Unraveling temporal processes using probabilistic graphical models
de Paula Bueno, M.L.

Cover Page

The handle http://hdl.handle.net/1887/85168 holds various files of this Leiden University dissertation.

Author: De Paula Bueno, M.L.
Title: Unraveling temporal processes using probabilistic graphical models
Issue Date: 2020-02-11

1

INTRODUCTION

1.1 THE RELEVANCE OF TEMPORAL INFORMATION

The comprehension of real-world phenomena is often challenging, as their characterization might depend on some notion of time. A resulting lack of insight may stem from the fact that a single snapshot of a temporal process reveals only a part of its behavior, which may be insufficient for its complete understanding. This is the case, for example, when one contrasts a single instance of the observation of symptoms of a patient with a chronic disease against the longitudinal view of multiple instances of observations of symptoms: whereas the latter will offer a temporal view of the underlying disease process, the former will not shed any light upon disease dynamics.

In many everyday tasks, such as walking, cooking, sleeping and so on a role of temporal information can be identified. In professional fields, such as for example in psychiatry, the efficacy of pharmacological interventions in mental disorders only can be properly studied when the research is supported by collecting temporal data [3]. Such temporal information will tell for example how long it takes before a treatment becomes effective and how long and how often a patient should take a particular drug. Also in many other diseases, in particular those with a chronic duration, temporal information is of paramount importance to gain insight into speed of progress or recovery.

Thus, given the importance of information about time in everyday and professional life, when one wishes to mathematically *model* processes of human artifacts, for example cyberphysical systems, or processes in the life sciences, usually time will be one of the parameters that need to be taken into account. It is not surprising that predictions about the future are often more accurate when taking into account the history than when not relying on such information [63]. Of course, reasoning with time not only is concerned with predicting the future given the past, but can go in any other direction: from the present going back in time to understand the past, or from assumptions in the future going back in time to understand which past conditions are needed to make a particular future feasible. Whether these kinds of temporal reasoning are possible is determined by the nature and capabilities of the mathematical models and reasoning methods employed.

In this thesis, temporal processes are modeled as stochastic processes. Such processes typically involve one or more random variables that can be repeatedly observed. More importantly to their characterization, however, is that past observations have influence on future observations, which assigns to a temporal process a sequential nature. In other words, it is not just a matter of merely observing variables at different moments, or making repeated measurements of variables. On the other hand, by considering the sequential nature of such processes, additional challenges are introduced due to the increased modeling complexities that come along.

1.2 PROBABILISTIC GRAPHICAL MODELS

An innate property of temporal processes is change, which renders them a stochastic nature. This makes probability theory a suitable tool for modeling such processes. Probabilistic models naturally take into account uncertainty, and can be used for multiple purposes: to predict the behavior of process variables in the future, discover associations between variables (e.g. which variables have more or less influence on a certain variable), and to pinpoint causes that could explain abnormal behavior.

Deriving models from data is a reality nowadays. This is because not only data storage technology has advanced (e.g. hardware capacity), but also more data is currently being generated, by means of sensor devices, content posted on the Internet, hospitals, health care services, etc. Obtaining statistical models from data which are expressive enough and can provide answers in reasonable running time is, however, not trivial. One major reason is that the process variables might interact in a very large number of ways. Without prior knowledge on the problem at hand, there is typically no obvious way as to how to reduce the space of models that might be of interest. In the past, researchers often relied on overly simplistic models (see e.g. [70]) to make model building feasible.

One solution to the parsimony problem faced by researchers is found with the adoption of probabilistic graphical models (PGMs, for short) [104, 136]. PGMs combine probabilities with graph theory for providing a graphical representation of probability distributions. With the representation of PGMs it becomes much easier to represent statistical properties suitable for the problem at hand. Well-known PGMs include Bayesian networks, hidden Markov models, Markov random fields, among others. PGMs allow for a move from probability distributions, which are rich in detail, to graphs that abstract away from such details by encoding independence relationships. This occurs by means of a *qualitative* semantics entailed by the graphical structure of PGMs.

