Probabilistic Reasoning with Temporal Indeterminacy

Maarten van der Heijden and Peter J.F. Lucas
Radboud University Nijmegen, The Netherlands
{m.vanderheijden,peterl}@cs.ru.nl

Abstract

Temporal indeterminacy, the lack of specific knowledge about the timing of events, occurs often in temporal reasoning in practical applications and is connected to the concept of time granularity. Although logical properties of granularities have been described by several researchers in the literature, the implications of temporal indeterminacy and granularities for probabilistic representation and reasoning have received little attention. Given the widespread occurrence of problems, specifically in medicine, where one has to cope with both temporal indeterminacy and uncertainty, it is somewhat surprising that methods that handle both do not exist as yet. In this paper we propose a formalism to model granularities in temporal Bayesian networks in order to deal with temporal indeterminacy in predictive Bayesian-network models. In addition, some of the properties of multigranular models are explored. Finally, we study a medical use case.

1 Introduction

In clinical medicine, as in many other fields, one frequently has to deal with data that is uncertain and records temporal progression. The time aspect is often also uncertain, which is sometimes referred to as *temporal indeterminacy* (Combi and Pozzi, 2001). To build useful predictive models we need to deal with these two kinds of uncertainty. A typical example in clinical medicine would be diagnosis based on a patient report of symptoms, e.g. over the last week. From a modelling perspective it may be uncertain *which* symptoms occurred, and, while our knowledge of *when* the symptoms occurred is constrained to last week, within that period we may not be able to be more specific.

There are many methods that support modelling processes of time; dynamic Bayesian networks (of which hidden Markov models are a special case) are an example (Dean and Kanazawa, 1989; Murphy, 2002). The problem of temporal indeterminacy and its relation to temporal granularities has been recognised in the database community (e.g. Combi and Pozzi (2001)). However, the fundamental problem of dealing with temporal indeterminacy in predictive models is mostly unsolved.

The representational power of Bayesian networks makes them a useful tool in a clinical context to represent (causal) relations between symptoms, signs and diseases. We have developed a decision support system for chronic obstructive pulmonary disease (COPD), where patients are monitored at home to detect exacerbation events. In this system a Bayesian network is used for automatic data interpretation (van der Heijden et al., 2011). The dynamic nature of the disease process leads naturally to the desire to employ a model that mirrors the disease and clinical practice more closely than a static Bayesian network. The first step is then to extend the Bayesian network to a dynamic Bayesian network (DBN) which allows us to model dependences through time.

Temporal indeterminacy results from the harsh reality of patient monitoring in a home environment, in the sense that it may not always be possible to obtain precise information at the right time. We want to be capable of dealing with indeterminacy of the kind "symptom S appeared between 2 and 4 days ago". This problem is not limited to monitoring, but ap-

pears often in clinical practice. We can identify similar aggregation problems: when lots of data is available (e.g. on the intensive care unit), aggregation as a summary is sometimes useful; a situation that occurs frequently is that measurements are taken at different times but the clinician is interested in an aggregated result that represents the health status of the patient over the period the measurements were taken. These problems have been studied in the context of temporal abstraction (Shahar, 1997), but not in probabilistic graphical models.

We aim to develop a framework to specify temporal indeterminacy, and related aggregation problems, probabilistically. In this paper we focus on this representational problem in the context of dynamic Bayesian networks, with the objective to make predictive models that handle indeterminacy.

