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

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

8

DISCUSSION

This thesis dealt with the discovery of the underlying structure of temporal processes. The underlying structure of a process can be captured by different mathematical representations, depending on which aspects one wishes to focus on. Yet, the task of choosing suitable models to represent a particular real-world problem is challenging.

In this work, we advocate that the proposed methods based on probabilistic graphical models are advantageous when solving several real-world problems. Just as PGMs, the methods introduced in this work can work well under situations of varied data scarcity and are often interpretable. One advantage of being interpretable is that insight can often be obtained from such models without a lot of barriers. Although models such as deep neural networks have been recently very successful at tasks such as supervised learning, they tend to require higher amounts of data, large amounts of computational resources (e.g., computing time) and are considerably less interpretable, characteristics which go in an opposite direction to PGMs.

8.1 CONTRIBUTIONS

The central theme covered in this work is the different viewpoints on dynamics of temporal processes, which we summarize in the following sections.

8.1.1 *Asymmetry in models*

Hidden Markov models have been used for providing a viewpoint of processes based on latent variables. In order to increase the problem insight and model fit that one can obtain from HMMs, we proposed asymmetric HMMs (HMM-As, for short) in Chapter 3.

HMM-As allow for more expressive representation of observations by capturing distribution asymmetries (also known as local structure). As a result, HMM-As often need fewer latent states to achieve model fit at least as good as that of other HMM models. The parsimonious representation of HMM-As is also valuable in model interpretation. Due to the large number of variations of existing HMMs, it

is often difficult to decide which model class would be suitable for the problem at hand. The flexibility of the HMM-A representation reduces the need for deciding which class of symmetric and other asymmetric HMMs one should use. In Chapter 3, we demonstrated the aforementioned advantages of HMM-As by means of simulated and real-world data from several domains.

8.1.2   *Generation of hypotheses on processes*

We proposed in Chapter 4 a method that helps the selection of hypotheses to further investigate about disease processes. This method is semi-automatic and is based on structured HMMs, particularly with the goal of generating insight on disease processes. The formulation of outcomes is aided by means of *state reachability* that one can build by looking at model aspects such as state probabilities at different time points. The state reachability notion was shown to help understanding patient trajectories in a compact way.

A case study of psychotic depression was considered in Chapter 4, for which hypotheses were generated. One of the main results discussed is that patients undergoing psychotic depression treatment showed to be sensitive to treatment based on their initial psychotic symptoms. By using this methodology, new knowledge about this mental disorder was acquired, which potentially helps doctors to prescribe more efficient medication in the future.

8.1.3   *Capturing hidden (non-observed) aspects of processes*

Besides the independence structure of processes given by asymmetric HMMs (Chapter 3), we investigated in Chapter 5 how to gain insight in health care data by means of latent-variable modeling. In this chapter we introduced the notion of cluster of hidden states based on hidden Markov models. Clusters of states can be learned from datasets with a single event produced at each instant.

Based on health care data from Dutch practices, clusters of latent states were learned. The clusters were shown to provide additional characterization to the latent states by suggesting that states from each cluster are correlated to different patient severity. Ultimately, the results of Chapter 5 also allow for gaining insight on multimorbidity by analyzing the clusters of states learned for different disease codes.

8.1.4   *Taking into account the size of datasets*

Learning insightful models from small datasets is known to be a challenging problem. In this work, we proposed new methodologies and model classes that showed to be useful in real-world situations of the limited availability of data. In Chapter 6, we proposed partitioned dynamic Bayesian networks for representing dynamic Bayesian networks with regime change over time. A heuristic was

proposed for identifying regime cut-offs of PDBNs in a parsimonious way, which favors the situation of small datasets.

Experiments based on simulations showed that PDBNs learned by the proposed heuristic search dealt well with different situations, in particular with small datasets generated by different underlying models. PDBNs learned from psychotic depression treatment data provided a better model fit than DBNs and more insight on the interaction between psychotic and depressive features. Different viewpoints on disease dynamics for psychotic depression treatment were provided in Chapters 4 and 6 by investigating different variables and problem aspects.

8.1.5 *Temporal subgroups*

A new problem called temporal exceptional model mining (TEMM, for short) was defined in Chapter 7. TEMM aims to discover exceptional behavior associated to subsets of the data that can be described by a configuration of variables.

We proposed a method that allows for the discovery of exceptional dynamic Bayesian networks by means of a distance measure and a search algorithm. The proposed temporal representation allows for more accurate retrieval of exceptional subgroups than that based on simpler temporal models such as Markov chains. The method for identifying exceptional DBNs was evaluated based on real-world process data and was able to discover exceptional subgroups in a principled fashion. The results were validated in comparison to previous research, based on different techniques.

