



Universiteit
Leiden
The Netherlands

Unraveling temporal processes using probabilistic graphical models

de Paula Bueno, M.L.

Citation

De Paula Bueno, M. L. (2020, February 11). *Unraveling temporal processes using probabilistic graphical models*. *SIKS Dissertation Series*. Retrieved from <https://hdl.handle.net/1887/85168>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/85168>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



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

SUMMARY

Temporal processes, such as walking, sleeping, eating, and so on, are ubiquitous in daily life as well as in more intricate situations such as medical treatment, seasonal climate variation, events in a workflow, and so on. We are often interested in understanding how aspects of objects of study evolve, such as signs and symptoms of disorders of patients. On the one hand, we need *expressive* enough models for capturing complex behavior. On the other hand, such models should provide *parsimonious* descriptions of processes if one wishes to gain insight. This balance is not trivial, as by increasing expressivity one often arrives at more complicated models, which might make them less interpretable. It is often the case that suitable descriptions of processes also need that uncertainty be explicitly recognized.

In this thesis, we aim to increase model expressivity inspired by complex and real-life problems, while retaining model interpretability. To this end, we describe three new different viewpoints on processes based on probabilistic graphical models.

We first provide a new process viewpoint based on *latent states*, which can be seen as abstract representations of the observable data. Latent states can help interpretation as they act as a dimensionality reduction tool.

In Chapter 3, we introduce asymmetric hidden Markov models for capturing local structure in the space of observable variables. This is done by associating each latent state to a Bayesian network. Asymmetric hidden Markov models often lead to better model fit and increased insight into the domain, while reducing the need for selecting an *a priori* model architecture. Simulated and real-world datasets are used for empirical evaluation.

In Chapter 4, we propose a semi-automatic framework for understanding disease dynamics based on the dynamics of latent states within hidden Markov models. We apply the framework to psychotic depression treatments, where latent states act as patient groups and are shown to reveal predictive symptoms to patient prognosis.

In Chapter 5, we learn hidden Markov models from health-care event data. A case study based on atherosclerosis events is used. The size of such datasets, in contrast with a small number of events, makes the same event be associated to multiple hidden states, a notion we call clustering of hidden states. We show that events in a cluster associate to patients with different disease severity.

The second viewpoint on processes is based on the identification of process *change-points* or *regime change*. The challenge lies in how to extend models that are time invariant (such as dynamic Bayesian networks) for capturing regime change in a parsimonious way, which can be suitable when the available dataset is small.

In Chapter 6, we propose partitioned dynamic Bayesian networks for representing models for which the time homogeneity assumption is not suitable. Partitioned dynamic Bayesian networks are a collection of dynamic Bayesian networks for which cut-off points are built heuristically. The resulting models are evaluated in a wide set of experiments.

The last process viewpoint tries to discover subsets of temporal data associated to models that deviate substantially from the model obtained from the whole dataset. This can be seen as identifying significant *subprocesses*.

In Chapter 7, we introduce dynamic Bayesian networks for representing exceptional temporal models of data subgroups. This provides a general representation for subprocesses within the context of subgroup discovery and exceptional model mining. We evaluate the proposed approach by means of simulated and event data on farmer financial support applications.