2 Preliminaries

2.1 Representing time

In order to reason about temporal indeterminacy we need a formalism to represent time and granularity. The work on representing time in temporal databases seems useful as a starting point (Combi and Pozzi, 2001). We take a linearly ordered set of points $(\mathcal{T}, \leq)$ as the primitive representation of a time line, where $\mathcal{T}$ is a subset of the natural numbers $\mathbb{N}$. A *determined instant* at the lowest time granularity (e.g. seconds) is a point $t \in \mathcal{T}$. Due to temporal indeterminacy and granularities, it may be useful to represent an instant as a set. For example, an instant at the scale of minutes is a single minute, but may be represented as a set of 60 seconds at a finer granularity. Similarly, when the exact time of an instant is unknown, a set of points can be used to represent the uncertainty in time, that is, to represent the temporal indeterminacy. An *indeterminate instant* is defined as a set $a \subseteq \mathcal{T}$, with the property that a is contiguous: $\forall t \in \mathcal{T} : \inf a \leq t \leq \sup a \Leftrightarrow t \in a$, using $\inf a$ and $\sup a$ to denote the lower and upper bound of the interval of indeterminacy. Two instants a, b are called non-overlapping if $\sup a < \inf b$ or $\sup b < \inf a$.

In studies on logical representations of granularities (e.g. Bettini et al. (1998)), different time structures are allowed as granularities. This allows one to specify hierarchies of granularities for example for the Gregorian calendar. Here we define a granularity G in a more restricted sense as a partition of $\mathcal{T}$, i.e. a set of subsets such that $\bigcup G = \mathcal{T}$ and if $g, g' \in G$ then $g \cap g' = \varnothing$ or $g = g'$. We further consider only granularities that are contiguous, uniform and comparable, which means that every time point can be expressed in each granularity and time points within a granularity have the same size (which excludes for example 'month'). Examples of possible granularities with these restrictions include 'second', 'hour', 'day' etc.

Before we go on we introduce some notation. We are often interested in events associated with time. In particular we consider events of observing a value of a certain variable of interest. We write E_a for an event E that occurs in a, where a is an indeterminate instant. It is sometimes convenient to specify a certain time point t within a. For example, to denote the probability that E occurs at $t \in a$, with a an instant, we write $P(E_a(t))$. The complement event of E_a is denoted $\bar{E}_a$.

2.2 Bayesian networks

To be able to add temporal indeterminacy to temporal Bayesian networks, we first define the usual way to explicitly incorporate time in Bayesian networks.

A *dynamic Bayesian network* is a pair $(G, \mathbf{F})$, with G a graph $G = (\mathbf{V}, \mathbf{A})$ and $\mathbf{F}$ a set of factors over random variables $\mathbf{X}$ corresponding to the vertices $\mathbf{V}$. The set of arcs in the graph $\mathbf{A} \subseteq \mathbf{V} \times \mathbf{V}$ represents (in)dependences of the variables. For a two-slice dynamic Bayesian network we subdivide the vertices in an initial slice and a repeated transition slice $\mathbf{V} = \mathbf{V}_0 \cup \mathbf{V}_{1:T}$, which implies a similar subdivision $\mathbf{X} = \mathbf{X}_0 \cup \mathbf{X}_{1:T}$. An arc $(v, w) \in \mathbf{A}$ with $v \in \mathbf{V}_t$ and $w \in \mathbf{V}_{t'}$, $t < t'$, denotes a temporal dependence. Let $pa(X)$ denote the parents of $X \in \mathbf{X}$ in the graph G. The set $\mathbf{F}$ contains factors for each $X \in \mathbf{X}$ such that $f(x, \mathbf{p}) = P(X = x \mid pa(X) = \mathbf{p})$. The joint

distribution over $\mathbf{X} = \bigcup_i^T \mathbf{X}_i$, factorises over the graph such that $P(\mathbf{X}) = \prod_{X \in \mathbf{X}} P(X \mid pa(X))$.

To simplify the models, often only first-order Markov models are considered, which means that the future is independent of the past given the present:

$$P(X_{t+1} \mid X_t, X_{1:t-1}) = P(X_{t+1} \mid X_t).$$

Furthermore, a usual assumption is homogeneity, also called stationarity, which means that the probabilities are equal between time steps:

$$P(X_t \mid pa(X_t)) = P(X_{t'} \mid pa(X_{t'})) \quad \forall t, t' : X_t \in \mathbf{X}_t, X_{t'} \in \mathbf{X}_{t'}.$$

This may be too strong an assumption for real processes, which has lead to recent work on learning non-stationary dynamic Bayesian networks (Robinson and Hartemink, 2010; Grzegorczyk and Husmeier, 2011). We leave these assumptions intact here, and focus only on the indeterminacy problem.

