AFRL-AFOSR-VA-TR-2019-0276

Scalable Representational Structures for Complex Multivariate Time Series

Emily Fox
UNIVERSITY OF WASHINGTON

09/18/2019
Final Report

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/ RTA2
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
18-09-2019

2. REPORT TYPE
Final Performance

3. DATES COVERED *(From - To)*
01 Nov 2015 to 31 Oct 2018

4. TITLE AND SUBTITLE
Scalable Representational Structures for Complex Multivariate Time Series

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-16-1-0038

5c. PROGRAM ELEMENT NUMBER
61102F

6. AUTHOR(S)
Emily Fox

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
UNIVERSITY OF WASHINGTON
4333 BROOKLYN AVE NE
SEATTLE, WA 98195-0001 US

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AF Office of Scientific Research
875 N. Randolph St. Room 3112
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/AFOSR RTA2

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-AFOSR-VA-TR-2019-0276

12. DISTRIBUTION/AVAILABILITY STATEMENT
A DISTRIBUTION UNLIMITED: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This grant led to developments in approaches to learn interpretable and
exible representa-
tions of interactions in complex time series arising in a range of applications, including the study of brain networks from neuroimaging modalities, the dynamics of homelessness, gene regulatory networks, climate and economic datasets. Three fundamental challenges were tackled: (i) modeling the intricate contemporaneous interactions between groups of observed dimensions (e.g., each group representing the cortically localized signals in a region of interest in the brain), (ii) capturing lagged, directed interactions in complex time series through Granger causality detection, and (iii) developing models of functional connectivity from neuroimaging data that exploit related methods for learning interactions in time series. Our results demonstrate the potential to uncover interpretable notions of both static and dynamic interactions in nonlinear, non-Gaussian time series.

15. SUBJECT TERMS
Bayesian inference, machine learning, statistics and probability, time series analysis, artificial intelligence, data fusion

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
Unclassified	Unclassified	Unclassified	UU		RIECKEN, RICHARD

					19b. TELEPHONE NUMBER *(Include area code)* 703-941-1100

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

DISTRIBUTION A: Distribution approved for public release.

Subject: Final Report to Dr. Doug Riecken

Contract/Grant Title: Scalable Representational Structures for Complex Multivariate Time Series
Contract/Grant #: FA9550-16-1-0038
Reporting Period: 1 November 2015 to 31 October 2018

Abstract:

This grant led to developments in approaches to learn interpretable and flexible representations of interactions in complex time series arising in a range of applications, including the study of brain networks from neuroimaging modalities, the dynamics of homelessness, gene regulatory networks, climate and economic datasets. Three fundamental challenges were tackled: (i) modeling the intricate contemporaneous interactions between groups of observed dimensions (e.g., each group representing the cortically localized signals in a region of interest in the brain), (ii) capturing lagged, directed interactions in complex time series through Granger causality detection, and (iii) developing models of functional connectivity from neuroimaging data that exploit related methods for learning interactions in time series. Our results demonstrate the potential to uncover interpretable notions of both static and dynamic interactions in nonlinear, non-Gaussian time series.

Final accomplishments:

Below we detail the accomplishments arising from this funded project. In particular, we tackled three fundamental challenges: (i) modeling the intricate contemporaneous interactions between groups of observed dimensions (e.g., each group representing the cortically localized signals in a region of interest in the brain), (ii) capturing lagged, directed interactions in complex time series through Granger causality detection, and (iii) developing models of functional connectivity from neuroimaging data that exploit related methods for learning interactions in time series.

Contemporaneous and Lagged Interactions in Complex Time Series We developed methods to flexibly learn both contemporaneous, undirected and lagged, directed interactions in complex multivariate time series by studying both (i) nonlinear latent factor models of complex covariance structures and (ii) Granger causality networks in non-traditional settings. Specifically, for the latter, we considered the following set of important settings to greatly expand the reach of Granger causal analyses: (1) the subsampled and mixed frequency setting where each component series might be sampled at a different rate, (2) multivariate categorical time series, and (3) settings where the series have potentially nonlinear dynamics. These developments are outlined below.

