Artificial Intelligence In Medicine 95 (2019) 104–117

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed

A comparison between discrete and continuous time Bayesian networks in learning from clinical time series data with irregularity

Manxia Liu[a,1,*], Fabio Stella[d], Arjen Hommersom[a,c], Peter J.F. Lucas[a,b], Lonneke Boer[e], Erik Bischoff[e]

[a] Radboud University, ICIS, Nijmegen, The Netherlands
[b] Leiden University, LIACS, The Netherlands
[c] Open University, Heerlen, The Netherlands
[d] University of Milano-Bicocca, Italy
[e] Radboud University Nijmegen Medical Centre, The Netherlands

ARTICLE INFO

Keywords:
Dynamic Bayesian networks
Continuous-time Bayesian networks
Point evidence
Interval evidence
Irregular time-series data
COPD

ABSTRACT

Background: Recently, mobile devices, such as smartphones, have been introduced into healthcare research to substitute paper diaries as data-collection tools in the home environment. Such devices support collecting patient data at different time points over a long period, resulting in clinical time-series data with high temporal complexity, such as time irregularities. Analysis of such time series poses new challenges for machine-learning techniques. The clinical context for the research discussed in this paper is home monitoring in chronic obstructive pulmonary disease (COPD).

Objective: The goal of the present research is to find out which properties of temporal Bayesian network models allow to cope best with irregularly spaced multivariate clinical time-series data.

Methods: Two mainstream temporal Bayesian network models of multivariate clinical time series are studied: dynamic Bayesian networks, where the system is described as a snapshot at discrete time points, and continuous time Bayesian networks, where transitions between states are modeled in continuous time. Their capability of learning from clinical time series that vary in nature are extensively studied. In order to compare the two temporal Bayesian network types for regularly and irregularly spaced time-series data, three typical ways of observing time-series data were investigated: (1) regularly spaced in time with a fixed rate; (2) irregularly spaced and missing completely at random at discrete time points; (3) irregularly spaced and missing at random at discrete time points. In addition, similar experiments were carried out using real-world COPD patient data where observations are unevenly spaced.

Results: For regularly spaced time series, the dynamic Bayesian network models outperform the continuous time Bayesian networks. Similarly, if the data is missing completely at random, discrete-time models outperform continuous time models in most situations. For more realistic settings where data is not missing completely at random, the situation is more complicated. In simulation experiments, both models perform similarly if there is strong prior knowledge available about the missing data distribution. Otherwise, continuous time Bayesian networks perform better. In experiments with unevenly spaced real-world data, we surprisingly found that a dynamic Bayesian network where time is ignored performs similar to a continuous time Bayesian network.

Conclusion: The results confirm conventional wisdom that discrete-time Bayesian networks are appropriate when learning from regularly spaced clinical time series. Similarly, we found that time series where the missingness occurs completely at random, dynamic Bayesian networks are an appropriate choice. However, for complex clinical time-series data that motivated this research, the continuous-time models are at least competitive and sometimes better than their discrete-time counterparts. Furthermore, continuous-time models provide additional benefits of being able to provide more fine-grained predictions than discrete-time models, which will be of practical relevance in clinical applications.

* Corresponding author.
E-mail addresses: m.liu@cs.ru.nl (M. Liu), stella@disco.unimib.it (F. Stella), arjenh@cs.ru.nl (A. Hommersom), peterl@cs.ru.nl (P.J.F. Lucas), Lonneke.Boer@radboudumc.nl (L. Boer), Erik.Bischoff@radboudumc.nl (E. Bischoff).

[1] ML is supported by the China Scholarship Council, by a grant from project NanoSTIMA [NORTE-01-0145-FEDER-000016], which was financed by the North Portugal Regional Operational Programme [NORTE 2020], under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund [ERDF].

https://doi.org/10.1016/j.artmed.2018.10.002
Received 19 September 2017; Received in revised form 12 September 2018; Accepted 3 October 2018

1. Introduction

The aging of the population is pushing governments and health-care organizations towards improving health-care quality, yet within the boundaries of strict budgetary constraints. At the same time, many governments and health-care organizations are increasingly investing into the use of electronic health technology, often referred as *eHealth*, with the expectation that it will make health-care delivery cheaper, while offering greater control by patients (patient empowerment) [1,2]. The general trend in most parts of the world is that health-care costs are increasing, sometimes quite steeply, and eHealth is seen as a way to move part of the health care burden from expensive institutional organizations, such as hospitals, to the home environment, thus contributing to the reduction in health-care costs.

It is now becoming clear that eHealth is shifting the health-care field to an increasingly data-driven way of working, yielding substantial *quantities* of patient data. However, the data-driven paradigm also renders the *quality* of the collected data of paramount importance to build and deploy sufficiently accurate models that support both patients and doctors. A common means to collect data in many clinical studies are paper diary cards [3,4]. Patients are encouraged to fill out diary cards, thereby documenting the status of their symptoms in the form of their responses to a questionnaire. Major drawbacks associated with using paper diary cards, in general, are that the dates and times of paper diary entries are often missing, due to the patient's poor compliance [5]. Therefore, the quality of the data collected from paper diaries has its limitations.

Another drawback of using paper diaries is the lack of generalizability. The time points of observations collected by paper diaries can be viewed as regular random samples from the timeline with a certain rate, also known as *observation rate*, e.g., once every day. However, having a fixed observation rate restricts eHealth studies to shorter periods and a smaller scale. It is unrealistic to expect that many patients are willing to collect clinical data on a regular basis as part of a long-term study. A more realistic assumption is that the regularity of recording an observation by the patient will vary and may be affected by many factors, such as whether or not the patient feels ill. This implies that methods to handle different time-regularity patterns are greatly needed.

Besides differences in time regularity, time *irregularity* is another common phenomenon of clinical time series. In clinical trials, the patient's health status, in terms of physiological data, may be observed only at irregularly spaced points in time. In addition, it is very unlikely that different patients are observed at the same points in time. Most of the current literature is based on statistical analysis of periodic snapshots of physiological measurements with a fixed time interval, such as daily [6,7] or weekly [3]. In this research, we aim to learn accurate and useful models from irregularly spaced clinical time series using temporal Bayesian networks.

To provide a more concrete clinical context for this research, we pay attention to chronic obstructive pulmonary disease (COPD) as an application area. COPD is a progressive disease where a patient's deterioration manifests itself in worsening symptoms, known as an *exacerbation*. It is of clinical interest to predict whether and when an *exacerbation event* will occur for a given patient. However, an exacerbation can not be directly observed. It is defined either in terms of specific worsening symptoms for consecutive days or if there is evidence of a patient's hospital admission due to an exacerbation. Unfortunately, clinicians have so far not been able to agree on a clinical definition of an exacerbation [3,8].

Rather than focusing our research on automatically deciding on the presence or absence of an exacerbation, using multiple definitions, we aim at trying to understand the dynamic behavior of the symptoms of COPD. In principle, the probability of having an exacerbation in the future can be computed based on the presence of the relevant worsening symptoms in the past, for example by rephrasing an exacerbation as the disjunction of all possible combinations of the symptoms for a given definition. The advantage is that we do not have to bother about the lack of a definition of a COPD exacerbation at the learning stage. The prediction, however, can still incorporate the different definitions of an exacerbation without relearning models. In that sense, our research can be extended in several directions.

The main contribution of our work consists of two parts. One contribution lies in capturing COPD symptom dynamics, which we see as representative for many other diseases that are being monitored in the home environment. So far it is unknown which particular method best captures disease dynamics using data from home monitoring. The second contribution lies in the in-depth investigation of two temporal Bayesian network methods to model the dynamics: *dynamic Bayesian networks* (DBNs), where time is assumed to be discrete, and *continuous-time Bayesian networks* (CTBNs), where time is assumed to be continuous. We also believe that this study sheds some light on the practical requirements of using DBNs and CTBNs in general.

The performance of DBNs and CTBNs for modeling the dynamics of COPD symptoms is investigated given COPD time series in three forms:

- when observations are made regularly at time points but with different observation rates;
- when time points of observations are unevenly spaced over time as a consequence of two missing data mechanisms, i.e.,
  (1) the probability of having variables observed at a time point is independent from other time points where variables are observed or unobserved, also known as *missing completely at random* (MCAR);
  (2) the probability of having variables observed at a time point is dependent on other time points where variables are observed, also known as *missing at random* (MAR). More specifically, the values for variables in the system are missing at time $t + 1$ if the values at time $t + 1$ are identical to those at time t.

In the rest of the paper, we only focus on the situation where variables are either fully observed or completely missing at a given time point. We investigate the performance of DBNs and CTBNs to learn from regular and irregular COPD time series. Within CTBNs, we also study the impact of the *evidence type*, i.e., point and interval evidence, on the performance of CTBNs. In addition, we also give an analysis of the impact of hyper-parameters on the performance of CTBNs.

To the best of our knowledge, this is also the first work where hyperparameters in CTBNs are taken into consideration in the modeling process. Within DBNs, we study the performance of DBNs interpreting time series in three ways, namely, (1) viewing time series as a sequence; (2) imputing values at discrete time points with the *Last-Observation-Carried-Forward* (LOCF) method (See Section 4.4); (3) filling in missing values at discrete time points by *Expectation Maximization* (EM). Our final aim is to gather information about potential factors that practitioners of temporal Bayesian networks need to take into account to learn a model from unevenly spaced clinical multivariate time series.

The rest of the paper is organized as follows. In the following section, we devote ourselves to describing the related works about predicting COPD exacerbation using machine-learning techniques and the state of the art of continuous-time Bayesian networks. It is followed by the description of two COPD time series we used to conduct experiments. Then we provide some theoretical background of temporal Bayesian networks and evidence type, i.e., point and interval evidence in Section 4. The experimental setup is described in Section 5.1, the evaluation methods, and the implementations. Comprehensive results are given in Section 5.2, where we compare the performance of dynamic Bayesian networks and continuous-time Bayesian networks both for simulated time series and for a real-world time series. Finally, we discuss our work's contribution, limitation, and future work in Section 6.