A technique often used to model the interaction between multiple variables (to prevent an exponential growth of parameters) are *causal independence* models (Pearl, 1988). The main assumption is that different causes of an effect can be assumed to be independent and only interact through a particular deterministic function. Here we will use this method to model the interaction between random variables at different time granularities.

3 Events and granularity

We first state some properties of indeterminate events, and their consequences. Consider an event E_a to be a point-like occurrence at the finest granularity under consideration somewhere at a. We then have:

$$E_a = \bigvee_{t \in a} E_a(t), \tag{1}$$

i.e., event E_a is defined in terms of events at the finer granularity. Analogously, the complement event is:

$$\bar{E}_a = \neg \bigvee_{t \in a} E_a(t) = \bigwedge_{t \in a} \bar{E}_a(t). \tag{2}$$

The following property ensures that only a single point event $E_a(t)$ is true:

$$\forall t, t' \in a : t \neq t' \rightarrow E_a(t) \wedge E_a(t') = \bot. \tag{3}$$

called the *single event assumption*. This means that the event at coarser granularity is caused by exactly one event at finer granularity. Conversely, if multiple events could have occurred at the finer granularity, this is called the *multiple event assumption*.

When using probability theory to capture the principles of an indeterminate instant a, we need to specify a distribution for the occurrence of an event E associated with the instant a. Under the single event assumption a distribution for E_a has the property:

$$P(E_a) = \sum_{t \in a} P(E_a(t)),$$

which follows from the two properties above. From Property (3) it also follows directly that:

$$\forall t, t' \in a : t \neq t' \rightarrow P(\dots, E_a(t), \dots, E_a(t'), \dots) = 0 \tag{4}$$

and

$$P\Big(\bigwedge_{t \in a \setminus \{t'\}} \bar{E}_a(t) \mid E_a(t') \Big) = 1.$$

From this, it follows that

$$\begin{aligned} &P\Big(\bigwedge_{t \in a \setminus \{t'\}} \bar{E}_a(t), E_a(t') \Big) \\ &= P\Big(\bigwedge_{t \in a \setminus \{t'\}} \bar{E}_a(t) \mid E_a(t') \Big) P(E_a(t')) \\ &= 1 \cdot P(E_a(t')) = P(E_a(t')). \end{aligned} \tag{5}$$

However, when the multiple event assumption is adopted, only

$$P(E_a) = P\left(\bigvee_{t \in a} E_a(t) \right)$$

holds. This means that at the aggregated level we cannot distinguish between what are multiple events at the precise granularity. We are interested in somehow relating the representations at both granularities.

Further constraints result from domain specific patterns of dependence between points in a, which will be considered later.

4 Temporal indeterminacy in DBNs

Now we focus on modelling temporal indeterminacy in dynamic Bayesian networks. In general the problem is representing a temporal process at multiple granularities.

4.1 Probabilistic aggregation

A conceptually simple representation consists of separate models for the granularities. This is shown in Figure 1. Basically, the lower part of the model aggregates the top part, although here it is not specified how the aggregation takes place. A one-to-one translation from the possible states of the finer granularity to the coarser granularity results in:

$$P(O_a|X_a)P(X_a) = P(\bigwedge_{i=1}^{n} O_i | \bigwedge_{j=1}^{n} X_j) P(\bigwedge_{k=1}^{n} X_k),$$

where basically the domain of O_a is the Cartesian product of the domains of $O_1, \ldots, O_n$; similarly, the domain of X_a is the Cartesian product of the domains of $X_1, \ldots, X_n$. The result is a representation that is exponential in the size of the domains.