Flexible and interpretable descriptors of contemporaneous interactions Deep generative models have recently yielded encouraging results in producing subjectively realistic samples of complex data. Far less attention has been paid to making these generative models interpretable. In many scenarios, ranging from scientific applications to finance, the observed variables have a natural grouping. It is often of interest to understand systems of interaction amongst these groups, and latent factor models (LFMs) are an attractive approach. However, traditional LFMs are limited by assuming a linear correlation structure. In [4], we presented an output interpretable VAE (oi-VAE) for grouped data that models complex, nonlinear latent-to-observed relationships. We combined a structured VAE comprised of group-specific generators with a sparsity-inducing prior. We demonstrated that oi-VAE yields meaningful notions of interpretability in the analysis of motion capture and MEG data. In particular, even in a very low SNR regime, we learned networks of interactions between regions of interest that coincide with known networks, such as the default mode and dorsal attention networks. We further showed that in these situations, the regularization inherent to oi-VAE can actually lead to improved generalization and learned generative processes.

Granger causality in subsampled and mixed frequency time series Causal inference in multivariate time series is challenging because the sampling rate may not be as fast as the time scale of the causal interactions, so the observed series is a subsampled version of the desired series. Furthermore, series may be observed at different sampling rates, yielding mixed-frequency series. To determine instantaneous and lagged effects between series at the causal scale, we took a model-based approach that relies on structural vector autoregressive models [2]. We presented a unifying framework for parameter identifiability and estimation under subsampling and mixed frequencies when the noise, or shocks, is non-Gaussian. By studying the structural case, we developed identifiability and estimation methods for the causal structure of lagged and instantaneous effects at the desired time scale. We further derived an exact expectation-maximization algorithm for inference in both subsampled and mixed-frequency settings. We validated our approach in simulated scenarios and on a climate and an econometric dataset.

Granger causality in categorical time series In [8] (journal version under review), we presented a new framework for learning Granger causality networks for multivariate categorical time series, based on the mixture transition distribution (MTD) model. Traditionally, MTD is plagued by a non-convex objective, non-identifiability, and presence of many local optima. To circumvent these problems, we recasted inference in the MTD as a convex problem. The new formulation facilitates the application of MTD to high-dimensional multivariate time series. As a baseline, we also formulated a multi-output logistic autoregressive model (mLTD), which while a straightforward extension of autoregressive Bernoulli generalized linear models, had not been previously applied to the analysis of multivariate categorial time series. We developed novel identifiability conditions of the MTD model and compared them to those for mLTD. We further devised a novel and efficient optimization algorithm for the MTD based on the new convex formulation, and compared the MTD and mLTD in both simulated and real data experiments. Our approach simultaneously provides a comparison of methods for network inference in categorical time series and opens the door to modern, regularized inference with the MTD model.

Neural Granger causality While most classical approaches to Granger causality detection assume linear dynamics, many interactions in applied domains, like neuroscience and genomics, are inherently nonlinear. In these cases, using linear models may lead to inconsistent estimation of Granger causal interactions. In [6] (journal version under review), we proposed a class of nonlinear methods by applying structured multilayer perceptrons (MLPs) or recurrent neural networks (RNNs) combined with sparsity-inducing penalties on the weights. By encouraging specific sets of weights to be zero—in particular through the use of convex group-lasso penalties—we can extract the Granger causal structure. To further contrast with traditional approaches, our framework naturally enables us to efficiently capture long-range dependencies between series either via our RNNs or through an automatic lag selection in the MLP. We showed that our neural Granger causality methods outperformed state-of-the-art nonlinear Granger causality methods on the DREAM3 challenge data. This data consists of nonlinear gene expression and regulation time courses with only a limited number of time points. The successes we showed in this challenging dataset provide a powerful example of how deep learning can be useful in cases that go beyond prediction on large datasets. We likewise demonstrated our methods in detecting nonlinear interactions in a human motion capture dataset.

Connectivity Analysis of MEG Data Connectivity analyses of MEG have traditionally relied on a two-stage procedure, with the first stage involving an ill-posed mapping of sensor measurements to cortical signals. This inverse solution does not take signal dynamics into account and introduces other artifacts. For our functional connectivity analysis, we instead devised a state space linear dynamical system approach that overcomes these issues, and provides a modeling framework in which sharing of information between multiple subjects is possible. Performing multi-subject analyses in this way boosts the signal-to-noise ratio, fundamental to robust inference of underlying networks. Crucially, the model involves learning *time-varying directed* connections between regions of interest. This moves beyond the contemporaneous interactions learned by the oi-VAE described above [4].

In a separate approach to learning functional connectivity networks, in [7] we considered the *undirected and static* interactions described by a graphical model. In contrast to typical approaches to learning graphical model structures—here representing our functional connectivity network of the cortically-localized signals—we treat the MEG signals as time series rather than i.i.d. observations. We also incorporated a low-rank component that accounts for latent signals that would otherwise lead to inferring spurious connections. We developed a penalized likelihood approach using convex regularizers and developed a suite of efficient optimization approaches to perform the structure learning.