The qualitative information encoded in a PGM is appealing for domain experts, who can read off relationships from the graph, such as whether a variable A becomes irrelevant for the prediction of B when C is known. The graphical representation allows for answering such queries often in a computationally efficient way. This represents an alternative way to identify independence properties by not relying on calculating numerical probabilities that can be computationally

demanding. PGMs also have a *quantitative* semantics by assigning numerical parameters to nodes in the graph, which allows for computing probability queries with full detail.

When it comes to model building, a number of advantages result from using graphs to represent distribution properties. At the domain-expert level, it is possible to specify the desired level of restrictions on the way variables can interact probabilistically. For example, one might say that the variables should be independent (leading to an empty graph), or that interactions should follow a tree-like pattern, or even specify detailed interactions. The language of graphs is intuitive enough to allow one to easily specify such patterns. Independence relationships between variables can also be derived in a completely algorithmic manner without any expert knowledge, as methods have been developed for learning the graphical structure of PGMs from data [163, 169]. A hybrid approach is also possible by combining expert knowledge with data-driven learning.

1.3 MODELING SEQUENTIAL BEHAVIORS

Dynamic Bayesian networks (DBNs, for short) [68, 104, 124] are well-known PGMs used for representing temporal processes. The process that a DBN models is assumed to be a first-order process (or memoryless), which means that the future state of the process depends only the present state. The process is also assumed to be time homogeneous, which means that probabilities for transitioning from time t to time $t + 1$ are the same for every $t \geq 0$. As a result, DBNs offer a parsimonious representation of temporal processes by requiring the specification of typically a small number of probability parameters. DBNs can also be seen as extensions of discrete Markov chains to multiple variables. Example 1.1 discusses the dynamics of a mental disorder treatment based on DBNs.

Example 1.1. *Suppose we want to model the symptom dynamics of patients with psychotic depression [147], which is a depressive disorder with psychotic features. On a regular basis, two psychotic features (Delusions and Hallucinations) and Depression are measured for each patient. Due to the DBN assumptions aforementioned, the structure of a DBN for this problem is given by:*

- *A graph over the variables {Delusions, Hallucinations, Depression} indicating the symptom interactions at $t = 0$.*

- *A graph over $\{Delusions^{(t)}, Hallucinations^{(t)}, Depression^{(t)}, Delusions^{(t+1)}, Hallucinations^{(t+1)}, Depression^{(t+1)}\}$ indicating the transitioning interaction of symptoms at two any time points, where $t \geq 0$.*

The transition structure of a DBN for patient dynamics is shown in Figure 1.1. Although Figure 1.1 shows only one transition, this model would normally be unrolled for any discrete time horizon $\{0, 1, \ldots\}$. The transition structure and numerical parameters are the same for every transition due to the homogeneity assumption.

Time t Time $t+1$

Delusions ————————————→ Delusions

Hallucinations ————————————→ Hallucinations

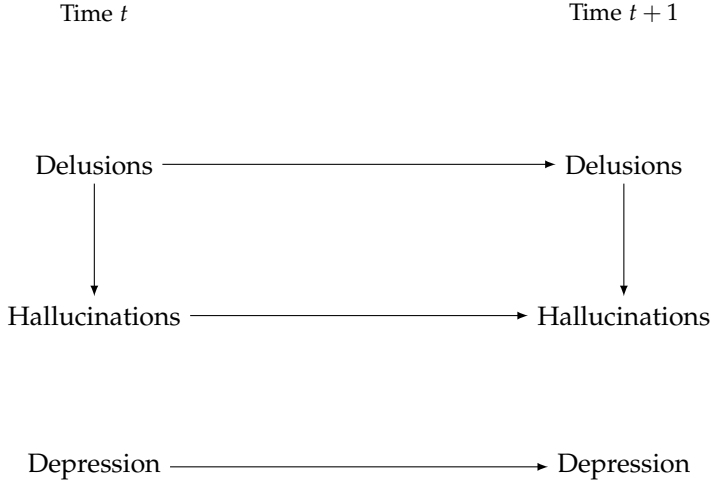Depression ————————————→ Depression

Figure 1.1: Transition structure of a dynamic Bayesian network that represents symptom interaction of psychotic depression patients.