With the single event assumption it is easily possible to compute the aggregated probabilities. However, when represented graphically, these logical constraints would require dropping the first-order Markov assumption yielding a factorisation of $P(X_1, \ldots, X_n)$ without any independence information. However, in the transformation it is not necessary to use a graphical representation as there is a very specific relation between the granularities. Let X_a take on values in $\{0, 1, \ldots, n\}$ then $P(X_a = i) = P(X_i = 1)$ and $P(X_a = 0) = 1 - \sum_{i=1}^{n} P(X_a = i)$, because by definition for i with $1 \leq i \leq n$:

$$\begin{aligned}
&P(X_a = i) \\
&= P(X_1 = 0, \ldots, X_i = 1, \ldots, X_n = 0) \\
&= P(X_n = 0, \ldots, X_{i+1} = 0 | X_i = 1, \ldots, X_1 = 0) \\
&\quad \cdot P(X_i = 1, X_{i-1} = 0, \ldots, X_1 = 0) \\
&= 1 \cdot P(X_i = 1, X_{i-1} = 0, \ldots, X_1 = 0) \\
&= P(X_i = 1)
\end{aligned}$$

using Equation (5). Note that it follows that

$$P(X_a \neq i) = P(X_i = 0).$$

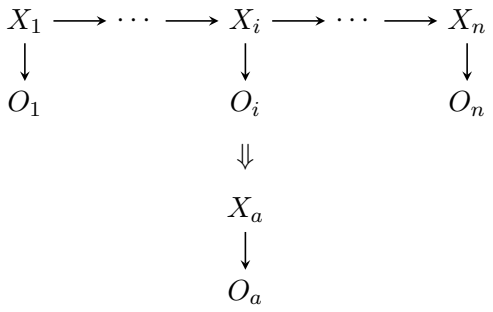

Figure 1: A general depiction of a model at two granularities (without the single event assumption).

$$\begin{aligned}
P(X_i = 1) &= \sum_{X_j, i \neq j} P(X_i = 1, \bigwedge\nolimits_j X_j) \\
&= P(X_i = 1, \bigwedge\nolimits_j X_j = 0),
\end{aligned}$$

where the second equality follows from mutual exclusivity, as is easily seen from Equation (4). Since this is true for X_i with $1 \leq i \leq n$ and $P(X_a)$ is normalised by setting $P(X_a = 0) = 1 - \sum_{i=1}^{n} P(X_a = i)$, we obtain the aggregation $P(X_a) = P(\bigwedge_{i=1}^{n} X_i)$.

To aggregate the observation variables O_i we take the domain of O_a to be binary. Since we assumed stationarity we have only two parameters: for $1 \leq i \leq n$ $P(O_i = 1 \mid X_i = 1) = p$ and $P(O_i = 1 \mid X_i = 0) = q$. Still assuming that the first-order Markov assumption does not hold, conditioned on the X_i's the O_i variables are conditionally independent, allowing us to aggregate the observations by multiplying the conditional probabilities; for $1 \leq i \leq n$, it holds that:

$$\begin{aligned}
P(O_a = 1 \mid X_a = i) &= P(O_i = 1 \mid X_i = 1) \\
&\quad \cdot \prod_{j \neq i} P(O_j = 1 \mid X_j = 0) = pq^{n-1}
\end{aligned}$$

and in addition

$$\begin{aligned}
&P(O_a = 1 \mid X_a = 0) \\
&\qquad = \prod_{j} P(O_j = 1 \mid X_j = 0) = q^n.
\end{aligned}$$

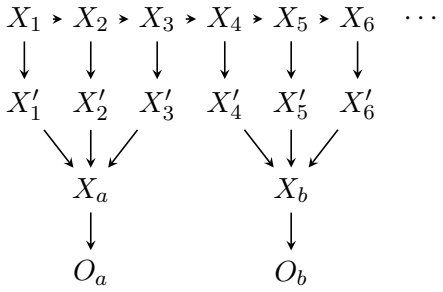

Figure 2: Incorporating granularity within the model. Observation variables for the finer granularity have been omitted.

