafer Theory, is a theory of uncertainty developed specially for modelling complex systems. It is based on a special fuzzy measure called *belief measure*. Beliefs can be assigned to propositions to express the uncertainty associated to them being discerned. Given a finite universal set $\mathcal{U}$, the *frame of discernment*, the beliefs are usually computed based on a density function $m : 2^{\mathcal{U}} \rightarrow [0,1]$ called *basic probability assignment* (bpa):

$$m(\emptyset) = 0, \text{ and } \sum_{A \subseteq \mathcal{U}} m(A) = 1.$$

$m(A)$ represents the belief exactly committed to the set A. If $m(A) > 0$, then A is called a *focal element*. The set of focal elements constitute a core:

$$\mathcal{F} = \{A \subseteq \mathcal{U} : m(A) > 0\}$$

The core and its associated bpa define a *body of evidence*, from where a belief function $Bel : 2^{\mathcal{U}} \rightarrow [0,1]$ is defined:

$$Bel(A) = \sum_{B | B \subseteq A} m(B)$$

For any given measure Bel, a *dual measure*, $Pl : 2^{\mathcal{U}} \rightarrow [0,1]$ can be defined:

$$Pl(A) = 1 - Bel(\overline{A}).$$

So, this measure called *plausibility measure*, can be also defined:

$$Pl(A) = \sum_{B | B \cap A \neq \emptyset} m(B).$$

It can be verified [13] that the functions Bel and Pl are, respectively, a possibility (or necessity) measure if and only if the focal elements form a nested or consonant set. A weaker statement consists in establishing the equivalence when the intersection of the focal subsets are not empty.

5.2 Calculating the possibility measures of temporal relations

In our proposal the core is the sequences base $\mathcal{BS}^{D,\sigma,\omega}$ which is formed by a set of focal elements or sequences.

Let Ω be the set of event types presented in the dataset, that is,

$$\Omega = \{(A[i], v) | v \in dom(A[i])\}.$$

Taking into account the *maxspan* constraint, the set of events is defined as an extension of the Ω set in this way:

$$\Omega^{\omega} = \{(A[i], v, t) | v \in dom(A[i]) \wedge 0 \leq t \leq w)\}$$

This set is our frame of discernment, that is, $\Omega^{\omega} = \mathcal{U}$. So, the set of focal elements, the sequence base, is defined:

$$\mathcal{BS}^{D,\sigma,\omega} = \{S_i \subseteq \Omega^{\omega} | m(S_i) > 0\},$$

where m is the bpa function derived from the frequencies of the sequences, such that $m : 2^{\Omega^{\omega}} \rightarrow [0,1]$,

$$m(\emptyset) = 0, \sum_i m(S_i) = 1$$

We will denote a temporal relation between two events e_1, e_2 as $e_1 \Theta e_2$. Since we are only interested in the basic temporal relations,

$$\Theta \in \{<, =, >\}.$$

For each pair of event types, we need to obtain the possibility degree of each basic temporal relation between them. In order to compute the possibility of a temporal relation, it is necessary to consider all focal elements, that is, all sequences which make the temporal relation possible.

Proposition 4 *Let suppose the qualitative temporal relation $e_1 \Theta e_2$. This relation induces a parameterized set:*

$$\mathcal{X}_{e_1 \Theta e_2} = \{(e_i e_j)\},$$

where $e_i, e_j \in \Omega^{\omega}, e_i.type = e_1$, $e_j.type = e_2$, and $e_i.t \Theta e_j.t$.

Proposition 5 *In order to obtain the set of sequences involved in the temporal relation, we introduce the assessment operator Γ, defined as:*

$$\Gamma(\mathcal{X}_{e_1 \Theta e_2}) = \{S_i | S_i \cap \mathcal{X}_{e_1 \Theta e_2} \neq \emptyset\},$$

where $S_i \in \mathcal{BS}^{D,\sigma,\omega}$.