In our review paper [1], we presented a literature review on model-based approaches to estimating functional connectivity from neuroimaging data, including the classes of methods described above. In contrast to the typical focus on a particular scientific question, we reframed a wider literature in terms of the underlying statistical model used. We distinguished between directed versus undirected and static versus time-varying connectivity. There are numerous advantages to a model-based approach, including easily specified inductive bias, handling limited data scenarios, and building complex models from simpler building blocks.

In addition to these efforts, we also explored the use of modern variational autoencoders (VAEs) to analyze replicate network data [5]. In contrast to a recently proposed mixture model, the VAE approach addresses two problems: i) representing networks that individually exhibit multiple types, and ii) representing all observations in the same latent space. Finding an embedding of the observed networks in the same latent space is useful for a variety of reasons, including visualizing the distribution of observed networks and for further predictive tasks.

Dynamics of Homelessness As a separate study of learning interpretable dynamics in a noisy, complex dataset, we analyzed the dynamics of homelessness in [3]. The relationship between housing costs and homelessness has important implications for the way that city and county governments respond to increasing homeless populations. Though many analyses in the public policy literature have examined inter-community variation in homelessness rates to identify causal mechanisms of homelessness [J. Urban Aff. 35 (2013) 607625; J. Urban Aff. 25 (2003) 335356; Am. J. Publ. Health 103 (2013) S340S347], few studies have examined time-varying homeless counts within the same community [J. Mod. Appl. Stat. Methods 15 (2016) 15]. To examine trends in homeless population counts in the 25 largest U.S. metropolitan areas, we developed a dynamic Bayesian hierarchical model for time-varying homeless count data. Particular care was given to modeling uncertainty in the homeless count generating and measurement processes, and a critical distinction was made between the counted number of homeless and the true size of the homeless population. For each metro under study, we investigated the relationship between increases in the Zillow Rent Index and increases in the homeless population. Sensitivity of inference to potential improvements in the accuracy of point-in-time counts was explored, and evidence was presented that the inferred increase in the rate of homelessness from 20112016 depends on prior beliefs about the accuracy of homeless counts. A main finding of the study was that the relationship between homelessness and rental costs is strongest in New York, Los Angeles, Washington, D.C., and Seattle.

Archival publications (published) during reporting period:

- [1] N. Foti and E.B. Fox (2019) Statistical model-based approaches for functional connectivity analysis of neuroimaging data. Current Opinion in Neurobiology. `https://doi.org/10.1016/j.conb.2019.01.009`

- [2] A. Tank, E.B. Fox, and A. Shojaie (2019) Identifiability and estimation of structural vector autoregressive models for subsampled and mixed-frequency time series. Biometrika. `https://doi.org/10.1093/biomet/asz007`

- [3] C. Glynn and E.B. Fox (2018) Dynamics of homelessness in urban America. Annals of Applied Statistics. `https://projecteuclid.org/euclid.aoas/1554861661`

- [4] S. Ainsworth, N. Foti, A.K.C. Lee, and E.B. Fox (2018) oi-VAE: Output Interpretable VAEs for Nonlinear Group Factor Analysis. Proc. of the International Conference on Machine Learning (ICML). `https://arxiv.org/abs/1802.06765`

- [5] N. Foti and E.B. Fox (2017) Comment: Nonparametric Bayes modeling of populations of networks. Journal of the American Statistical Association. `http://dx.doi.org/10.1080/01621459.2017.1388245`

- [6] A. Tank, I. Covert, N. Foti, A. Shojaie, and E.B. Fox (2017) An Interpretable and Sparse Neural Network Model for Nonlinear Granger Causality Discovery. NIPS Time Series Workshop. `https://arxiv.org/abs/1802.05842`

- [7] N. Foti, R. Nadkarni, and E.B. Fox (2016) Sparse plus low-rank graphical models of time series for functional connectivity in MEG. SIGKDD Workshop on Mining and Learning from Time Series. `http://www-bcf.usc.edu/~liu32/milets16/paper/MiLeTS_2016_paper_22.pdf`

- [8] Tank, E.B. Fox, and A. Shojaie (2016) Granger Causality Networks for Categorical Time Series. SIGKDD Workshop on Mining and Learning from Time Series. `http://www-bcf.usc.edu/~liu32/milets16/paper/MiLeTS_2016_paper_24.pdf`

Changes in research objectives, if any: None

Change in AFOSR program manager, if any: None

Extensions granted or milestones slipped, if any: None

Include any new discoveries, inventions, or patent disclosures during this reporting period (if none, report none): None