4.2 Structural aggregation

A different perspective on modelling indeterminacy is including the aggregate temporal structure within the model. This turns out to be useful as it allows us to model explicitly the interaction between the variables at different granularities. The mechanism that we will employ to model multiple interactions in a systematic way is *causal independence*. This is depicted in Figure 2. A causal independence model allows us to model the projection by means of a particular deterministic function of the variables at the finer granularity.

Although separating the state and observation model is common practice to model measurement errors, we will for now assume that state variables are directly observable in order to focus on modelling the indeterminacy. Hence, disregarding the O variables in Figure 2 we obtain the following joint distribution for the aggregation over a

$$P(X_a, \bigwedge_{i=1}^{n} X'_i, X_i) = P(X_a | \bigwedge_{i=1}^{n} X'_i) \prod_j P(X'_j | X_j)$$
$$P(X_1) \cdots P(X_i \mid X_{i-1}) \cdots P(X_n \mid X_{n-1}). \quad (6)$$

In a causal independence model the term $P(X_a \mid \bigwedge_{i=1}^{n} X'_i)$ is represented by a deterministic function. Equation (1) implies that the logical OR is a natural interaction function.

As the single event assumption is easily represented by the construction in the previous section, we now focus on the multiple event assumption. The causal independence model allows us to combine the events, even when the independence assumption on the causes is not met (as can be seen to be the case in Figure 2). Although the model complexity increases, the number of parameters is still linear in the number of variables at the fine granularity, as opposed to the exponential increase which results when connecting the granularities directly.

The model depicted in Figure 2 is not the only possible aggregation. An aggregation variable can summarise arbitrary sets of state variables, which is related to the work by Evers et al. (2008). The statement "symptom S occurred between 2 and 4 days ago" can be modelled with an aggregation variable that summarises 3 variables at the granularity of days. These aggregations can be made to overlap, that is X_a summarises $X_{1:3}$, X_b summarises $X_{2:4}$ etcetera.

Aggregation patterns

Starting from the aggregated level we can use domain knowledge to specify a pattern on the finer granularity. It seems worthwhile to identify patterns that have intuitive appeal, defining which variables have more influence on the aggregated variable, or, from the perspective of the coarser granularity, which variables are more likely to have caused the aggregated state. Specifically, we consider four situations: the *uniformity* assumption implies that we have no reason to believe that there is a particular structure at the precise granularity, the maximum entropy choice. A *discretised normal* distribution is useful to model the uncertainty around a particular observation point. The *decreasing* and *increasing* patterns convey the idea that some event is more likely to occur at the start or the end of the aggregated interval.

If a point in the coarser granularity X_a has corresponding points in the finer granularity $X_1, \ldots, X_i, \ldots, X_n$, the patterns have the following properties:

Uniform: $P(X_1) = P(X_i) = P(X_n)$

Normal: $P(X_i) = P(X_{n-i+1}) < P(X_{n/2})$

Decreasing: $P(X_1) \geq P(X_i) \geq P(X_n)$

Increasing: $P(X_1) \leq P(X_i) \leq P(X_n)$

Note that for the *normal*-pattern some care needs to be taken when aggregating an even number of variables. For the *increasing* and *decreasing* patterns one could also consider the strict cases, but especially for large n the current versions are more flexible.