Proposition 6 *The possibility degree of the temporal relation $e_1 \Theta e_2$ is defined as:*

$$\Pi(e_1 \Theta e_2) = Pl(e_1 \Theta e_2) = \sum_{S_i \in \Gamma(\mathcal{X}_{e_1 \Theta e_2})} m(S_i)$$

Proceedings of the Thirteenth International Symposium on Temporal Representation and Reasoning (TIME'06)
1530-1311/06 $20.00 © 2006 IEEE

Table 1. An example dataset

T	C_1	C_2	T	C_1	C_2
0	0	0	15	1	1
1	1	2	16	0	1
2	2	1	17	1	2
3	0	1	18	0	1
4	1	0	19	2	1
5	1	1	20	1	2
6	1	0	21	1	1
7	1	1	22	1	2
8	1	2	23	2	1
9	2	1	24	1	1
10	1	2	25	0	0
11	0	1	26	1	1
12	2	1	27	2	1
13	1	1	28	1	0
14	0	0	29	0	1

6 Example

Let us see an example using a synthetic dataset to illustrate the technique proposed in this work. Suppose that, after some preprocessing, we have obtained the dataset D shown in Table 1, where T represents the temporal attribute expressed in temporal units, and C_1 and C_2 two descriptive attributes $the 0 and 1 attributes, respectively$. We assume that the user set $\sigma = 25\%$ and $\omega = 2$. We have that

$$\Omega = \{(0,0),(0,1),(0,2),(1,0),(1,1),(1,2)\}$$

, and therefore, $|\Omega^\omega| = |\Omega| * (\omega + 1) = 18$. Using the formulae defined in the Section 3, we obtain $F_r(\mathcal{S}^{D,\sigma,\omega}) = 603$. Considering the given user-defined parameters setting, $TSET$ discover the following set of frequent sequences, $\mathcal{S}_f^{D,\sigma,\omega}$:

$S_1 = \{(0,1,0)\}, sup = 0.5333$
$S_2 = \{(0,2,0)\}, sup = 0.2000$
$S_3 = \{(1,1,0)\}, sup = 0.6000$
$S_4 = \{(1,2,0)\}, sup = 0.2000$
$S_5 = \{(0,1,0),(1,1,0)\}, sup = 0.2333$
$S_6 = \{(0,1,0),(1,2,0)\}, sup = 0.2000$
$S_7 = \{(0,1,0),(0,1,1)\}, sup = 0.2000$
$S_8 = \{(0,1,0),(1,1,1)\}, sup = 0.3667$
$S_9 = \{(0,1,0),(0,1,2)\}, sup = 0.3333$
$S_{10} = \{(0,1,0),(1,1,2)\}, sup = 0.3000$
$S_{11} = \{(0,2,0),(1,1,0)\}, sup = 0.2000$
$S_{12} = \{(1,1,0),(0,1,1)\}, sup = 0.3333$
$S_{13} = \{(1,1,0),(1,1,1)\}, sup = 0.2333$
$S_{14} = \{(1,1,0),(0,1,2)\}, sup = 0.3333$
$S_{15} = \{(1,1,0),(1,1,2)\}, sup = 0.3667$
$S_{16} = \{(1,2,0),(1,1,1)\}, sup = 0.2000$
$S_{17} = \{(0,1,0),(1,2,0),(1,1,1)\}, sup = 0.2000$
$S_{18} = \{(0,1,0),(1,1,1),(0,1,2)\}, sup = 0.2333$
$S_{19} = \{(1,1,0),(0,1,1),(1,1,2)\}, sup = 0.2667$
$S_{20} = \{(1,1,0),(0,1,2),(1,1,2)\}, sup = 0.2000$

As $F_r(\mathcal{S}_f^{D,\sigma,\omega}) = 172$, $F_r(\mathcal{S}_{\neg f}^{D,\sigma,\omega}) = 431$. Now, we can obtain the basic assignment function from the frequencies of the sequences. In Table 2 we show the structure of the Sequences Base $\mathcal{BS}^{D,\sigma,\omega}$, that is, the body of evidence. In this example, there are 4 frequent event types, $a = (0,1)$, $b = (0,2)$, $c = (1,1)$, and $d = (1,2)$. Let us calculate now the uncertain vectors for each pair of type events. For the pair a, b, we show the basic temporal relations, the set of sequences in which the temporal relations holds and the value of the possibility degrees.