2. Related work

2.1. The clinical setting: COPD symptomatology

The availability of a widely accepted definition of an exacerbation of COPD in the medical community would definitely help to facilitate public communication and designing guidelines. Unfortunately, as said above, such a definition is still not available [9,10,8]. There is some work in the literature that studies the diagnostic impact of various definitions of an exacerbation [6,3]. So far, clinicians use a variety of clinical features to describe the COPD-related health status of a patient [4,11]. Even for the most accepted definition of an exacerbation at this moment, the *Anthonisen criteria* (AC) [12], the required major and minor symptoms are not always available partly due to design of the clinical study (see [6]). In [13,14], an exacerbation is defined in terms of a patient's hospital admission, or a non-scheduled visit to the emergency unit or to the specialists because of respiratory symptoms, or self-treatment of the patient by antibiotics. Because of their limited medical knowledge, patients are prone to misuse of antibiotics, i.e., they use antibiotics for a viral infection or a bacterial non-pulmonary infection. Instead, work in [7] chooses the worsening of respiratory symptoms at two consecutive days as an indicator of an exacerbation. It is clearly easier to predict an exacerbation for the next day when an exacerbation is currently observed than when it has not been observed. However, it seems that the authors provide no clear distinction between these two situations.

In the context of COPD management, a telehealth system [15] has been described previously that supports decision making. However, its decision support is limited to rule-based detection of abnormal values and to simple trend analysis. In contrast, predicting a COPD exacerbation, i.e., when the patient's health condition gets worse, can help to support the patient and doctor by providing an opportunity for early intervention before it is too late. To this end, some work has focused on the development of classifiers, e.g., using K-nearest neighbors (K-NN) [13] and K-means clustering [14] to predict the onset of an exacerbation given the patient's signs and symptoms. Nevertheless, there is still a lack of an explicit description of the underlying dynamics of the clinical symptoms in these models. Capturing temporal dynamics of signs and symptoms is the main goal of [7], where time and uncertainty are also considered for the first time. Given a limited amount of temporal clinical data, the work chooses to use dynamic Bayesian networks to capture the dynamics of COPD symptoms. As a consequence, the approach suffers from the need of finding the finest time interval. Usually this is undesirable both from a modeling and inference perspective. Thus, the models described above are unable to capture the dynamics of symptoms [13,14], or they do not take time as a parameter [7].

We conclude that a data-driven temporal model capturing the COPD dynamics is of clinical interest. It would not suffer from the subjective nature of a definition of an exacerbation, and may yield much more valuable insight into the nature of COPD in comparison to what can be achieved by a classifier model.

2.2. Model development: temporal Bayesian networks

In the previous section, we have already mentioned the related work by Van der Heijden et al. [7], which uses dynamic Bayesian networks (DBNs) for the detection of exacerbations of COPD. Another way to model the dynamics of symptoms using Bayesian networks is offered by continuous time Bayesian networks, i.e., time is used as a continuous parameter. The states of the symptoms satisfy a multinomial distribution, whereas the time when a transition occurs, e.g., a symptom change from one state to another, is modeled as an exponentially distributed parameter. Early work has demonstrated the powerful expressiveness of CTBNs to model the dynamics of systems where variables are observed at time points that are unevenly spaced over time [16–18]. In the specific domain of medical applications, CTBNs have been used to diagnose cardiogenic heart failure and have been shown to anticipate its likely evolution [19,20]. They have also been used to construct gene networks [21] to generate hypotheses for biological experiments [22]. Nevertheless, the current clinical applications significantly suffer from the unavailability of temporal patient data. The quantitative component of the CTBN model in [19], i.e. the parameters, are so far mainly elicited on the basis of clinical expertise. In our work, however, CTBNs are both applied to clinical synthetic data and real-world data.

Like standard Bayesian networks, evidence in CTBNs is also associated with a probability to incorporate the uncertainty nature of observations [23]. In addition, evidence entails the amount of time that a variable stays in a state. While *point evidence* claims that a variable holds a value for an infinitely small amount of time $\Delta t \to 0$, *interval evidence* states that a variable in the system can hold on a certain state throughout an interval of time. The concept of interval evidence is firstly introduced in [24], where it is originally called negative evidence.

Another relevant extension of Bayesian networks are irregular-time Bayesian networks [25]; they aim to increase the expressiveness of the temporal dynamics to handle irregular time series, with variables having a continuous state space. In the present work, however, we focus on discrete state spaces, which are typical for CTBNs and DBNs. Acerbi *et al.* [21] attempt to study the difference in performance between DBNs and CTBNs in a specific problem in the realm of molecular biology, where gene expressions are unevenly distributed over time. In their work, the focus is on the reconstruction of a gene network using DBNs and CTBNs, where solely simulated gene data are used. In addition, it still remains unclear whether there is a difference in the performance of CTBNs by using point and interval. Our work, however, clarifies this difference in the practical use of CTBNs.

Algorithm 1. Generating regular time series D_{REG}.

Data: A time series $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ in the dataset D_L, where n is the number of observations in D_L;
observation rates $R = \{1, 2, 3, 4, 5, 6, 7, 14, 21, 28\}$
Result: D_{REG}
$D_{\text{REG}} = \varnothing$;
foreach $r \in R$ **do**
  $D_r = \varnothing$;
  foreach $i \in \{1, \ldots, r\}$ **do**
    create a time series S_i with observations $\mathbf{x}_i, \mathbf{x}_{i+r}, \ldots, \mathbf{x}_{i+mr}$, where $m = \max\{s \mid i + sr \leq n\}$;
    $D_r = D_r \cup \{S_i\}$;
  $D_{\text{REG}} = D_{\text{REG}} \cup \{D_r\}$;
return D_{REG};

2.3. Data: irregular longitudinal clinical data

The temporal representation of clinical data has been extensively investigated by researchers in Artificial Intelligence in Medicine for more than two decades [26–29]. Most of this research deals with the use of time in clinical reasoning, e.g., for treatment planning and decision support, which is not of immediate relevance for our research. However, the reason why there is so much research on temporal reasoning in medicine is due to the significance of time in medical decision-making. In the context of the current research, we are dealing with a special kind of clinical temporal data, i.e., data that are being recorded by patients at home.

In many longitudinal clinical trials, patients are followed over a period of time and are scheduled to be assessed at a prespecified visit time after being enrolled in the study. However, patients often selectively miss their visits or return at non-scheduled points in time. As a result, the times of measurement are irregular, yielding a highly

Table 1
The number of observations in the dataset D_L for thirteen patients.

Patient identifier												
1	2	3	4	5	6	7	8	9	10	11	12	13
111	209	173	145	257	189	145	155	147	203	152	198	137

Table 2
The time difference in days between consecutive observations for each of the thirteen patients in the database D_L.

	Patient identifier												
	1	2	3	4	5	6	7	8	9	10	11	12	13
Min	1	1	1	1	1	1	1	1	1	1	1	1	1
Mean	1	2	1	1	1	1	1	1	1	1	1	1	1
Max	22	130	1	6	3	2	14	1	25	9	2	8	25

imbalanced time series. Some medical examples of this phenomenon are given by studies on the incident rate of sexual maturation [30] and by homeless people with mental illness [31].

Advances in computer technology have turned mobile devices into efficient data collecting tools in many scientific disciplines. One advantage of a mobile, network-linked digital tool is that patients are less likely to miss the time window of taking a measurement. However, the novel tools also create new mechanisms for obtaining irregular time series. One characteristic of the new mechanism is that all measured variables are either fully observed or fully missing at a give time point. The apparent irregularity pattern can be, in part, due to patients being instructed to report their symptoms only when abnormal symptoms are detected.

3. Materials

In this section, the time-series datasets used in the research are described.

3.1. Synthetic datasets

Data of a COPD patient cohort in London, which we obtained from the research group of Wedzicha et al. [4], were employed to generate time series with variables observed at equally and unequally spaced time points. The symptoms and signs in the original dataset were recorded by the patients on a daily basis. The methodology of the data collection process was previously extensively discussed in [32]. The dataset, denoted as D_L (where '*L*' stands for 'London'), consists of time series of thirteen COPD patients; each of them had at least one exacerbation. The data contains a total of 2849 data entries with values for the variables dyspnea (*D*), sputum volume (*SV*) and purulence (*SC*), wheeze (*W*), cough (*C*), temperature (*Temp*), and oxygen saturation (*O*). The number of observations and time difference between consecutive observations for thirteen patients in the dataset D_L are presented in Tables 1 and 2, respectively.

To study the behavior of CTBNs and DBNs for time series $\{\mathbf{x}_t \mid t \in T\}$, where $\mathbf{x}_t$ denotes a vector of values for the variables (symptoms and signs of COPD) at time point *t*, we generated time series from the dataset D_L given an *observation rate r*: the time interval between two successive observations. Ten time series were generated according to Algorithm 1, collectively denoted as D_{REG}, with the observation rate ranging from 1–6 day(s) to 1–4 week(s). Concerning modeling these variables, an earlier attempt to capture their temporal interactions was made by Heijden et al. [7]. For illustrative purpose, a dataset D_{REG} is generated from the fragment given in Fig. 1a by Algorithm 1 where $r = 2$, consisting time series S_1 and S_2 given in Fig. 1b and c.

Algorithm 2. Generating irregular time series D_{MCAR}.

Data: A time series $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ in the dataset D_L, where n is the number of observations in D_L;
% of removed entries $P = \{5, 10, 30, 50, 60, 70, 80, 90, 95\}$
Result: D_{MCAR}
$D_{\text{MCAR}} = \varnothing$;
foreach $p \in P$ **do**
  Randomly sample a set of time index D from $\{1, 2, \ldots, n\}$, where $|D| = n * p\%$;
  $S = \{1, 2, \ldots, n\} \setminus D$;
  Select observations from D_L with time indices S, yielding time series D_S;
  $D_{\text{MCAR}} = D_{\text{MCAR}} \cup \{D_S\}$;
return D_{MCAR};

Algorithm 3. Generating irregular time series D_{MAR}.