5 Temporal indeterminacy in COPD

The model shown in Figure 3 is an example of how temporal aggregation can be used in the context of COPD-monitoring. The data has been gathered at various hospitals and general practices in the Netherlands, with the smartphone-based monitoring system as described in (van der Heijden et al., 2011). The model is an unrolled dynamic Bayesian network for four symptoms: dyspnea, cough, sputum production and activity capacity. Each variable indicates a daily measurement over the course of a week, and takes on binary values (False = normal, True = worse than normal). The daily symptom variables are aggregated in a variable representing the whole week by means of a generalised causal independence model (Srinivas, 1993). Instead of a binary outcome variable, the aggregate has values representing the number of symptom days (0 through 8). The intermediate variables model the inhibition, or noise, parameters. Finally, we are interested in the outcome variable *Exacerbation*, which provides an indication of a clinical significant event. Here again a causal independence model is utilised to aggregate the individual symptoms, by means of a counting function that classifies the number of symptoms into the categories [0-7],[8-15],[16-23],[24-32], that is, it is a deterministic function of the symptom aggregates.

The parameters of the model have been learned from the monitoring data of 8 COPD patients that participated in a pilot study. As before we assumed stationarity, and with a Dirichlet uninformative prior distribution with $\alpha = 1.1$ we estimated the probabilities of a symptom variable $X \in \{D, C, S, A\}$ as:

$$P(X_i = k \mid pa(X_i) = j) = \frac{\alpha_{ijk} + N_{ijk} - 1}{\alpha_{ij} + N_{ij} - |X_i|},$$

where N_{ijk} is the number of cases that variable i takes value k and the parent variables of i take on configuration j; α_{ijk} denotes the pseudo-counts; $|X_i|$ is the number of values that variable X_i can take on and $N_{ij} = \sum_k N_{ijk}$. The parameters of the intermediate variables were all equal with $P(XIn_i = 1 \mid X_i = 1) = 0.9$ and $P(XIn_i = 1 \mid X_i = 0) = 0.1$. All the aggregate variables have deterministic parameters as explained above.

In the pilot study, participants were asked to fill in the Clinical COPD Questionnaire (CCQ) as a control measure for the data gathered with the monitoring system. The questionnaire consists of 10 questions on a 7 point scale (lower scores are better); 4 questions are about dyspnea, 1 about cough, 1 about sputum production and 4 about activity; we omitted the 2 psychological questions about dyspnea for the current analysis; the version used asked patients to answer with respect to last week. For our current purposes this gives us a good example to look at aggregation, since we can compare the monitoring data from the preceding week with the results of the CCQ. Because not all data was available at the time of writing and due to missing data (patients not filling in the CCQ), we have data of 5 patients for comparison.

The procedure is as follows, we entered the monitoring data as evidence in the DBN (missing values are easily taken care of by leaving them as non-evidence variables) and noted the probabilities of the aggregate variables. For the CCQ scores we took the mean of the scores for the same aspect, rounded to the nearest integer. The results of the comparison are shown in Table 1, where the distribution over the aggregate variables is shown, which basically means the probability of a certain number of days of that particular symptom during the preceding week. Each value of an aggregate variable is mapped to the same CCQ score (0 to 6), except 7 which is mapped to CCQ score 6 (i.e. '(almost) always' is taken to mean 6 or 7 days).

Using the most likely value of the variables as a simple measure to assess agreement with the CCQ scores we see that we end up with the same value in 30% of the cases and are one value