8.2 FUTURE WORK

In this section, we discuss limitations and directions for future work that we believe might be relevant. We also discuss potential approaches that could be relevant for achieving these goals.

8.2.1 *Asymmetry in models*

The number of states of HMM-As was selected by a trial and error approach. It is worthwhile to investigate more principled ways to make this selection. One alternative is to use infinite HMMs [8] to automatically determine the number of hidden states. Another alternative is to predict the number of states by means of Bayesian optimization [156], where the performance (e.g. the goodness of fit) of each state could be seen as a complex black-box function.

8.2.2 *Generation of hypotheses on processes*

We would like to apply the hypothesis generation methodology proposed in Chapter 4 in a more *automatic* fashion. This means automatizing the selection of

baseline and target states, the computation of state reachabilities, and so on. This is already possible given the proposed method, which would likely result in a set of candidate hypotheses to further investigate.

By applying the method of Chapter 4 to other datasets, it would be possible to evaluate the effectiveness of automatic generation of hypothesis. The advantage of using multiple datasets is a more effective evaluation of the effect of different definitions for selecting baseline and target states, including those definitions proposed in Chapter 4.

One new challenge that would likely arise by automatizing hypothesis generation is how to properly assess different hypotheses that would be generated. To that end, one could benefit from currently available electronic health records [30, 142], which might lead to additional sources of information, such as clinical notes written by medical doctors (often in natural language). Such data could perhaps give direction to which hypotheses could be promising to be investigated, e.g., by assigning some kind of utility to the generated hypothesis.

8.2.3 *Capturing hidden (non-observed) aspects of processes*

In Chapter 5, the analysis of clusters based on a medical outcome (in this case, disease counts) was carried out after the clusters of states were identified. We would like to integrate medical outcomes as part of the models, so that the relationship with the clusters of states and outcome measure could be direct and more general, which would allow one to look at such relationship by different angles. One could consider doing this by integrating such outcomes during model learning, e.g., by means of covariates along the lines of Input-Output HMMs [11].

A small number of diagnosis variables was considered when learning clusters of states based on the case study of Chapter 5. It would be interesting to look at a larger set of variables and the resulting clusters, which seems a natural extension to the presented method since health care data often have hundreds or more diagnosis codes.

To make multimorbidity analysis more effective, we would like to consider data representations where multiple diseases occur at the same time. This would, however, result in a significant departure from the modeling assumptions considered of Chapter 5, as the observation space would likely have several active variables, which could create the need for a different notion of cluster of states. The advantage, however, is that multimorbidity could be analyzed in a more direct way.

8.2.4 *Taking into account the size of datasets*

We would like to extend PDBNs to make it possible to model recurrent regimes over time, such that the identified regimes could be seen as states similarly to HMM states. This could make PDBNs more compact and more explainable

models. Related research that considers recurrent regimes includes, e.g., gated networks [10] and DBNs with an HMM-based dependence structure [78].

We also would like to broaden the evaluation of PDBNs by comparing them to other non-homogeneous models, which are usually based on different assumptions. This would allow for evaluating different classes of models under different assumptions and how they perform when some (or all) assumptions are not met.

8.2.5 *Temporal subgroups*

In TEMM, no relationship between discovered subgroups is explicitly computed. Nevertheless, one can argue that some kind of relationship between the discovered exceptional subgroups might exist. One simple example is when multiple subgroups represent (approximately) the same exceptional behavior, i.e., they deviate from the population in (approximately) the same way.

By identifying subgroups that are exceptional and yet similar, not only redundancies could be reduced, but also more insight about the problem would be obtained as to how exceptional behavior might occur. Redundancies could be eliminated by rejecting specialized subgroups that are similar to more general subgroups (see, e.g., [112]). Instead of rejecting subgroups, an alternative is to introduce merge and split operations in the search algorithm. In that case, the language for expressing subgroups could be extended to represent other description patterns beyond pairs of attribute-value, which could allow for a more general understanding between description patterns and exceptional behavior.

Another research direction that might be worthwhile to explore in TEMM is what makes a model an exceptional model. While this might be more or less evident when one deals with exceptional models based on just a few variables (as in standard subgroup discovery, for example), this is no longer the case for more complex models such as dynamic Bayesian networks. For DBNs, one could be interested in knowing whether particular pieces of the model structure or particular parameters are relevant for explaining why a DBN is considered as exceptional. Sensitivity analysis [34, 72] might be of help for pursuing this research direction.

The computation of distributions of false discoveries based on DBNs is expensive. We would like to investigate whether further optimizations can be used to reduce these computations. One idea is to try to predict new distributions of false discoveries based on previously computed ones, e.g., by predicting its parameters such as the mean and standard deviation. One method that might help is Bayesian optimization [156, 160], as obtaining the distributions of false discoveries can be seen as evaluating an expensive black-box function.