Data: A time series $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ in the dataset D_L, where n is the number of observations in D_L;
Result: D_{MAR}
$D_{\text{MAR}} = \{\mathbf{x}_1\}$
foreach $i \in \{2, \ldots, n\}$ **do**
  if $\boldsymbol{x}_i \neq \boldsymbol{x}_{i-1}$ **then**
    $D_{\text{MAR}} = D_{\text{MAR}} \cup \{\mathbf{x}_i\}$;
return D_{MAR};

Similarly, we investigate the capability of CTBNs and DBNs to handle time *irregularities* by generating irregular time series. First, we generated time series $\{\mathbf{x}_t \mid t \in T\}$, where observations are made at discrete-time points that are completely randomly and irregularly sampled in time. Nine datasets, collectively denoted as D_{MCAR}, were generated by randomly removing entries with percentage of 5, 10, 30, 50, 60, 70, 80, 90, and 95 from the dataset D_L, as shown in Algorithm 2. An illustrative dataset D_{MCAR} as shown in Fig. 2a is generated from the fragment given in Fig. 1a with $p = 50$. Second, we generated one time

	Time (day(s))									
	1	2	3	4	5	6	7	8	9	10
D	0	0	0	0	1	1	1	0	0	0
SV	0	0	0	0	1	1	1	0	0	0
SC	0	0	0	0	1	1	1	0	0	0
W	0	0	0	0	1	1	1	0	0	0
C	0	0	0	0	1	1	1	0	0	0
Temp	0	0	0	0	0	0	0	0	0	0
O	0	0	1	0	1	1	1	0	0	0

(a)

	Time (day(s))				
	1	3	5	7	9
D	0	0	1	1	0
SV	0	0	1	1	0
SC	0	0	1	1	0
W	0	0	1	1	0
C	0	0	1	1	0
Temp	0	0	0	0	0
O	0	1	1	1	0

(b)

	Time (day(s))				
	2	4	6	8	10
D	0	0	1	0	0
SV	0	0	1	0	0
SC	0	0	1	0	0
W	0	0	1	0	0
C	0	0	1	0	0
Temp	0	0	0	0	0
O	0	0	1	0	0

(c)

Fig. 1. A fragment of dataset D_L over ten consecutive days (a); a dataset D_{REG} consists of time series S_1 (b) and S_2 (c) generated from the fragment given in (a) according to the Algorithm 1 where $r = 2$. For an explanation of the meaning of the mentioned variables: see text. The value '0' stands for normal and '1' for abnormal.

	Time (day(s))				
	1	3	5	6	8
D	0	0	1	1	0
SV	0	0	1	1	0
SC	0	0	1	1	0
W	0	0	1	1	0
C	0	0	1	1	0
Temp	0	0	0	0	0
O	0	1	1	1	0

(a)

	Time (day(s))				
	1	3	4	5	8
D	0	0	0	1	0
SV	0	0	0	1	0
SC	0	0	0	1	0
W	0	0	0	1	0
C	0	0	0	1	0
Temp	0	0	0	0	0
O	0	1	0	1	0

(b)

	Time (day(s))							
	1	3	4	5	6	7	9	10
D	0	0	0	1	1	1	0	0
SV	0	0	0	1	1	1	0	0
SC	0	0	0	1	1	1	0	0
W	0	0	0	1	1	1	0	0
C	0	0	0	1	1	1	0	0
Temp	0	0	0	0	0	0	0	0
O	0	0	1	0	1	1	0	0

(c)

Fig. 2. Illustrative examples of generated datasets from the fragment of D_L given in Fig. 1a according to Algorithm 2, Algorithm 3 and Algorithm 4 respectively: (a) D_{MCAR} where $p = 50$ and selected time indices $S = \{3, 4, 6, 7, 10\}$; (b) D_{MAR}; (c) D_{L2A} with duplicated entries at time 2 and 8 removed.

between 46 and 89 years (mean (sd): 69.5 (8.9)), from June 2015 to Feb 2017. All patients were recruited from those attending Radboud University Medical Centre based on their willingness to participate in a long-term study. The patients had at least 2 self-reported exacerbations in the previous 12 months of the time of recruitment and had no severe co-morbidity. Inclusion criterion were: post-bronchodilator FEV_1/FVC < 0.70. Diseases such as diabetes, kidney diseases, and smoking habits were also recorded.

The ACCESS dataset, denoted as D_A, consists of a total of 1138 data entries with the same variables as in D_L. The number of observations and time difference between two consecutive observations for individual patient in D_A are given in Tables 3 and 4 , respectively. As part of the training, the patients were instructed to daily self-report using the ACCESS system for the first two weeks. As a consequence, the observations at the beginning of the study are relatively regularly spaced over time with a time interval of a day. Later in the study, most patients were reluctant to comply to the registration of their health status on a daily basis over a period of many months. To deal with this problem, after the two-weeks initial registration, the patients were instructed to take the initiative to make a registration of symptoms and signs when they detected something abnormal in their symptomatology. However, not all patients followed exactly the instruction, in particular, three patients randomly registered their respiratory symptoms in the entire study period. As a consequence, the time intervals between two consecutive observations (mean (sd): 7.4 (21.5) days) varied considerably from patient to patient.

Table 3
The number of observations in the dataset D_A for forty patients. 'ID' stands for patient identifier and 'Nob' for the number of observations.

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Nob	24	13	21	47	11	17	14	13	11	15	43	217	20	27	17	42	47	7	7	10
ID	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Nob	17	22	40	23	20	18	21	4	105	25	29	25	18	19	22	31	22	20	18	16

Table 4
The time difference in days between consecutive observations for each of the forty patients in the dataset D_A. 'ID' stands for patient identifier.

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Min	1	1	2	4	1	1	1	1	1	1	4	1	1	1	1	1	1	3	1	2
Mean	5	8	7	7	1	10	1	9	1	1	9	2	1	5	13	9	7	16	7	18
Max	55	80	47	101	3	38	1	57	1	1	137	90	3	91	175	44	35	39	31	63
ID	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Min	1	1	2	4	1	1	1	1	1	1	4	1	1	1	1	1	1	1	1	1
Mean	6	9	9	13	15	14	15	4	3	6	12	6	5	7	1	1	9	5	6	9
Max	66	77	95	155	199	166	285	7	15	89	65	29	52	53	2	2	81	57	89	65

series where the missingness is dependent on the observations. We removed consecutive identical entries in the dataset D_L (see Algorithm 3), resulting in an irregular time series D_{MAR}. The corresponding D_{MAR} for the fragment described earlier in this section is given in Fig. 2b.

3.2. ACCESS dataset

A real clinical dataset, independent of the London dataset discussed in the previous section, was subsequently used to study the behavior of DBNs and CTBNs on real-world irregular time-series. The dataset was collected using the recently developed "*Adaptive Computerized COPD Exacerbation Self-management Support (ACCESS)*" [2] system by Radboud University. The data were collected over two years for 40 patients, aged

Algorithm 4. Generating irregular time series D_{L2A}.

Data: A time series $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ in the dataset D_L, where n is the number of observations in D_L and time series D_A;
Result: D_{L2A}
Calculate the percentage of missing entries in D_A, denoted as p;
$S = \varnothing$;
foreach $i \in \{1, \ldots, n\}$ **do**
  if $\boldsymbol{x}_i$ *indicates the symptoms are normal* **then**
    $S = S \cup \{\mathbf{x}_i\}$;
Random select a subset D_p from S with the number of entries $p * n$;
$D_{\text{L2A}} = D_{\text{L}} \setminus D_p$;
return D_{L2A};

[2] see https://clinicaltrials.gov/ct2/~show/NCT02553096

Given an irregular time series for a specific medical problem, adopting an appropriate temporal technique requires a better understanding of the cause of the time irregularity. For the irregular time series D_A from the ACCESS study, it is reasonable to assume that the patients did not encounter any worsening symptoms for the days when no values were filled out for the variables. This is in accordance with the instructions they received at the beginning of the study which stated that symptoms only had to be recorded when something abnormal occured. To have a better understanding of the behavior of CTBNs and DBNs to handle such time irregularity, we generated another irregular time series in accordance to time series D_A. More specifically, we removed the same amount of entries if symptoms were normal from D_L in comparison to time series D_A, resulting in a time series denoted as D_{L2A} ('London to ACCESS', see Algorithm 4). An illustrative D_{L2A} shown in Fig. 2c is generated from the fragment given in Fig. 1a where the duplicated entries at time 2 and 8 are removed.

Note that variables in all synthetic and real-time series, namely, D_L, D_{REG}, D_{MCAR}, D_{MAR}, D_{L2A}, and D_A, are either fully observed or fully missing at a given time point.

4. Methods

In this section, a brief introduction to the techniques used in the remainder of the paper is given. We cover the technical background of Bayesian networks, their temporal variants, i.e., dynamic Bayesian networks (DBNs) and continuous time Bayesian networks (CTBNs), the interpretation of unevenly spaced time series using the temporal models, and the choice of hyperparameters in CTBNs.

4.1. Bayesian networks

In this paper, upper-case letters, e.g., X, Y, denote random variables. We denote the values of a variable by lower-case letters, e.g. x or $\bar{x}$ are short for $X = true$ and $X = false$, respectively. Note that all random variables are assumed to have a finite number of possible values.

A Bayesian network is a probabilistic graphical model which represents a joint probability distribution over a set of random variables. A *Bayesian network* $\mathcal{B}$ is defined as a pair $\mathcal{B} = (G, P)$, where G is an acyclic directed graph with $G = (\mathbf{V}, \mathbf{E})$, where $\mathbf{V}$ is a set of *nodes*, and $\mathbf{E} \subseteq \mathbf{V} \times \mathbf{V}$ a set of *directed edges* or *arcs*. The *joint probability distribution* P over random variables corresponding 1-1 with the nodes of the graph G is defined by a product of conditional probabilities of each random variable X given its *immediate parents* $\pi(X)$ in G, formally:

$$P(\mathbf{V}) = \prod_{X \in \mathbf{V}} P(X|\pi(X))$$

In the following, we do not distinguish between nodes and variables, and give them the same name.

The standard interpretation of the graph of a Bayesian network is that when two of its disjoint subsets of nodes $\mathbf{U}, \mathbf{W} \subseteq \mathbf{V}$ are connected by a path (ignoring the direction of the edges) that contains nodes v from a third, disjoint set $\mathbf{Z} \subseteq \mathbf{V}$ that only are serial nodes ($\cdots \rightarrow v \rightarrow \cdots$ or $\cdots \longleftarrow v \longleftarrow \cdots$) or divergent nodes ($\cdots \longleftarrow v \rightarrow \cdots$), and none of the nodes v or its descendants in $\mathbf{Z}$ have two incoming arcs $\cdots \rightarrow v \longleftarrow \cdots$, the path is called *blocked*. If every path between node sets $\mathbf{U}$ and $\mathbf{W}$ is blocked by $\mathbf{Z}$, it is said that $\mathbf{U}$ and $\mathbf{W}$ are *d-separated* given $\mathbf{Z}$ [33]. D-separation implies that the two corresponding variable sets of $\mathbf{U}$ and $\mathbf{W}$ are conditionally independent given the variables corresponding to $\mathbf{Z}$. If sets of nodes $\mathbf{U}$ and $\mathbf{W}$ are *not* d-separated given $\mathbf{Z}$, they are called *d-connected*.