r_{ab} :

	Γ	Π
$\Pi_{ab}^{<}$	S_0	0.715
$\Pi_{ab}^{=}$	S_0	0.715
$\Pi_{ab}^{>}$	S_0	0.715

So, $r_{ab} = (0.715, 0.715, 0.715) = (1,1,1)$, which is equivalent to the *total ignorance* state. This possibilistic distribution means that we do not known anything about the relative position between the events a and b. Let us see the uncertain vector for the temporal relation between a and c:

r_{ac} :

	Γ	Π
$\Pi_{ab}^{<}$	$S_0, S_8, S_{10}, S_{17}, S_{18}, S_{19}$	0.783
$\Pi_{ab}^{=}$	S_0, S_5, S_{20}	0.737
$\Pi_{ab}^{>}$	$S_0, S_{12} S_{14}, S_{18}, S_{19}, S_{20}$	0.784

We obtain that $r_{ac} = (0.783, 0.737, 0.784)$, which is transformed into the normalized vector $r_{ac} = (1, 0.94, 1)$. Following the same calculi, we obtain the uncertain vectors:

$$r_{ad} = (0, 0.735, 0) = (0,1,0),$$

$$r_{bc} = (0, 0.725, 0) = (0,1,0),$$

$$r_{bd} = (1,1,1), and$$

$$r_{cd} = (0,0,0.035) = (0,0,1).$$

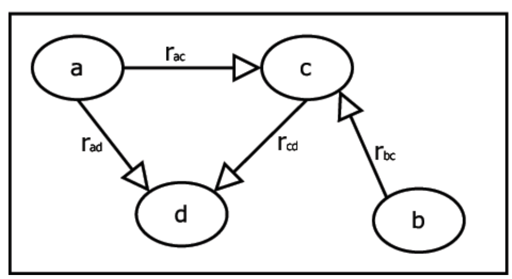

Figure 2. The obtained network

In Figure 2 we can see the structure of the network from the second-order data mining process.

Proceedings of the Thirteenth International Symposium on Temporal Representation and Reasoning (TIME'06)
1530-1311/06 $20.00 © 2006 IEEE

Table 2. Sequences Base

i	S_i	f_r	m
0	$\{?\}$	431	0.715
1	$\{(0,1,0)\}$	16	0.027
2	$\{(0,2,0)\}$	6	0.010
3	$\{(1,1,0)\}$	18	0.030
4	$\{(1,2,0)\}$	6	0.010
5	$\{(0,1,0),(1,1,0)\}$	7	0.012
6	$\{(0,1,0),(1,2,0)\}$	6	0.010
7	$\{(0,1,0),(0,1,1)\}$	6	0.010
8	$\{(0,1,0),(1,1,1)\}$	11	0.018
9	$\{(0,1,0),(0,1,2)\}$	10	0.017
10	$\{(0,1,0),(1,1,2)\}$	9	0.015
11	$\{(0,2,0),(1,1,0)\}$	6	0.010
12	$\{(1,1,0),(0,1,1)\}$	10	0.017
13	$\{(1,1,0),(1,1,1)\}$	7	0.012
14	$\{(1,1,0),(0,1,2)\}$	10	0.017
15	$\{(1,1,0),(1,1,2)\}$	11	0.018
16	$\{(1,2,0),(1,1,1)\}$	6	0.010
17	$\{(0,1,0),(1,2,0),(1,1,1)\}$	6	0.010
18	$\{(0,1,0),(1,1,1),(0,1,2)\}$	7	0.012
19	$\{(1,1,0),(0,1,1),(1,1,2)\}$	8	0.013
20	$\{(1,1,0),(0,1,2),(1,1,2)\}$	6	0.010
		603	1.000