In Figure 1.1, an arc between two variables indicates that these variables may be statistically dependent. On the other hand, two variables that are indirectly linked in the graph can still be statistically dependent, however this depends on the configuration of other variables in between them. When two variables are not linked directly nor indirectly, they are statistically independent [136].

Each arc of a DBN can refer to an *instantaneous* interaction, i.e. an interaction within the same time point, or a *temporal* interaction, i.e. an interaction that occurs at different time points [104]. The temporal arcs should satisfy the natural temporal order, i.e., there must be no arcs with direction from future to past. In Example 1.1, the instantaneous arc from Delusions to Hallucinations indicates that at any point where measurements were made, delusions and hallucinations are statistically correlated. On the other hand, the temporal arcs for current depression to future depression means that the depression score of a patient at a certain week has influence on the patient's depression the week after (as one would expect in general).

What makes the model of Figure 1.1 a temporal model is the temporal arcs, because otherwise variables at a time point t_2 would be statistically independent of any variable at time $t_1 < t_2$. The instantaneous interactions are not strictly needed for the model to be temporal, but they are often used to represent more complex statistical relationships.

As the process is assumed to be time homogeneous, in the model of Example 1.1 not only the graphical structure is fixed over time, but also the numerical parameters that describe the transition probabilities. Just as the initial graph structure, the initial numerical parameters might also differ from the transition parameters.

1.4   ADDING MORE EXPRESSIVE POWER

One consequence of the compactness of models as DBNs is that they capture the *average* process behavior over time. This is because by having a transition model that is time invariant, the structure and numerical parameters of a DBN are the same for every time point. However, this might not always be desirable, because processes might change over time. By restraining ourselves to models that represent average behaviors, we might lose the opportunity to learn important insight about processes.

Process change might be captured by different ways. One would expect that real processes are in constant change. However, we would like to capture sensible process changes in our models, which would allow us to arrive at parsimonious explanations of the process at hand. We discuss 3 situations (or *challenges*) where it is desirable to represent processes in a more expressive way.

1.4.1   *Time-dependent representation*

One situation where we might be interested in more expressive models occurs when we wish to capture process change that manifests by varying variable interaction over time. For example, if the instantaneous and temporal interactions between Hallucinations and Depression are substantially different over time in Example 1.1, one would likely obtain a better understanding of patient evolution by modeling symptom interaction change in an explicit fashion.

In the case of models such as DBNs, process change could manifest as changes in the graphical structure, numerical parameters, or both. One challenge that arises in modeling is the identification of process change-points (or regime change) which take into account reasonable process change assumptions.

1.4.2   *Factor-dependent representation*

A different way of looking at process change is by identifying latent variables. Such factors can be seen as unmeasured quantities that are missing or difficult to be measured [178]. Latent variables can also be seen as categories or abstract concepts derived from observable data, such as intelligence and extraversion in the context of human behavior [17]. In the latter case, latent variables might act as a dimensionality reduction tool of observable data as it might be easier to understand the data in terms of the usually more compact latent representation.

In Example 1.1, latent factors that might be associated to process change could be medical treatment, environmental factors (age, gender, climate, etc.), genetic expression, etc. Some of these factors might actually happen to be measured (e.g. medication and dosage), which would allow for explicitly including them as observable variables. At other times we are interested in learning latent concepts from observable data, such as patient clusters in Example 1.1, which might help us understand patient dynamics by means of a succinct representation.

Hidden Markov models (HMMs, for short) [138, 141] are PGMs that use latent variables to model sequential processes. HMMs can capture very complex distributions by using a suitable latent state space [13]. The standard assumption in HMMs is that observable variables are independent given the latent state. However, representing complex observed behavior based on such assumption can require too many states [76]. This is undesirable for several reasons, e.g., computing cost and less interpretable models.

Extensions of HMMs that are able to represent more general variable interaction have been proposed [12, 76, 102]. One challenge that arises is a better understanding of how such HMMs compare in theory and in practice. Another relevant challenge is how to generate problem insight from more expressive HMMs which can help one understand the dynamics of processes (e.g. disease processes) in an effective way. This is relevant because the task of interpreting latent states is usually not straightforward.