Example 1. Consider the Bayesian network $\mathcal{B}$ shown in Fig. 3, with nodes and variables $\mathbf{V} = \{W, D, C, O\}$ with meaning: W: wheeze, C: cough, D: dyspnea, and O: oxygen saturation. According to the standard interpretation of Bayesian networks, the graph tells us that W and D are independent (d-separated), but they become conditionally dependent (d-connected) when C is known. In addition, whereas C and O are d-connected given the empty set $\emptyset$, they become d-separated given D, because D has a divergent connection on the path between C and O.

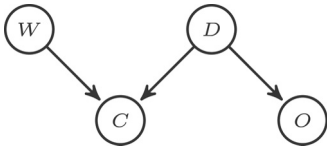


Fig. 3. A simplified COPD Bayesian network with four variables; W: Wheeze; C: Cough; D: Dyspnea; O: Oxygen saturation.

We obtain the joint probability distribution P over the variables $\mathbf{V}$ by multiplying the conditional distributions according to the structure of the graph:

$$P(\mathbf{V}) = P(C|D, W)P(W)P(O|D)P(D)$$

4.2. Temporal Bayesian networks

We use X_t to represent the instantiation of the variable X at time t, as time becomes an additional parameter in temporal Bayesian networks, i.e., dynamic (t is discrete) and continuous-time (t is continuous) Bayesian networks. These are two common methods for modeling clinical time series using probability theory. Different ways for interpreting time-series by these models are discussed, being an important practical consideration for both modeling and reasoning. In particular, in CTBNs, observations can be interpreted as point evidence or as interval evidence. In the latter case, an observation is assumed to persist within a time interval. In DBNs, we will distinguish three possible interpretations of an irregularly collected time series, either by interpreting a time series as a sequence only, or imputing the missing values or other values at discrete time points.

4.2.1. Dynamic Bayesian networks

A *dynamic Bayesian network* [34], DBN for short, is defined as a pair $(\mathcal{B}_0, \mathcal{B}_{\rightarrow})$ over variables $\mathbf{V}$, where $\mathcal{B}_0$ is a Bayesian network over variables $\mathbf{V}_0$ representing the *initial distribution* over states, and $\mathcal{B}_{\rightarrow}$ is defined as a *conditional distribution* for a *2-time-slice Bayesian network* (2-TBN) given by:

$$P(\mathbf{V}_{t+1}|\mathbf{V}_t) = \prod_{X \in \mathbf{V}} P(X_{t+1}|\pi(X_{t+1}))$$

for every time-point t.

For any desired time span $T \geq 0$, the joint distribution over $\mathbf{V}_{0:T}$ is defined by a product of the conditional probability distributions in the time 0 model and in the 2-TBN:

$$\mathrm{P}(\mathbf{V}_{0:T}) = \prod_{X \in \mathbf{V}} P_{\mathcal{B}_0}(X_0|\pi(X_0)) \prod_{X \in \mathbf{V}} \prod_{t \in \{0:T-1\}} P_{\mathcal{B}_{\rightarrow}}(X_{t+1}|\pi(X_{t+1}))$$

where X_{t+1} is the random variable X at time $t+1$. The parent set $\pi(X_{t+1})$ includes variables from the same or the previous time slice. As the equation shows, we can obtain a standard Bayesian network by 'unrolling' the DBN over T steps. This also clarifies that a DBN is a homogeneous discrete-time factorized Markov chain.

Example 2. Consider a dynamic Bayesian network over variables $\mathbf{V}$ with $\mathbf{V} = \{D, W, C\}$ (see Fig. 4), with the time 0 model and a transition model as shown in Fig. 4a and b, respectively. Then the joint distribution for the DBN over time $\mathbf{V}_{0:T}$ with the corresponding Bayesian network as shown in Fig. 4c is:

$$\begin{aligned} P(\mathbf{V}_{0:T}) &= P(D_0)P(W_0)P(C_0|D_0, W_0) \\ &\quad \prod_{t=0}^{T-1} P(D_{t+1}|D_t)P(W_{t+1}|W_t)P(C_{t+1}|D_{t+1}, W_{t+1}, C_t) \end{aligned}$$

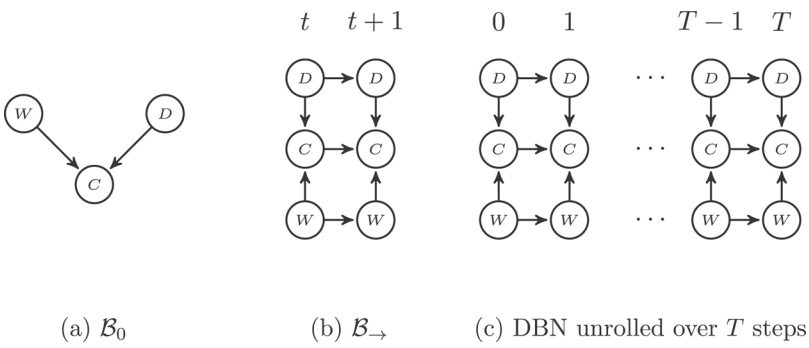


Fig. 4. A highly simplified DBN for COPD problem: (a) the time 0 model; (b) the 2-TBN; (3) the resulting unrolled DBN over T steps.

4.2.2. Continuous-time Bayesian networks

A *continuous-time Bayesian network*, CTBN for short, [35] represents a temporal stochastic model with variables as a homogeneous *continuous-time* factorized Markov chain, parameterized by intensity matrices. Formally, a CTBN over variables $\mathbf{V}$ consists of two components: the first component is an *initial distribution* specified as a Bayesian network; the second one is the continuous transition model specified as a graph and a *conditional intensity matrix* (CIM) $Q_{X|\pi(X)}$ for each variable $X \in \mathbf{V}$:

$$Q_{X|\pi(X)} = \begin{pmatrix} q_{11} & q_{12} & \cdots & q_{1n} \\ q_{21} & q_{22} & \cdots & q_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ q_{n1} & q_{n2} & \cdots & q_{\mathrm{nn}} \end{pmatrix}$$

The conditional intensity matrix describes the dependence of a variable X on the current values of its parents $\pi(X)$.

An entry q_{ij} in a conditional intensity matrix $Q_{X|\pi(X)}$ models the transition rate from state i to state j, $i \neq j$. Furthermore, the main diagonal q_{ii}, $i = 1, \ldots, n$, makes each row sum to zero, i.e., $q_{ii} = -\Sigma_{j \neq i} q_{ij}$. Intuitively, the reciprocal of the diagonal elements $-1/q_{ii}$ gives the expected time that variable X will remain in the state i, and once it transitions, it shifts from state i to state j with probability $-q_{ij}/q_{ii}$.

Two Bayesian *hyperparameters* are used in estimating the exponential parameter q_{ij} in CTBNs from data: (1) α: an imaginary count of the number of transitions in which a variable transits from one state to another, and (2) τ: an imaginary count of the amount time for a given state. Details can be found in [36].

Example 3. Suppose we want to model the dynamics of COPD symptom cough as the variable C with two possible states, i.e., c and $\bar{c}$, with an intensity matrix Q_C as follows:

$$Q_C = \begin{array}{c} \\ c \\ \bar{c} \end{array} \begin{array}{c} \begin{array}{cc} c & \bar{c} \end{array} \\ \begin{pmatrix} -1 & 1 \\ 2 & -2 \end{pmatrix} \end{array}$$

For example, the entry (2, 2) indicates that we expect on average a patient will cough in half hour if we take hours as the time unit.

A *full amalgamation* product operator * [35] is defined over a set of CIMs to compute the joint intensity matrix, resulting in a single continuous-time Markov process for the entire system.

Example 4. Suppose we have a CTBN with graph $W \rightarrow C$ with CIMs:

$$Q_W = \begin{pmatrix} -1 & 1 \\ 2 & -2 \end{pmatrix} \quad Q_{C|w} = \begin{pmatrix} -3 & 3 \\ 4 & -4 \end{pmatrix} \quad Q_{C|\bar{w}} = \begin{pmatrix} -5 & 5 \\ 6 & -6 \end{pmatrix}$$

We can compute the joint intensity matrix of variables W and C using the amalgamation operation on their CIMs, i.e., Q_W and $Q_{C|W=w}$ and $Q_{C|W=\bar{w}}$. The resulting joint intensity matrix is shown below:

$$Q_{WC} = \begin{array}{c} \\ wc \\ w\bar{c} \\ \bar{w}c \\ \bar{w}\bar{c} \end{array} \begin{array}{c} \begin{array}{cccc} wc & w\bar{c} & \bar{w}c & \bar{w}\bar{c} \end{array} \\ \begin{pmatrix} -4 & 3 & 1 & 0 \\ 4 & -5 & 0 & 1 \\ 2 & 0 & -7 & 5 \\ 0 & 2 & 6 & -8 \end{pmatrix} \end{array}$$

For a homogeneous Markov process over variables $\mathbf{V}$ with an intensity matrix $Q_{\mathbf{V}}$ and an initial distribution $P(\mathbf{V}_0)$, we can compute the distribution over the values of $\mathbf{V}$ at a particular time point or the joint distribution at different time points. The joint distribution at time point t is given by:

$$P(\mathbf{V}_t) = P(\mathbf{V}_0) \exp Q_{\mathbf{V}} t$$

where exp is the exponential operator that acts on a matrix.

4.3. Interpretation of unevenly-spaced time series

A *time series* is a set of observations $\{X_t = x \mid t \in T\}$, where the observation of variable X takes the form $X_t = x$, with x the value at time t. Unlike sequential data, the actual time stamp t is an important aspect of a time series. For example, we may have a time series recording whether or not a patient coughs at discrete time points 9:00, 11:00, 11:30, 12:00 in the morning.

In the statistical literature, there is some recent work on converting unevenly-spaced time series to equally-spaced data, or to directly analyze and manipulate the unevenly-spaced time series without an equally-spaced transformation (see e.g. [37,38]). In this paper, we will consider some natural choices when employing probabilistic graphical models for analyzing unevenly spaced clinical time series, in particular by transformation (e.g. by imputing the last observation) or by directly analyzing the irregular time series.

4.4. Interpretation of time series by DBNs

For discrete-time models, we consider three ways for interpreting a time series. An unevenly-spaced time series can be directly viewed as just a sequence, ignoring time differences between consecutive observations. Alternatively, the occurrence of a transition can be specified at discrete time points with a fixed time interval. In this case, two common methods are considered to handle irregular time series, as there will be missing data at some of the discrete time-points. These methods are (1) Expectation Maximization (EM) to learn from time series with missing data; (2) imputing the last observation (Last Observation Carried Forward, LOCF): values are filled in for variables at discrete time points where there is missing data.