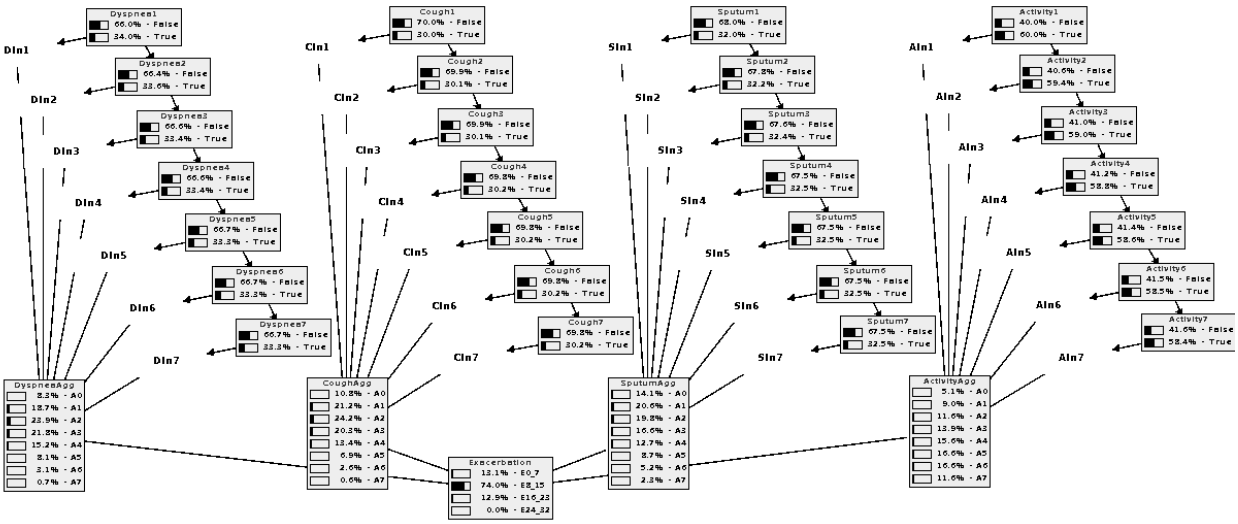

Figure 3: COPD-monitoring aggregation network.

removed in 45%. But given the currently limited amount of data it is hard to say whether this could be improved by a better aggregation model or whether there exists a discrepancy between the monitoring data and the CCQ scores. The sputum value of patient 3 favours the latter explanation, as the CCQ indicates sputum was produced almost never whereas the monitoring data provide evidence to the contrary; clearly an aggregation model cannot solve these inconsistencies. With respect to the outcome variable *Exacerbation*, visual inspection of Table 1 indicates that patients 1 and 2 are stable, patients 3 and 5 are in the course of an exacerbation and patient 4 is somewhere in between. This shows that the aggregation provides a useful and more easily interpreted summary of the monitoring data.

6 Related work

A well known representation of time is Allen's algebra (Allen, 1983). It provides a set of relations between time intervals and operations to reason over the intervals. The relations allow expressing information about the order of intervals and some different types of overlap, for example 'X starts or is during Y'. In some sense this allows indeterminacy, but in a qualitative way. Temporal indeterminacy and its relation to representations at different granularities is a topic studied in the context of temporal databases (Combi and Pozzi, 2001); but the problem is also relevant for planning and scheduling and information systems that deal with temporal data. For medical information systems there has been quite some work on temporal reasoning: on multigranular representations (Keravnou, 1999); and on temporal abstraction (Shahar, 1997). The latter deals with aggregating temporal data to meaningful higher level concepts, which requires taking care of temporal representation issues like granularities. Our probabilistic graphical model approach provides a different view on a part of these problems. Also related are irregular-time Bayesian networks (Ramati and Shahar, 2010), which generalise DBNs to time-slices with changing size. Bettini et al. (1998) studied granularities in temporal constraint satisfaction problems; and Combi et al. (2004) in a more general linear time logic context, both deal extensively with properties of multiple granularities but not with probabilistic representations.

7 Conclusion

Modelling temporal indeterminacy in probabilistic graphical models creates the opportunity to deal with different kinds of uncertainty

Table 1: Distributions (percentage) over the aggregate variables (D=dyspnea, C=cough, S=sputum, A=activity) for 5 patients. In bold the probability of the value that corresponds to the CCQ score.