1.4.3   *Subprocess representation*

A different viewpoint on process change concerns the identification of *subprocesses* which deviate considerably from the main model (or main process). We refer to 'main process' and 'subprocess' as processes associated to the model of the whole dataset and the model of a subset of the data respectively. In this situation, we would like to identify subsets of the data that are also representative enough, because it is easy to come up with very specific subprocesses made out of just a few data points (hence, not representative).

In Example 1.1, there could exist a subset of patients with certain psychotic symptoms that have a substantially slower response to treatment than the average patient response. We would like to identify such subsets (or subprocesses) in an automatic fashion. However, we cannot directly identify such subprocesses by models such as standard DBNs and HMMs.

One approach to identifying deviating subprocesses is the exceptional model mining (EMM, for short) [57, 110], whose goal is the discovery of exceptional models (in the sense of significantly different) associated to subsets of the data. However, EMM has been limited to either static data or univariate temporal data [112]. It would be desirable to extend the EMM framework for more general temporal data.

1.5   THESIS OUTLINE

This work addresses the discovery of structure from temporal data that can aid the comprehension of the underlying processes. For convenience, this work is divided into three parts as follows. In Chapters 3-5, we investigate the underlying structure of processes by means of models that use latent variables. In Chapter 6, we investigate how temporal processes can be better understood by identifying process changepoints or regime change. Finally, in Chapter 7 we investigate

how temporal data can be decomposed based on data subgroups that have a substantially different characterization in terms of a set of target variables compared to the distribution of those targets in the whole data. As a result, Chapter 7 provides a different characterization of process structure compared to those of other chapters.

In order to demonstrate the usefulness of the proposed methods, we use real-world data from several domains, including medical data (e.g. primary care and clinical trials), industrial processes and business processes. We also discuss problem insight that can be obtained by the application of the methods to such datasets. We summarize the content of each chapter in the following.

CHAPTER 2: *Preliminaries*, where notions on PGMs relevant for representing temporal processes are discussed.

CHAPTER 3: *Asymmetric hidden Markov models*, where we introduce the family of *asymmetric hidden Markov models* (HMM-As, for short) for representing local structure of distributions in the hidden Markov model framework. An algorithm for learning HMM-As from temporal data is proposed. HMM-As are empirically evaluated based on simulated data and real-world data from several domains. This chapter is based on the publications [25] and [23].

CHAPTER 4: *Predicting disease dynamics: a case study of psychotic depression*, in which a methodology is proposed for aiding the generation of medical hypotheses based on structured hidden Markov models learned from data. The methodology is used to uncover insight on the dynamics of different pharmacological therapies undertaken by psychotic depression patients. This chapter is based on the publications [27] and [28].

CHAPTER 5: *Understanding multimorbidity through clusters of hidden states*, where we analyze the problem of disease interaction and multimorbidity in terms of patterns of transitions between latent states. We consider a study case of patients with disorders related to atherosclerosis based on a large primary care data, and show that multiple patient characterization can be associated to cluster of states. This chapter is based on the publication [26].

CHAPTER 6: *Partitioned dynamic Bayesian networks*, in which we propose *partitioned dynamic Bayesian networks* (PDBNs, for short) for representing temporal processes by means of a collection of dynamic Bayesian networks. We propose a learning algorithm for PDBNs which adds process cut-offs in a parsimonious way. PDBNs are evaluated experimentally based on simulations and real data. This chapter is based on the publication [24].

CHAPTER 7: *Exceptional model mining using dynamic Bayesian networks*, where we investigate how observable data can be decomposed in a way different than that pursed in the previous chapters. We propose a method to identify subgroups of data that are exceptionally different than the total data based on the framework of subgroup discovery and exceptional model mining.

Subgroups are characterized by dynamic Bayesian networks. This chapter is based on the paper [22], which was submitted for publication.

CHAPTER 8: *Discussion*, in which the results achieved in this work are summarized and future directions to be pursued are discussed.