Example 5. Consider a time series of observations for coughing at time points 9:00, 11:00, 11:30 and 12:00. The time series can be interpreted as a sequence by discarding the time stamps as shown in Fig. 5a, i.e., we assume we have complete data for this sequence of observations. If we select a half hour as the fixed time interval, the time series then contains a number of missing data points as shown in Fig. 5b. Using this

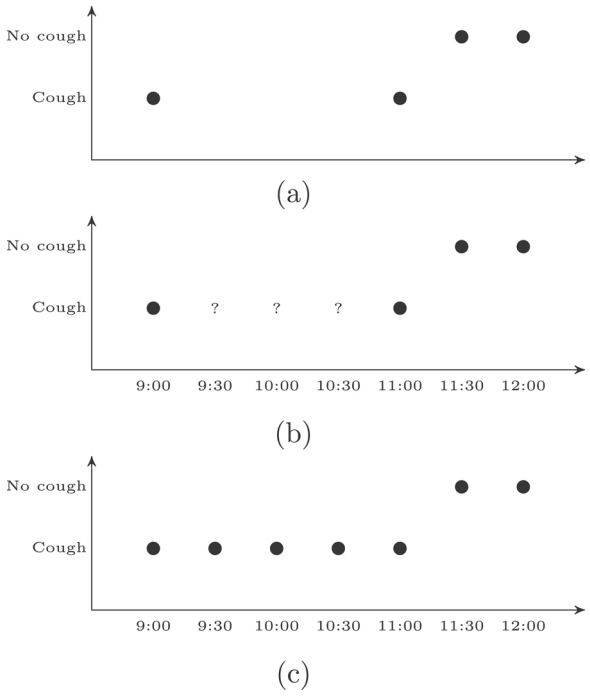


Fig. 5. Three DBN interpretations of clinical time series for the COPD symptom cough at time points 9:00, 11:00, 11:30 and 12:00. (a): just as a sequence with time stamps discarded; (b) missing data at some of the discrete time points; (c) with imputing the last observation at some of the discrete time points with missing data.

data, EM can be applied to learn from this time series directly. Imputing the last observation in the time series, which again leads to complete data (see Fig. 5c).

4.5. Interpretation of time series by CTBNs

In continuous-time models, there are two natural choices for interpreting time series. An observation is interpreted as *point evidence* when it only describes momentary behavior of a variable or system. For example, a patient is observed to have a cough at a certain point of time during the day. The counterpart of point evidence, *interval evidence*, on the other hand, states that a variable stays in a state for a given period of time. Formally, interval evidence on a right-open interval $[t_1, t_2)$, $t_1 < t_2$, can be denoted by $X_{[t_1,t_2)} = x$, asserting that X stays in state x during the given time interval $[t_1, t_2)$.

Example 6. Reconsider the time series from Example 5. The observations can be interpreted as point evidence (see Fig. 6a) or as interval evidence (see Fig. 6b). In the former, the interpretation states that the patient only coughs at time point 9:00, 11:00, 11:30 and 12:00. In the latter, however, the interpretation states that the patient keeps coughing in the time intervals [9 : 00, 9 : 30) and [11 : 00, 11 : 30) and does not stop in the time interval [11 : 30, 12 : 00).

4.6. Choice of hyperparameters in CTBNs

An issue that arises when learning CTBNs from clinical time series is choosing values for the associated hyperparameters. CTBN behavior is characterized by hyperparameters that are distinct from those of other temporal probabilistic models; together they gives a prior estimate of the time that a variable stays in a given state. When a time series is sufficiently large to capture all possible transitions, in particular for variables which change at a very slow rate, the hyperparameter τ has little impact on the learned models. However, this appears to be a rare

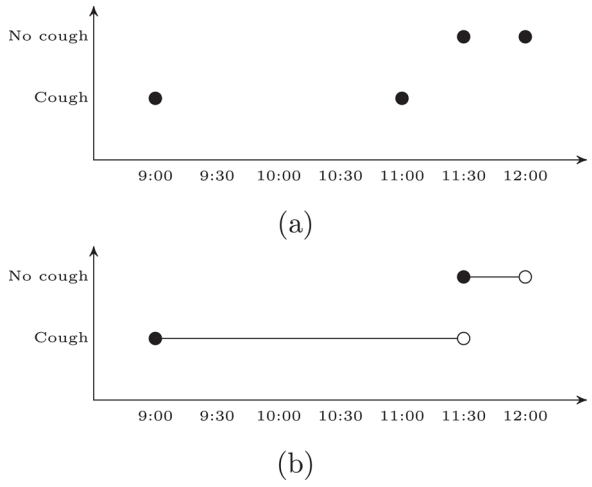


Fig. 6. Two CTBN interpretations of clinical time series for cough at time points 9:00, 11:00, 11:30 and 12:00. The observations are interpreted as point evidence in (a) and as interval evidence in (b).

situation for clinical time series. For chronic diseases, such as COPD in our case, a change of relevant symptoms can take so much time that it will never be observed in the limited clinical datasets that normally are available. This shortcoming of clinical time series is, in part, due to the difficulty of collecting data for many patients with *sufficient temporal detail* during a long period of time.

While some search algorithms have been devised to optimize hyperparameters (see e.g. [39,40]), they often come at the expense of high computational costs. Instead, a more cost-efficient approach can be used to constrain the state space of hyperparameters by the utilization of domain knowledge. In the present research, we use such an approach to select an appropriate hyperparameter for CTBNs to model the dynamics of COPD symptoms. Theoretically speaking, one symptom might have a hyperparameter configuration that differs from that of the other ones. In the paper, however, it is assumed that all COPD symptoms have the same configuration.

For COPD, we hypothesized that it is reasonable to assume that a change of symptoms expressed by the hyperparameter τ (See Section 4.2.2) takes less than 100 days but no less than 0.1 day. Therefore, we considered six possible values, i.e., $\tau \in \{0.1, 1, 10, 20, 50, 100\}$. Irrespective of the difference in value of the hyperparameter τ, the other hyperparameter α is set to 1. In Fig. 7, we present the result of this experiment where we subtract the log-likelihood from CTBNs with the hyperparameter having the value of 10 by those having one of the other five values. The results also show the impact of time granularity on the subtracted log-likehood. The results suggest that there is a relatively smaller difference in terms of log-likelihood for CTBNs using the values between 1 and 20. The results indicate that CTBNs in general achieve the best performance using the value of 20 for the hyperparameter τ in the given five choices. Nevertheless, they are still outperformed by those learned by using $\tau = 10$. Thus, in the following experiments in which CTBNs are compared to DBNs, the hyperparameter τ is fixed to 10, both when learning CTBNs from regular and irregular clinical time series.

5. Experiments

In this section we describe the experimental setup for learning and evaluating temporal probabilistic models, CTBNs and DBNs, of the evolution of COPD symptoms and signs.

5.1. Experimental settings

The purpose of the experiments is, firstly, to obtain insight into the

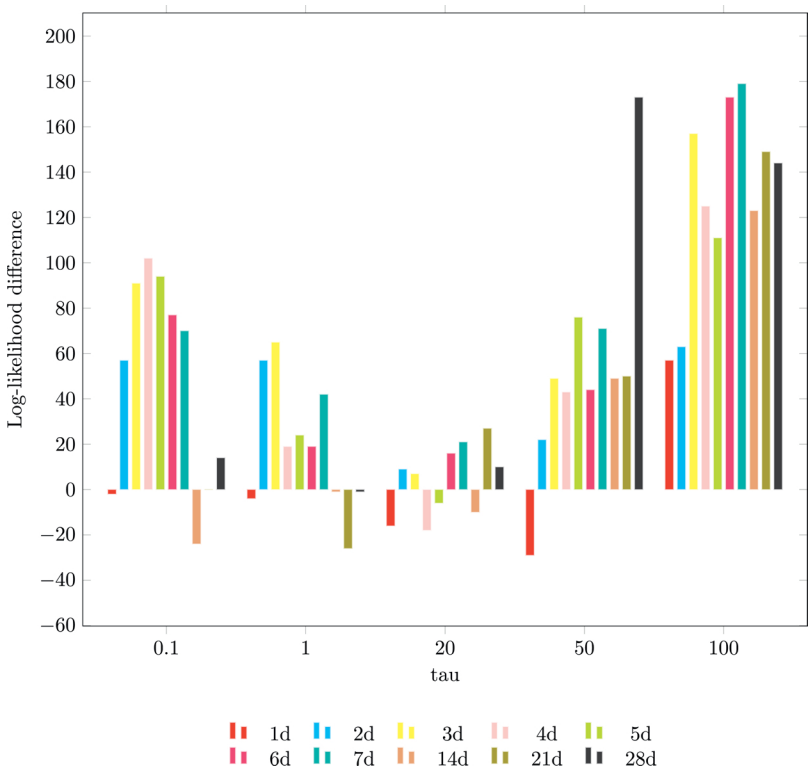


Fig. 7. Log-likelihood difference of CTBNs with interval evidence between a value of $\tau = 10$ and other values of $\tau \in \{0.1, 1, 20, 50, 100\}$ for various time granularities (1–7 days and 2–4 weeks). A higher positive value in log-likelihood indicates a performance for a CTBN with a value for $\tau \neq 10$ that is worse in comparison that that for $\tau = 10$.

behavior of CTBNs and DBNs for synthetic data, where observations are made at time points that are (1) equally spaced, and (2) unevenly spaced; secondly, the model types were also studied for their capability to handle time irregularity in a real-world situation. For this purpose, we generated several synthetic datasets, as described in Section 3.1, from an existing dataset that contained daily data, and consider one real-world time-series dataset, described in Section 3.2, for which patients entered data in an irregular way.

A number of software packages and tools were used in the experiments. With respect to learning DBNs, tools needed to learn from regular and irregular clinical time series were different. For the former, we first used *Banjo*[3] to learn a structure from regular time series D_{REG}, and subsequently used *bnlearn*[4] package in R to learn its parameters. For the latter, a choice to learn both structure and parameters using EM may seem reasonable. However, such an approach has been shown to offer limited capability to recover the underlying dependences between random variables [7,41]. In this paper, we chose to predefine a unique structure for the remaining irregular time series. We chose to learn the structure from the regular time series D_L, which ensures that differences in performance between models are not due to a poor structure obtained from using the EM algorithm. Given the learned structure, we used BNT tools[5] with the implementation of EM to learn parameters from irregular time series, i.e., D_{MCAR}, D_{MAR}, D_A, D_{L2A}. Furthermore, the R interface[6] for "*Continuous Time Bayesian Network Reasoning and Learning Engine (CTBN-RLE)*" was used to learn both structure and parameters for CTBNs.

