

University of Groningen

Advanced non-homogeneous dynamic Bayesian network models for statistical analyses of time series data

Shafiee Kamalabad, Mahdi

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2019

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Shafiee Kamalabad, M. (2019). *Advanced non-homogeneous dynamic Bayesian network models for statistical analyses of time series data*. University of Groningen.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Summary

One statistical challenge in many fields is to infer network topologies of interacting units from time series data. One class of statistical models, which has been widely applied to deal with this task, is dynamic Bayesian network models (DBNs). The underlying assumption of the conventional DBNs is that the underlying process is a homogeneous Markov process, so that DBNs do not allow the network parameters to change in time. Therefore, DBNs cannot deal with non-homogeneous and non-stationary regulatory processes, which arise in many important real world applications.

Recently, non-homogeneous dynamic Bayesian network models (NH-DBNs) have been introduced and become an important statistical tool to relax this restrictive assumption. NH-DBNs have been implemented with various allocation models to divide the temporal data into disjoint data subsets. Those models infer the data segmentation, the joint network structure and the segment- or component-specific interaction parameters from the data.

In this thesis we have focused on improving changepoint (CPS) divided NH-DBNs which have become the most widely applied NH-DBNs to model complex systems. These models infer changepoints which divide the data into disjoint segments and the segment-specific network parameters are learned for each segment separately. In many real world applications these NH-DBNs divide a time series into even shorter segments. Learning the network parameters for each segment separately ('uncoupled' NH-DBN models), leads to over flexibility and inflated inference uncertainties. Moreover, these models do not incorporate the reasonable prior assumption that neighboring segments are often more likely to have similar network interaction parameters than distant segments. To address these bottlenecks, Bayesian models with coupling mechanisms between the segment-specific parameters have been proposed.

Bayesian models with parameter coupling can lead to significantly improved network reconstruction accuracies when the segment-specific parameters are similar. However, recently we have found that coupling can become counter-productive when the segment-specific parameters are dissimilar. The reason for that is that neither the sequential nor the global coupling scheme has an effective mechanism for uncoupling. For many real-world applications this is a constraint. We have addressed these bottlenecks in this thesis by introducing four novel

NH-DBNs.

Another scenario corresponded to many real-world applications, happens when time series data are often collected under different experimental conditions. That is, instead of one single time series, which can be divided into segments with natural temporal order, there are K (short) time series with no natural order. They are exchangeable units and there is no need for inferring the segmentation. In this situation it is often unclear a priori whether the network parameters are actually component-specific or whether they are constant across components. Whereas in real-world applications there can be both types of parameters simultaneously. We, therefore, have addressed this problem by introducing novel partially NH-DBNs based on Bayesian regression models with partition design matrix.

In **chapter 1** we have given an introduction and an overview to existing network models and we have outlined this thesis.

In **chapter 2**, we have proposed two new models based on piecewise Bayesian regression models: The partially segment-wise coupled NH-DBN model and the generalized fully sequentially coupled model. Our empirical results have shown that the partially coupled model leads to improved network reconstruction accuracies. For the generalized coupled model we have not seen consistent improvements over the fully coupled NH-DBN model.

In **chapter 3**, we have therefore refined the generalized fully sequentially coupled model. For the refined model with a hyperprior onto the second hyperparameter of the coupling parameter prior we have seen improved network reconstruction accuracies.

In **chapter 4**, we have presented a novel NH-DBN model with partially edge-wise coupled segment-specific network parameters. This model operates on the individual edges. Instead of enforcing *all* edges to be coupled, our model operates edge-wise and infers for each individual edge from the data whether the associated parameters should be coupled or stay uncoupled across all segments. We have empirically shown on yeast gene expression time series that the new model reaches a highest network reconstruction accuracy. For Arabidopsis thaliana gene expression data, we have shown that our new model not only outputs a network prediction, but also allows to distinguish between edges whose regulatory effects stay similar across time and edges whose regulatory effects are subject to more substantial temporal changes.

In **chapter 5**, we have introduced a partially NH-DBN model, which is effectively a Bayesian regression model with partitioned design matrix. The new model aims to infer the best trade-off between a homogeneous model and a non-homogeneous model. For each network interaction there is a parameter, and the new model infers from the data whether this parameter is constant or whether it varies among segments. We, moreover, have proposed to employ a Gaussian process based approach to deal with non-equidistant measurements. Our applications to yeast data have shown that the new model improves the network reconstruction accuracy. We have used the new model to reconstruct the topologies of the mTORC1 data. The inferred network topologies showed features that are consistent with the biological literature.

Chapter 6 has been on a comparative evaluation study on popular non-homogeneous Poisson models for count data. For this study the standard homogeneous Poisson model (HOM) and three non-homogeneous variants, namely a Poisson changepoint model (CPS), a Poisson free mixture model (MIX), and a Poisson hidden Markov model (HMM) have been implemented in both frequentist and Bayesian framework. The first major objective has been cross-comparing the performances of the four aforementioned models independently for both modelling frameworks (Bayesian and frequentist). Subsequently, a pairwise comparison between the four Bayesian and the four frequentist models has been performed to elucidate to which extent the results of the two paradigms ('Bayesian versus frequentist') differ.