	0	1	2	3	4	5	6	7
D	47.8	37.2	**12.4**	2.3	0.3	0	0	0
C	0.6	**11.0**	53.8	27.9	6.1	0.7	0	0
S	**4.7**	45.7	35.8	11.6	2.0	0.2	0	0
A	**47.8**	37.2	12.4	2.3	0.3	0	0	0
D	30.6	**37.8**	23.0	7.2	1.3	0.1	0	0
C	32.7	**37.8**	21.7	6.6	1.1	0.1	0	0
S	36.3	36.9	**19.8**	5.9	1.0	0.1	0	0
A	**31.0**	35.3	23.8	8.2	1.5	0.2	0	0
D	0	0.5	4.8	20.7	33.5	**28.5**	11.1	0.9
C	0	0	0.5	4.3	**17.5**	34.0	29.9	13.7
S	0	**0**	0.3	2.4	11.5	28.4	33.7	23.6
A	0	0	0.1	1.2	6.7	**20.8**	36.6	34.7
D	0	1.0	7.8	23.4	**32.9**	25.1	9.0	0.7
C	1.2	13.8	29.7	**32.0**	18.1	4.7	0.5	0
S	33.2	37.1	**21.2**	7.0	1.3	0.1	0	0
A	0	0.2	1.9	**9.6**	24.6	37.8	23.7	2.2
D	0	0	0.2	2.4	13.5	37.8	**41.8**	**4.2**
C	0	1.0	**9.7**	36.8	37.3	13.3	1.8	0.1
S	0	0	0.4	3.9	19.4	44.3	**29.2**	**2.7**
A	0	0	0	0.5	3.7	**15.5**	37.4	42.8

that arises in many practical situations and in chronic disease monitoring in particular. We think that our initial results on probabilistic multigranular models by means of causal independence form a good starting point for further research in this direction. Particularly, the consequences and applicability of the single or multiple event assumption warrant attention. Future work also includes further analysis in the context of COPD-monitoring.

References

J. Allen. 1983. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843.

C. Bettini, X. Wang, and S. Jajodia. 1998. A general framework for time granularity and its application to temporal reasoning. *Annals of Mathematics and Artificial Intelligence*, 22:29–58.

C. Combi and G. Pozzi. 2001. HMAP – A temporal data model managing intervals with different granularities and indeterminacy from natural language sentences. *The VLDB Journal*, 9(4):294–311.

C. Combi, M. Franceschet, and A. Peron. 2004. Representing and reasoning about temporal granularities. *Journal of Logic and Computation*, 14(1):51–77.

T. Dean and K. Kanazawa. 1989. A model for reasoning about persistence and causation. *Computational Intelligence*, 5(3):142–150.

S. Evers, M. Fokkinga, and P. Apers. 2008. Probabilistic processing of interval-valued sensor data. In *Proceedings of the 5th workshop on Data management for sensor networks.*

M. Grzegorczyk and D. Husmeier. 2011. Non-homogeneous dynamic Bayesian networks for continuous data. *Machine Learning*, 83(3):355–419.

E. Keravnou. 1999. A multidimensional and multigranular model of time for medical knowledge-based systems. *Journal of Intelligent Information Systems*, 13:73–120.

Kevin P. Murphy. 2002. *Dynamic Bayesian Networks: Representation, Inference and Learning.* Ph.D. thesis, University of California, Berkeley.

Judea Pearl. 1988. *Probabilistic reasoning in intelligent systems: networks of plausible inference.* Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

M. Ramati and Y. Shahar. 2010. Irregular-time Bayesian networks. In *UAI '10: Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence.*

J. Robinson and A. Hartemink. 2010. Learning non-stationary dynamic Bayesian networks. *Journal of Machine Learning Research*, 11:3647–3680.

Y. Shahar. 1997. A framework for knowledge-based temporal abstraction. *Artificial Intelligence*, 90(1-2):79–133.

S. Srinivas. 1993. A generalization of the noisy-or model. In *UAI '93: Proceedings of the 9th conference on Uncertainty in Artificial Intelligence.*

M. van der Heijden, B. Lijnse, P.J.F. Lucas, Y.F. Heijdra, and T.R.J. Schermer. 2011. Managing COPD exacerbations with telemedicine. In M. Peleg, N. Lavrač, and C. Combi, editors, *AIME '11: Proceedings of the 13th Conference on Artificial Intelligence in Medicine*, volume 6747 of *LNAI*, pages 169–178.