To prevent overfitting, a K-fold cross-validation procedure was used where the data was randomly split into K partitions, with $K - 1$ partitions for learning and one partition for testing. The data for testing did not contribute to the learning of models. The number of folds K was set to 13 (the number of patients in the London data cohort) for time-series D_{REG}, D_{MCAR}, D_{MAR}, D_{L2A} and to 10 for the time series D_A, respectively. For each fold, the performance of DBNs and CTBNs was evaluated in terms of log-likelihood. However, the evaluation of CTBNs with point evidence was slightly different as the CTBN learning algorithm using point evidence led to a significant variation in the quality of the learned models. For this reason, we used an additional validation set to select a good-quality CTBN model when using point evidence.

5.2. Results

In this subsection, we will investigate and discuss the performance of DBNs and CTBNs when learned from regular and irregular time series. More specifically, their learning performance in terms of log-likelihood will be studied using a number of both synthetic and real-world time-series datasets. In the following, we will use the notation *iCTBN* and *pCTBN* for a learned CTBNs using interval and point evidence, respectively, and *sDBN*, *iDBN* and *emDBN* for DBNs learned from a sequence, from an imputed dataset, and using the EM algorithm, respectively.

5.2.1. Results for synthetic data

Synthetic time series are valuable for obtaining an understanding of how well CTBNs or DBNs can capture temporal knowledge when learned from a regular or irregular time series, as it is clearly impossible to obtain real-world time series that conform to any possible temporal pattern. In that sense, learning DBNs and CTBNs from synthetic data can act as a benchmark. The results were obtained by using the clinical time series previously described in Section 3.1.

Regular data with different observation rates. Given regular time series, we first study the impact of variations in observation rate on the performance of both DBNs and CTBNs. We learned *both the structures and parameters* from the data; the learned DBN and CTBN network structures, with the observation rate set to one day, are shown in Fig. 8. The results for the various models are shown in Fig. 9. A decrease in the number of observations from the time series D_{REG} accounts for an increase in the log-likelihood if the observation rate is higher than 14

[3] https://users.cs.duke.edu/~amink/software/banjo
[4] http://www.bnlearn.com
[5] https://www.cs.utah.edu/~tch/notes/matlab/bnt/docs/bnt_pre_sf.html
[6] http://rlair.cs.ucr.edu/ctbnrle/Rinterface

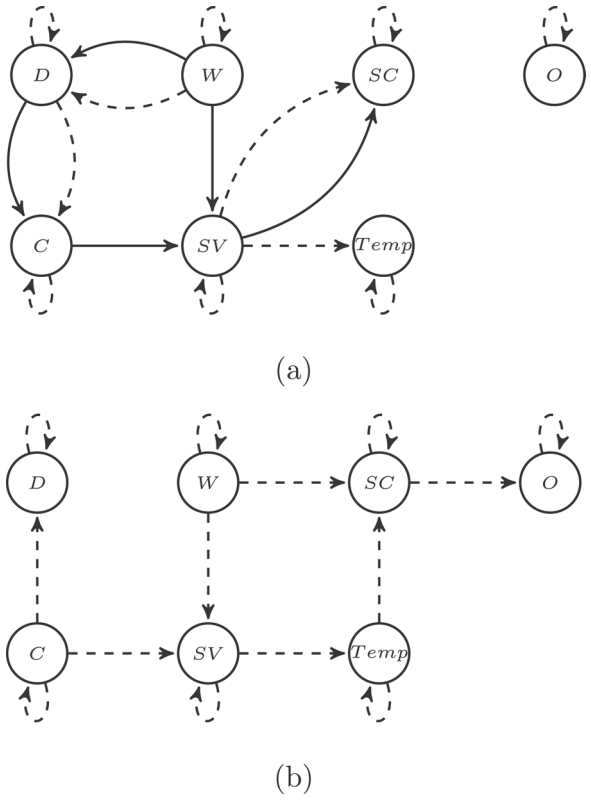


Fig. 8. Learned structures of a DBN and a CTBN from the London dataset with seven variables: O) Oxygen saturation; SC) sputum purulence; W) Wheeze; D) Dyspnea; C) Cough; SV) sputum volume; Temp) temperature. (a): A DBN; (b): A CTBN, where the value of hyperparameter τ is set to 20. Solid arcs indicate atemporal dependence and dashed arcs temporal dependence.

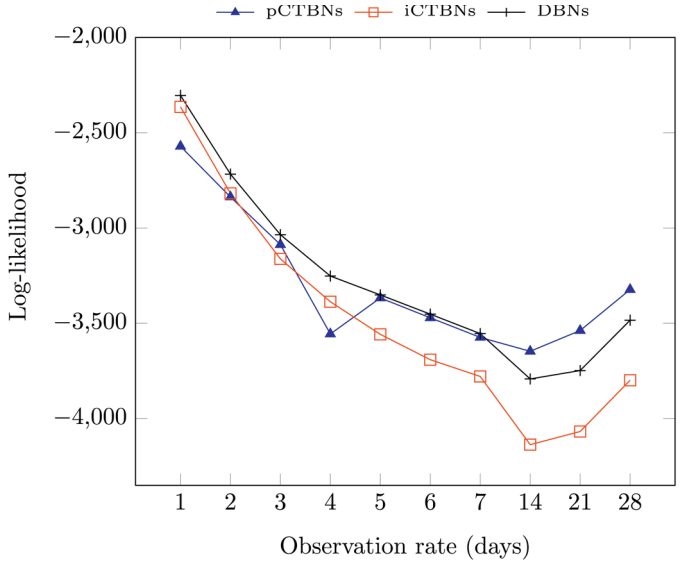
days. Overall, it is not surprising that the results show a declining performance for both DBNs and CTBNs when the observation rate increases. Irrespective of the observation rate, the results also show that DBNs have a higher performance than CTBNs in terms of log-likelihood, although the differences do not always reach statistical significance (see Fig. 9b).

Observations made completely at random. Given irregular time series with observations made completely at random in time, we study the impact of the data removal on the performance of DBNs and CTBNs. The log-likelihood of these models learned from a number of time series with a wide range of data removals is summarized in Fig. 10. Data removal is represented by a given percentage of removal of entries from the regular time series D_L.

Overall, the results suggest a positive correlation between the performance of DBNs and CTBNs with the data removal. When we take a closer look, the results indicate that it is significantly more difficult for emDBNs to capture the underlying dynamics when more than half of the entries are removed from D_L (see the p-value for emDBNs in Fig. 10b). In particular, when 60% of the entries are removed, the drop in the performance of emDBNs also indicates that it is most likely that the EM search does not reach a global optimum. In addition, DBNs using the imputation method have a significantly higher performance than CTBNs for the most of the removal percentages, irrespective of the evidence type.

When studying the behavior of DBNs alone, we also find significant differences of the performance of DBNs using the three distinct ways of interpreting time series. In particular, we find that the performance of DBNs using the EM algorithm declines at an increasing speed, while the performance of the other two DBNs declines relatively slowly. In DBNs, a consistent higher performance is also achieved by exploiting the imputation technique rather than simply discarding time stamps in irregular time series, although the difference is statistically insignificant based on the results on our synthetic datasets.

Now we switch our attention to the behavior of CTBNs. The results

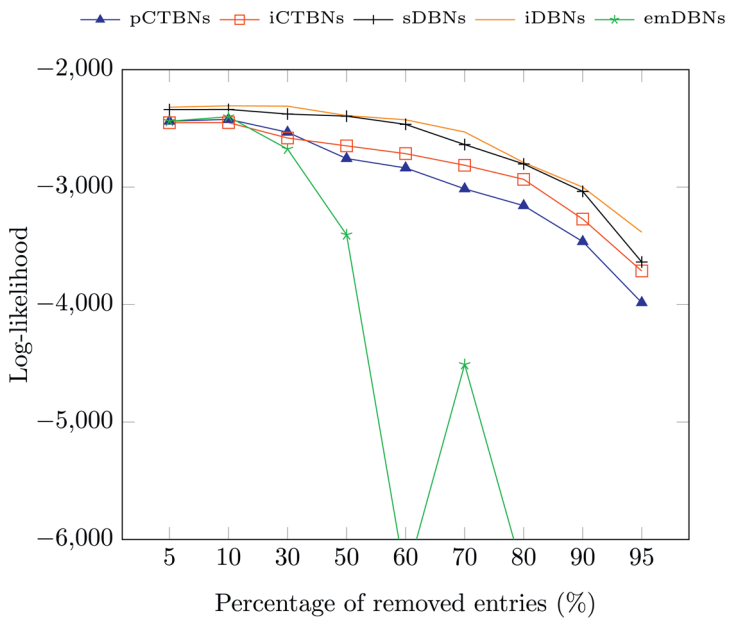


(a)

	Observation rates (day(s))									
	1	2	3	4	5	6	7	14	21	28
DBNs vs iCTBNs	0.31	0.24	0.25	0.26	0.18	0.08	0.14	0.09	0.10	0.10
DBNs vs pCTBNs	0.06	0.10	0.42	0.14	0.82	0.74	0.65	0.11	0.10	0.03
pCTBNs vs iCTBNs	0.24	0.86	0.52	0.56	0.08	0.12	0.15	0.01	0.00	0.01

(b)

Fig. 9. Performance of CTBNs and DBNs for regular time series D_{REG} where observations are made at different rates. (a) log-likelihood of CTBNs and DBNs; (b) p-values based on the paired t-test with bold text indicating a higher log-likelihood and with underline indicating a significant difference.



