How a Deep Neural Network Contributes to Learning Causal Graph and Forecasting Political Dynamics

Seo Eun Yang
The Ohio State University

Motivation

- Causal discovery on times series data is crucial for theory testing and forecasting in political economy and international conflict.
- In studies of political dynamics, however, scholars have paid little attention to modeling unknown and complex (usually nonlinear) relationship between variables that may exist inside the system.
- Existing time series modeling such as Bayesian Structural Vector Autoregression (B-SVAR) are based on linear system and appropriate distribution assumptions which require informed beliefs about the dynamics of the variables.
- Given unknown complex nonlinearity in political dynamics, I would like to show how deep neural networks approximates the causal graphs on time series and minimizes forecast errors.

Research Questions

1. Without a deeper understanding of domain knowledge, can deep neural networks well identify complex causal relationships among variables?
2. Does deep neural networks model produce smaller forecast errors than its competitor, B-SVAR model?

Main Objectives

1. I aim to present the improvements and extensions of time series modeling that have been made in deep neural networks.
2. As a data-driven approach, I aim to explore causality between variables without any informed beliefs about the dynamics of the variables.
3. Applying deep learning model to political dynamics, I aim to revisit two papers: democratic accountability in UK (2010), and conflict dynamics during the 2012 Gaza Conflict (2018).

Email: yang.3176@osu.edu

Scalable Causal Graph Learning

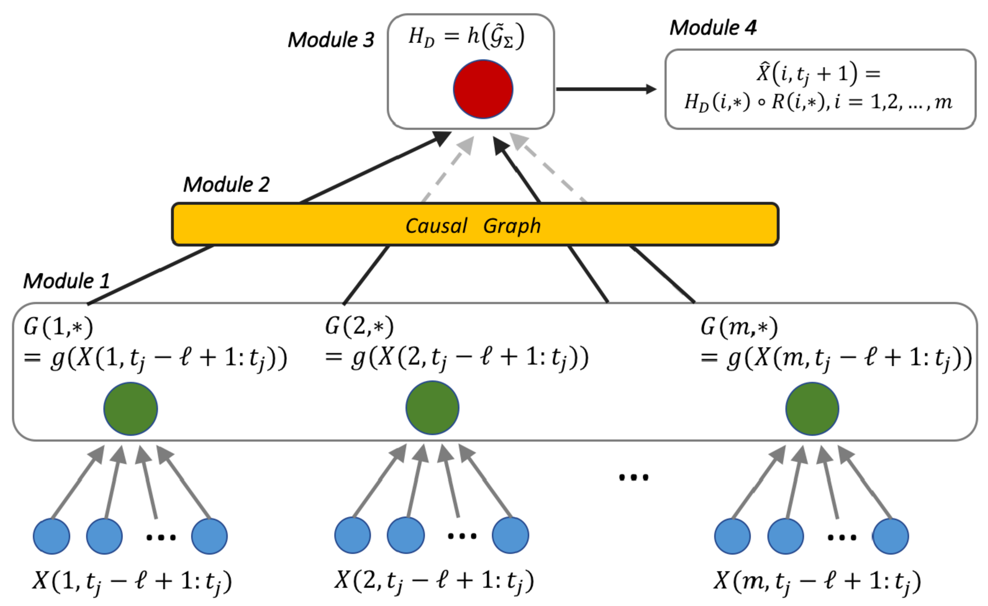


Figure 1:Overview of SCGL model (Xu et al., 2019, Figure 3)

Preliminaries

- Input: m time series with time length n.
$$X = \{X(1,*), X(2,*), ..., X(m,*)\} \in \mathbb{R}^{(m\times n)}$$
- Goal: to learn one global causal graph $A \in \mathbb{R}^{(m\times m)}$
  - $A(i,j) \gg 0$: the i-th time series is the cause of the j-th time series .
- Hyperparameters were optimized via grid search technique.

Module 1: Learning univariate nonlinearity

- A residual neural network (ResNet) is applied.
$$G_q = ReLU(\beta_q(G_{q-1}B_q) + id(G_{q-1})), q = 1, ..., Q$$
where G_0 is the raw input from each time series, and Q is the total number of residual blocks
- Final output $G_{1:Q}(i,*)$ represents different levels of univariate nonlinearity to the underlying causal graph.