7 Conclusions and Future Work

In this paper, we propose an initial approach for building temporal constraint networks from a set of mined frequent sequences with the aim of obtaining a more understandable, useful and manageable sort of knowledge. The selected temporal model belongs to a special class of model $proposedbyHadjAli, Dubois, andPrade$, which uses the Possibility Theory as an expressive tool for representing and reasoning with uncertain temporal relations between point-based events. The uncertainty is represented as a vector involving three possibility values expressing the relative plausibility of the three basic relations between two temporal points, that is, ”before”, ”at the same time” and ”after”. We propose Shafer’s Theory-based technique to obtain these plausibility degrees from the calculated basic probability assignment function from the frequencies of the sequences.

In future work, we intend to analyze in depth the networks obtained from the set of mined frequent sequences. We also propose to extend the model of temporal network in order to represent not only qualitative but also quantitative temporal relations, taking advantage of the temporal information presented in the time-stamped sequences extracted by $TSET$.

References

[1] R. Agrawal, T. Imielinski, and A. N. Swami. Mining association rules between sets of items in large databases. In P. Buneman and S. Jajodia, editors, *Proc. of the ACM SIGMOD Int. Conf. on Management of Data, Washington, D.C., May 26-28, 1993*, pages 207–216. ACM Press, 1993.

[2] R. J. Bayardo. Efficiently mining long patterns from databases. In L. M. Haas and A. Tiwary, editors, *Proc. of the ACM SIGMOD Int. Conf. on Management of Data (SIGMOD 1998), June 2-4, 1998, Seattle, Washington, USA*, pages 85–93. ACM Press, 1998.

[3] F. Coenen, G. Goulbourne, and P. Leng. Tree structures for mining association rules. *Data Mining and Knowledge Discovery*, 8:25–51, 2004.

[4] D. Dubois and H. Prade. *Possibility Theory*. Plenum Press, 1988.

[5] D. Dubois, H. Prade, and G. Yager. Merging fuzzy information. In *Fuzzy Sets in Approximate Reasoning and Information Systems*, pages 335–401. Kluwer Academic Publishers, 1999.

[6] U. Fayyad, G. Piatetky-Shapiro, and P. Smyth. From data mining to knowledge discovery in databases. *AIMagazine*, 17(3):37–54, 1996.

[7] F. Guil, A. Bosch, and R. Marín. TSET: An algorithm for mining frequent temporal patterns. In *Proc. of the First Int. Workshop on Knowledge Discovery in Data Streams, in conjunction with ECML/PKDD 2004*, pages 65–74, 2004.

[8] A. HadjAli, D. Dubois, and H. Prade. A possibility theory-based approach for handling of uncertain relations between temporal points. In *11th International Symposium on Temporal Representation and Reasoning (TIME 2004)*, pages 36–43. IEEE Computer Society, 2004.

[9] H. Lu, L. Feng, and J. Han. Beyond intra-transaction association analysis: Mining multi-dimensional inter-transaction association rules. *ACM Transactions on Information Systems (TOIS)*, 18(4):423–454, 2000.

[10] A. K. Pani. Temporal representation and reasoning in artificial intelligence: A review. *Mathematical and Computer Modelling*, 34:55–80, 2001.

[11] J. F. Roddick and M. Spiliopoulou. A survey of temporal knowledge discovery paradigms and methods. *IEEE Transactions on Knowledge and Data Engineering*, 14(4):750–767, 2002.

[12] V. Ryabov and S. Puuronen. Probabilistic reasoning about uncertain relations between temporal points. In *8th International Symposium on Temporal Representation and Reasoning (TIME 2001)*, pages 1530–1511. IEEE Computer Society, 2001.

[13] G. Shafer. *A Mathematical Theory of Evidence*. Princenton University Press, Princenton, NJ, 1976.

Proceedings of the Thirteenth International Symposium on Temporal Representation and Reasoning (TIME'06)