(a)

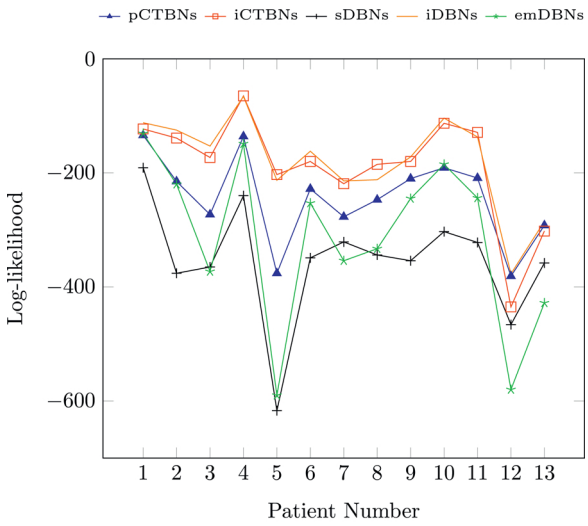
	Percentage of removed entries								
	5%	10%	30%	50%	60%	70%	80%	90%	95%
iDBNs vs emDBNs	0.05	0.12	0.02	0.01	0.03	0.00	0.00	0.00	0.00
sDBNs vs emDBNs	0.10	0.33	0.05	0.01	0.03	0.00	0.00	0.00	0.00
iCTBNs vs emDBNs	0.87	0.52	0.58	0.02	0.04	0.00	0.00	0.00	0.00
pCTBNs vs emDBNs	1.00	0.52	0.24	0.03	0.04	0.00	0.00	0.00	0.00
iDBNs vs sDBNs	0.26	0.18	0.31	0.97	0.78	0.63	0.97	0.91	0.59
sDBNs vs iCTBNs	0.13	0.22	0.20	0.20	0.29	0.54	0.69	0.58	0.90
sDBNs vs pCTBNs	0.01	0.12	0.15	0.01	0.01	0.01	0.12	0.00	0.22
iDBNs vs iCTBNs	0.05	0.06	0.02	0.02	0.01	0.02	0.15	0.03	0.09
iDBNs vs pCTBNs	0.01	0.02	0.02	0.02	0.02	0.05	0.20	0.13	0.09
iCTBNs vs pCTBNs	0.86	0.71	0.61	0.55	0.57	0.47	0.51	0.61	0.59

(b)

Fig. 10. Performance of CTBNs and DBNs for irregular time series D_{MCAR} where observations are made completely at random in time. (a) log-likelihood of CTBNs and DBNs; (b) p-values based on the paired t-test with bold text indicating a higher overall log-likelihood and with underline indicating a significant difference.

indicate that the learned CTBNs using interval evidence are better at capturing the underlying dynamics than these learned using point evidence, while the difference is not statistically significant.

Observations made at random. For irregular time series where observations are made at random in time, the performance of DBNs and CTBNs learned from the reduced irregular time series D_{MAR} is shown in Fig. 11. Unlike the previous two cases, the results from time series D_{MAR} are presented at the patient level. In general, the iDBNs learned by filling in missing values at discrete time points and CTBNs learned using interval evidence perform best. For the other two DBNs, however, the CTBNs with point evidence perform significantly better. When the focus is on the performance of iDBNs, the performance of these models appear to be rather vulnerable to the reduction of non-transition entries, i.e., when there is no transition between two consecutive entries. This is illustrated by their significantly worse performance on the time series D_{MAR} than on the time series D_{REG} with the observation rate set to one day. Moving to the performance at patient level, we also find that the performance of modeling individual patient using all the temporal models is similar, whereas the log-likelihood is much lower when evaluating on patient with index 5 and 12.

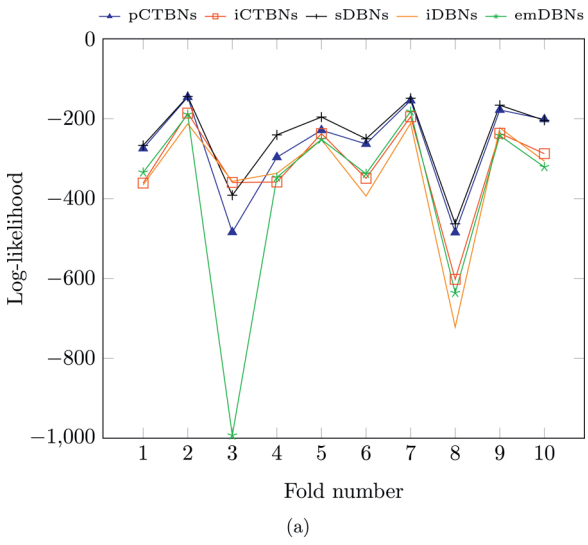


(a)

iDBNs vs emDBNs	1.6e-04
emDBNs vs sDBNs	9.6e-02
iCTBNs vs emDBNs	3.5e-04
pCTBNs vs emDBNs	5.1e-03
iDBNs vs sDBNs	1.2e-05
iCTBNs vs sDBNs	4.8e-05
pCTBNs vs sDBNs	4.8e-06
iDBNs vs iCTBNs	1.3e-01
iDBNs vs pCTBNs	2.4e-04
iCTBNs vs pCTBNs	3.5e-03

(b)

Fig. 11. Performance of CTBNs and DBNs for irregular time series D_{MAR} where observations are made at random in time. (a) log-likelihood of CTBNs and DBNs; (b) p-values based on paired t-test with bold text indicating a higher overall log-likelihood and with underline indicating a significant difference.



(a)

iDBNs vs emDBNs	5.1e-01
sDBNs vs emDBNs	3.1e-02
iCTBNs vs emDBNs	3.2e-01
pCTBNs vs emDBNs	3.7e-02
sDBNs vs iDBNs	4.0e-03
sDBNs vs iCTBNs	1.4e-03
sDBNs vs pCTBNs	3.1e-02
iDBNs vs iCTBNs	1.3e-01
pCTBNs vs iDBNs	4.7e-02
pCTBNs vs iCTBNs	5.9e-02

(b)

Fig. 12. Performance of CTBNs and DBNs for a real-world irregular time series D_A. (a) log-likelihood of CTBNs and DBNs; (b) p-values based on the paired t-test with bold text indicating a higher overall log-likelihood and with underline indicating a significant difference.

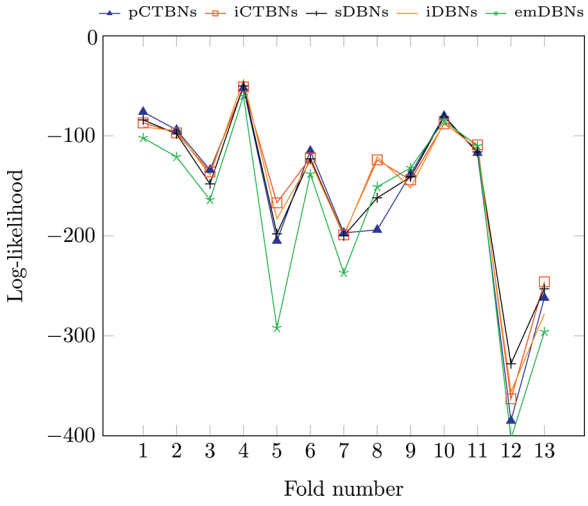


Fig. 13. Performance of DBNs and CTBNs on a synthetic time series D_{L2A}, derived from D_L based on D_A.

5.2.2. Results for the ACCESS dataset

Next, we investigate the performance of DBNs and CTBNs on the real-world irregular clinical time-series D_A; the results are summarized in Fig. 12. In general, the results show that sDBNs and pCTBNs have the best performance. By contrast, learning sDBNs performs poorly on the irregular time series D_{MAR}, as shown in Fig. 11. The difference can be explained by noting that the time series D_A differs from the time series D_{MAR} mainly in the number of non-transition entries. More specifically, the number of non-transitions in D_{MAR} is zero since it is generated by removing all the transitions from the time series D_L.

To further validate that the difference in the number of non-transitions may be the explanation for the better performance of sDBNS on the time series D_{MAR}, we considered its performance on the irregular time series D_{L2A}. Indeed, we do see an improved performance of sDBNs with an increasing number of non-transition entries in the time series D_{L2A} (see the increased log-likelihood of sDBNs in Fig. 13).

Therefore, the results from the two irregular time series D_{L2A} and D_A suggest that treating irregular time series as a sequence by leaving out time stamps may be sufficient to capture the transition probabilities of COPD symptoms. However, the pCTBNs provide a competitive and attractive alternative to these simple DBN models.

5.3. Discussion

Regular data with different observation rates. Given a regular time series, the performance of DBNs and CTBNs deteriorates when the number of transitions in the time series D_{REG} decreases. The time series D_{REG} were generated from the time series D_L by removing all transitions that occur within the amount time of a given observation rate. With fewer transitions, the task of learning DBNs becomes more difficult, and restoring the underlying transition probabilities degradates. In addition, it negatively affects the performance of CTBNs using interval evidence, as long time intervals can overestimate the length that a symptom stays in a particular state.

Conversely, increasing the observation rate can be expected to have a less negative impact on learning CTBNs with point evidence, thanks to

the presence of intact information at discrete time points in the time series D_{REG}. This is confirmed by the results as shown in Fig. 9b: the underlined numbers indicate that CTBNs with point evidence can achieve a significantly higher performance than interval evidence given a sufficiently high observation rate.

The performance difference between DBNs and CTBNs may be due to the ability of DBNs to represent atemporal dependence relative to a given observation rate. In particular, an atemporal dependence can be exploited by DBNs to represent correlations between symptoms that are not evolving over a long time span. For example, in the context of COPD, symptoms may influence each other within days. For a regular time series with larger time granularity (e.g. a week), the correlations within a week can not be captured by temporal dependences. Given that CTBNs do not contain atemporal dependences, we speculate that CTBNs using interval evidence may suffer most from this limitation in expressiveness for regular time series with a large time granularity.

Observations made completely at random. The iDBNs gain their advantage over sDBNs and emDBNs by imputing correctly most values of the symptoms at discrete time points. Discarding time stamps, sDBNs may suffer more from removing entries from the time series D_L, which results in a different probability distribution in the time series D_{MCAR}.

The lower performance of CTBNs with point evidence compared to interval evidence indicates that they may be more vulnerable to information loss than their counterpart with interval evidence. More specifically, the decrease in the number of observations in a time series can significantly increase the learning search space for CTBNs with point evidence. In contrast, this decrease has a less negative impact on CTBNs with interval evidence. This may attributed to a better approximation of the duration of the presence and absence of a symptom using interval evidence.

Observations made at random. The much lower log-likehood for patient 5 and 12 is partly because there are significant more observations for these patients in comparison to the others. More observations implies a higher log-likelihood. Moreover, patient 12 also differs from the other patients by experiencing a fever and having more often an abnormal level of oxygen saturation. Such a variance may not be captured by a model that was also learned from data of the other patients, leading to a model that fitting less well. Combined with the small study patient population, the results also indicate that there is a need for the development of personalized clinical models, which can be flexibly adapted to each individual patient's behavior.