Module 2: Learning causal graph

- The underlying causal graph is discovered through low-rank approximation.
- For each $G_q \in G_{1:Q}(i,*)$, the nonlinear variables relevant to regressing $X(i,*)$ are selected by a learned causal graph.
- Final output $\tilde{\mathcal{G}}_q^T(*,i) = G_q^T A_q(*,i), i = 1, ..., m$ represents linear combinations of all the univariate nonlinearity from the i-th time series $X(i,*)$.

Module 3: Learning intervariable nonlinearity

- Module 3 consists of a series of fully connected layers
$$H_j = \sigma(H_{j-1}W_j + b_j), j = 1, 2, ..., D$$
where $H_0 = \tilde{\mathcal{G}}_\Sigma^T$, and D is the total number of layers
- Final output H_D amounts to the intervariable nonlinearity set between time series.

Module 4: Approximate regression targets

- Using output H_D in Module 3, the regression target $X(i,t_j+1)$ is predicted.
$$\hat{X}(i,t_j+1) = H_D(i,*)R(i,*), i = 1, 2, ..., m$$
- $H_D(i,*)$ contains all of the possible contributing nonlinear forms from the causes of $X(i,t_j+1)$, $i = 1, ..., m$
- The residual between the predicted $\hat{X}(i,t_j+1)$ and the actual $X(i,t_j+1)$ is calculated by mean squared error (MSE) with regularization terms.

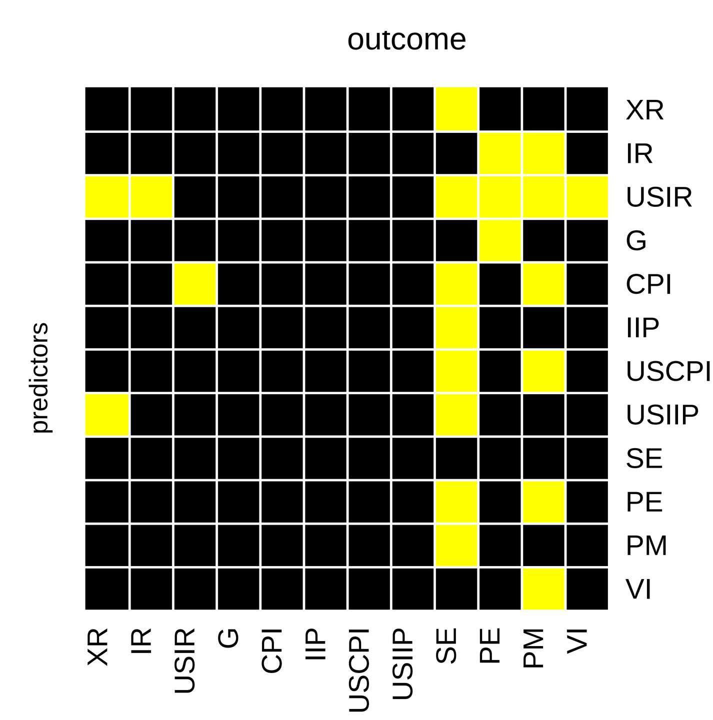
Simulation

- Purpose:
  1. Compare forecasting error from SCGL with B-SVAR
  2. Evaluate the estimated causal graph against true causal graph
- The data generating process (refer to Xu et al., p.1859)
  - 65 time series with time length 600 (train: 99%, test: 1%).
  - First, given causal graph, generate univariate nonlinearity
$$X(i,t_j) = cos(\sqrt{3}d(i)\times X(i,t_j-1)\times d(i)^2\times X(i,t_j-2)) + \epsilon(i,t_j)$$
where $d(i) \sim unif(0.95, 1)$, and $\epsilon(i,t_j) \sim N(0,1)$
  - Then, create intervariable nonlinearity
$$X(k,t_j) = \sum_{l=1}^{m} log((r(l\Rightarrow k)\times X(l,t_j - p(l\Rightarrow k)))^2) + \epsilon(k,t_j)$$
where $r(l\Rightarrow k) \sim unif(-1,1)$ if $A(l,k) = 1$ and 0 otherwise, and $p(l\Rightarrow k)$ determines the time lag of the temporal causal relationship from l to k and is randomly selected from $\{1,2,3\}$