6. Conclusion

The main motivation for this research was the wish to provide an alternative to the medically common way of managing symptom worsening of a chronic disease in terms of the *static* occurrence of particular symptoms. As an example we used COPD, a very worldwide common chronic disease that is increasingly managed in the home environment through eHealth technology [6,2]. Methods to capture symptom *dynamics* with their associated uncertainty were seen as a way to make progress here. However, as we discussed, there are significant challenges with respect to learning from clinical data that is collected in a home environment. In order to gain more insight into the most appropriate modeling technique for clinical time series with observations that are unevenly spaced over time, we have studied the performance of dynamic Bayesian networks and continuous-time Bayesian networks on both synthetic and real-world datasets with the final goal to build a predictive model for COPD. For simple cases, such as regularly-spaced time series, discrete-time Bayesian-network models are appropriate, as one might expect. However, for complex clinical time-series data that motivated this research, the continuous-time models are at least competitive and sometimes better than their discrete-time counterparts. Given that CTBNs also provide more fine-grained predictions over time, they are an attractive alternative to discrete time probabilistic models.

Besides this general conclusion, we also studied the usage of DBNs and CTBNs in more detail. Firstly, we have studied different manners to interpret data with temporal Bayesian networks. We showed that the evidence type has a significant impact on the results in different situations. Secondly, we have considered the impact of the hyperparameters (i.e. imaginary counts) in CTBNs. To the best of our knowledge, we are the first to explore the impact of hyperparameters on the learned models, which again has a significant impact on the results that one can obtain with CTBNs.

Our work also has some limitations. First, we only investigate one possible missingness case where variables are either all missing or all observed at each time point. If some random variables are missing and some others are observed at particular points in time, then this would create an additional complexity for learning and reasoning with temporal Bayesian networks.

Besides other types of missing data, we believe that there are a number of other interesting questions to investigate in the future. First of all, choosing the length of the interval when using interval evidence could have a significant impact on the results, as a larger interval provides a stronger bias. While in this paper we have chosen a fairly arbitrary interval determined by the time granularity of the data, it might be more sensible to assist the learning process with domain knowledge about the maximum length of the interval, e.g. a week. Second, it is an intriguing but challenging task to provide theoretical evidence to support the superiority of one method over the other under certain missingness conditions. Third, inspired by data with different time granularities, there is still some room to study the difference between short-term and long-term dynamics for COPD symptoms. Models with multiple time scales, such as provided by continuous-time models, can provide information about the evolution of symptoms from a different perspective. For the final purpose of COPD prediction, this is one of the attractive aspects of CTBNs compared to discrete-time models with a single time granularity.

To conclude, in this paper, the capability of DBNs and CTBNs to handle time series data was studied with a specific medical problem in mind, i.e., COPD patient management. However, the principles concerning their use in practice also apply to other real-world problems where multivariate time series are involved. For example, having an appropriate interpretation of time series is crucial for choosing between DBNs and CTBNs for a given problem. In particular for irregular time-series data it will be advantageous to use CTBNs as learning methods. In the medical domain, it will be of clinical interest to have a more fine-grained prediction for the evolution of a disease. In that case, CTBNs are a powerful modeling tool to deal with such predictions. In addition, it is better to use domain knowledge to choose between point and interval evidence where CTBNs are employed. The domain knowledge regarding the time that a variable stays in a state at most can assist the model learning process.

References

[1] Grando MA, Rozenblum R, Bates D, editors. Information technology for patient empowerment in healthcare. Berlin/Boston/Munich: Walter de Gruyter; 2015.
[2] Velikova M, Lucas PJF, van der Heijden M. Intelligent disease self-management with mobile technology. IEEE Comput 2015;48:32–9.
[3] Bischoff EW, Boer LM, Molema J, Akkermans R, van Weel C, Vercoulen JH, Schermer TR. Validity of an automated telephonic system to assess COPD exacerbation rates. Eur Respir J 2012;39(5):1090–6.
[4] Hurst JR, Donaldson GC, Quint JK, Goldring JJ, Patel AR, Wedzicha JA. Domiciliary pulse-oximetry at exacerbation of chronic obstructive pulmonary disease: prospective pilot study. BMC Pulmonary Med 2010;10(1):1–11.
[5] Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient compliance with paper and electronic diaries. Controlled Clin Trials 2003;24(2):182–99.
[6] van der Heijden M, Lucas PJF, Lijnse B, Heijdra YF, Schermer TRJ. An autonomous mobile system for the management of COPD. J Biomed Inform 2013 Jun;46:458–69.
[7] van der Heijden M, Velikova M, Lucas PJF. Learning Bayesian networks for clinical time series analysis. J Biomed Inform 2014;48:94–105.
[8] Effing TW, Kerstjens HA, Monninkhof EM, van der Valk PD, Wouters EF, Postma DS, Zielhuis GA, van der Palen J. Definitions of exacerbations: does it really matter in clinical trials on copd? CHEST J 2009;136(3):918–23.
[9] Sapey E, Stockley RA. COPD exacerbations. 2: aetiology. Thorax 2006

Mar;61:250–8.
[10] Wedzicha JA, Seemungal TA. Copd exacerbations: defining their cause and prevention. Lancet 2007;370:786–96. 2016/08/15.
[11] Sund ZM, Powell T, Greenwood R, Jarad NA. Remote daily real-time monitoring in patients with copd -a feasibility study using a novel device. Respir Med 2009 Sep;103:1320–8.
[12] Anthonisen N, Manfreda J, Warren C, Hershfield E, Harding G, Nelson N. Antibiotic therapy in exacerbations of chronic obstructive pulmonary disease. Ann Inter Med 1987;106(2):196–204.
[13] Jensen MH, Cichosz SL, Dinesen B, Hejlesen OK. Moving prediction of exacerbation in chronic obstructive pulmonary disease for patients in telecare. J Telemed Telecare 2012 Mar;18:99–103.
[14] Sanchez-Morillo D, Fernandez-Granero MA, Jiménez AL. Detecting COPD exacerbations early using daily telemonitoring of symptoms and k-means clustering: a pilot study. Med Biol Eng Comput 2015;53(5):441–51.
[15] Basilakis J, Lovell NH, Redmond SJ, Celler BG. Design of a decision-support architecture for management of remotely monitored patients. IEEE Trans Inform Technol Biomed 2010 Sept;14:1216–26.
[16] Boudali H, Dugan JB. A continuous-time Bayesian network reliability modeling, and analysis framework. IEEE Trans Reliability 2006;55(1):86–97.
[17] Xu J, Shelton CR. Continuous time Bayesian networks for host level network intrusion detection. in Machine Learning and Knowledge Discovery in Databases (ECML/PKDD), vol. 5212 of Lecture Notes in Artificial Intelligence 2008:613–27.
[18] Xu J, Shelton CR. Intrusion detection using continuous time Bayesian networks. J Artif Intell Res 2010;39:745–74.
[19] Gatti E, Luciani D, Stella F. A continuous time Bayesian network model for cardiogenic heart failure. Flexible Services Manuf J 2012;24(4):496–515.
[20] Liu M, Hommersom A, van der Heijden M, Lucas PJF. Hybrid time Bayesian networks. Int J Approximate Reasoning 2017;80:460–74.
[21] Acerbi E, Zelante T, Narang V, Stella F. Gene network inference using continuous time Bayesian networks: a comparative study and application to th17 cell differentiation. BMC Bioinform 2014;15(1):1–27.
[22] Acerbi E, Viganò E, Poidinger M, Mortellaro A, Zelante T, Stella F. Continuous time Bayesian networks identify Prdm1 as a negative regulator of th17 cell differentiation in humans. Sci Rep 2016;6.
[23] Sturlaugson L, Sheppard JW. Uncertain and negative evidence in continuous time bayesian networks. Int J Approximate Reason 2016;70:99–122.
[24] Gopalratnam K, Kautz H, Weld DS. Extending continuous time Bayesian networks. in Proceedings of the national conference on artificial intelligence, vol. 20. Menlo Park, CA; Cambridge, MA; London: AAAI Press; MIT Press; 1999. p. 981. 2005.
[25] Ramati M, Shahar Y. Irregular-time Bayesian networks. UAI'10: Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence 2010:484–91.
[26] Keravnou ET. Medical temporal reasoning (editorial). Artif Intell Med 1991;3(6):289–90.
[27] Keravnou ET. Medical temporal reasoning (editorial). Artif Intell Med 1996;8(3):187–91.
[28] Shahar Y, Combi C. Temporal reasoning and temporal data maintenance in medicine: issues and challenges. time-oriented systems in medicine. Comput Biol Med 1997;27(5):353–68.
[29] Adlassnig K-P, Combi C, Das AK, Keravnou ET, Pozz G. Temporal representation and reasoning in medicine: Research directions and challenges. Artif Intell Med 2006;38:101–13.
[30] Espeland MA, Platt OS, Gallagher D. Joint estimation of incidence and diagnostic error rates from irregular longitudinal data. J Am Stat Assoc 1989;84(408):972–9.
[31] Lin H, Scharfstein DO, Rosenheck RA. Analysis of longitudinal data with irregular, outcome-dependent follow-up. J Royal Stat Soc: Series B (Stat Methodol) 2004;66(3):791–813.
[32] Seemungal TA, Donaldson GC, Bhowmik A, Jeffries DJ, Wedzicha JA. Time course and recovery of exacerbations in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2000;161(5):1608–13.
[33] Pearl J. Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann; 1988.
[34] Murphy KP. Dynamic Bayesian Networks: Representation, Inference and Learning 2002. PhD thesis.
[35] Nodelman U, Shelton CR, Koller D. Continuous time Bayesian networks. Proceedings of the Eighteenth conference on Uncertainty in artificial intelligence 2002:378–87.
[36] Nodelman U, Shelton CR, Kollerf D. Learning continuous time Bayesian networks. Proceedings of the nineteenth international conference on uncertainty in artificial intelligence 2003:451–8.
[37] A. Eckner, "A framework for the analysis of unevenly spaced time series data," Preprint. Available at: http://www. eckner. com/papers/unevenly_spaced_time_series_analysis, 2012.
[38] Rehfeld K, Marwan N, Heitzig J, Kurths J. Comparison of correlation analysis techniques for irregularly sampled time series. Nonlinear Process Geophys 2011;18(3):389–404.
[39] Maclaurin D, Duvenaud D, Adams R. Gradient-based hyperparameter optimization through reversible learning. International conference on machine learning 2015:2113–22.
[40] Thornton C, Hutter F, Hoos HH, Leyton-Brown K. Auto-WEKA: Combined selection and hyperparameter optimization of classification algorithms. Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining 2013:847–55. ACM.
[41] Zhang Y, Deng Z, Jiang H, Jia P. Dynamic Bayesian network (DBN) with structure expectation maximization (SEM) for modeling of gene network from time series gene expression data. in BIOCOMP 2006:41–7.