	SCGL	B-SVAR
Forecast error (MSE), N=6	291.87	330.31
Causal Graph Prediction(AUC)	0.932	0.653

Democratic Accountability in UK

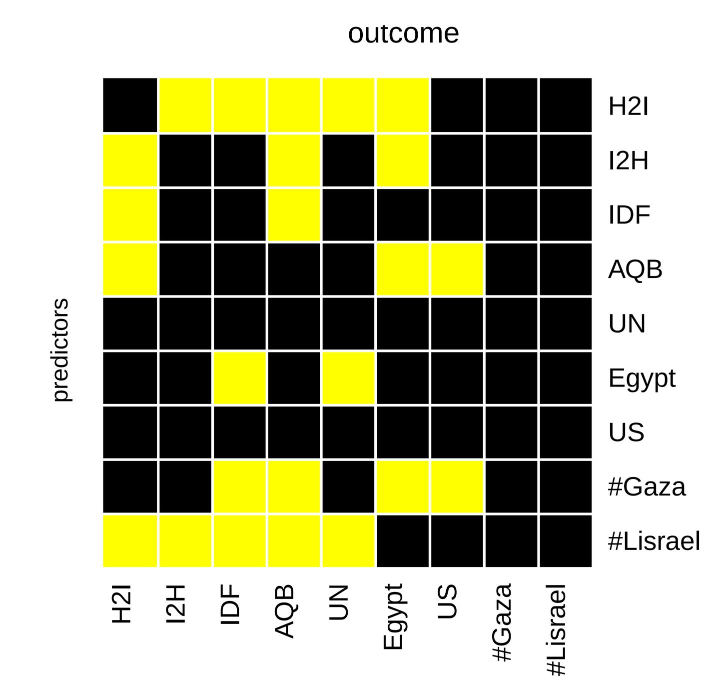
- Theoretical uncertainty still exists in endogenous relationships between the British economy, aggregate public opinion, and economic evaluations.
- Data: twelve monthly time series variables[a] were collected from 1984:4 to 1997:4 (m=12, T=157, train: 92%, test: 8%).
- Forecast error (MSE): (1) SCGL: 43, (2) B-SVAR: 45.84
- Learned causal graph:



[a]XR = exchange rate; IR = interest rate; CPI = consumer price index; IIP = index of industrial production; SE = sociotropic expectations; PE = personal expectations; PM = prime ministerial approval; VI = vote intentions

the 2012 Gaza Conflict Dynamics

- How did international audiences influence different actors during the 2012 Gaza Conflict?
- Data: nine hourly time series variables[a] with time length 179 were collected during the conflict (m=9, T=179).
- Forecast error (MSE): (1) SCGL: 94.24, (2) B-SVAR: 115.33
- Learned causal graph:



[a]$H2I$: Hamas conflict intensity, $I2H$: Israel conflict intensity, IDF: $@IDFSpokesperson$ aggressiveness on Twitter, AQB: $@AlQassamBrigade$ aggressiveness on Twitter, UN: UN attention, Egypt: Egyptian attention, US : US attention, #Gaza: $\#GazaUnderAttack$ mentions on Twitter, #Israel: $\#IsraelUnderFire$ mentions on Twitter

Conclusion

- Based on simulation and real-world dataset, SCGL model makes smaller forecast errors than B-SVAR model.
- It also well identifies complex causal relationships among variables without predefined distribution assumption.

References

- Sattler, T., Brandt, P. T., & Freeman, J. R. (2010). Democratic accountability in open economies. Quarterly Journal of Political Science, 5(1), 71-97.
- Xu, C., Huang, H., & Yoo, S. (2019, November). Scalable Causal Graph Learning through a Deep Neural Network. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management (pp. 1853-1862)
- Zeitzoff, T. (2018). Does social media influence conflict? Evidence from the 2012 Gaza Conflict. Journal of Conflict Resolution, 62(1), 29-63.