Nonlinear Causality Inference in Microarray Time Series

Fateme Nateghi Haredasht[1,2] and Mohammad Hassan Moradi[3]

[1] KU Leuven, Campus KULAK - Department of Public Health and Primary Care, Etienne Sabbelaan 53, 8500 Kortrijk, Belgium
[2] ITEC - imec and KU Leuven, Etienne Sabbelaan 51, 8500 Kortrijk, Belgium
[3] Department of Biomedical Engineering, Amirkabir University of Technology, Tehran, Iran

Abstract. Detecting causal relationships between time series data has been widely studied in many areas. One of the most popular causality inference methods is Granger causality that is a linear regression-based model for determining whether a single time series is useful in forecasting another; however, this approach cannot detect nonlinear relations in the data. In this work, we use Elastic-net regularization to infer linear Granger causalities in gene expression data. Moreover, we have proposed a method which uses the properties of kernel algorithms to infer nonlinear causalities.

Keywords: Causality inference · Granger causality · Gene regulatory networks

1 Introduction

Granger causality is a strong method for inferring causal relations among a number of time series data. According to the definition, if past values of X and Y together can predict future value of Y better that past values of Y alone, then time series X Granger-cause time series Y. One of the advantages of using Granger causality is to infer causal networks like gene regulatory networks (GRNs) from microarray time series data [1, 5, 3]. Real biological data sets have a problem which is about the big numbers of genes while we have few time steps ($n \gg T$). Due to this problem, directly applying Granger causality is not feasible. In this work, we use Elastic-net regularization [6] which has the sparsity feature and can be applied to high dimensional settings.

Moreover, in GRNs, genes inherently have nonlinear intersections with each other; however, Granger causality is a linear method which is not capable of inferring the nonlinear interactions. For nonlinear dynamics of the time series data, different extensions of Granger causality have been proposed [4, 2].

Using Elastic-net regularization, the problem of high dimensional data has been already solved; however, to infer nonlinear interactions, we need to apply a more flexible method. As a result, we propose to use kernelized Elastic-net approach which uses the properties of kernel algorithms to infer nonlinear causalities.

2 Experiments

We use Human Cancer Cell Line (HeLa) dataset for evaluation. Here, we use experiment 3 which has more than 1100 genes and 48 time points.

In Figure 1, we present a comparison between the two proposed methods, Elastic-net, and kernelized Elastic-net Granger causality.

The proposed method is capable of providing superior results in many cases. More specifically, kernelized Elastic-net provides high precision in the dataset.
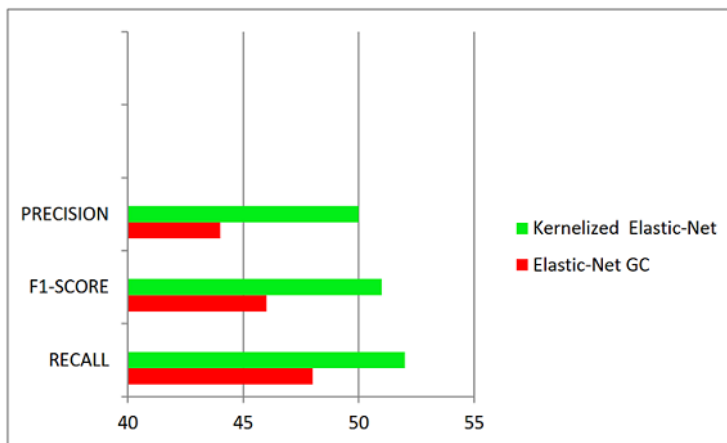

Fig. 1. Comparison between Elastic-net GC and kernelized Elastic-net GC

3   Conclusion

Inferring causal relations could lead to gene regulatory networks; however, in microarray time series the number of genes is far more than the number of time points. We have used Elastic-net regularization to have both sparse and smooth solutions. We have also proposed a method which can be applied to infer nonlinear interaction in high dimensional settings.

References

1. Geurts, P., et al.: dyngenie3: dynamical genie3 for the inference of gene networks from time series expression data. Scientific reports **8**(1), 3384 (2018)
2. Haredasht, F.N., Ghassemi, F., Moradi, M.H.: Causal inference of gene expression data using a clustering-based extension of kernel-granger causality. In: 2016 23rd Iranian Conference on Biomedical Engineering and 2016 1st International Iranian Conference on Biomedical Engineering (ICBME). pp. 84–88. IEEE (2016)
3. Lopes, M., Bontempi, G.: Experimental assessment of static and dynamic algorithms for gene regulation inference from time series expression data. Frontiers in genetics **4**, 303 (2013)
4. Marinazzo, D., Pellicoro, M., Stramaglia, S.: Kernel method for nonlinear granger causality. Physical review letters **100**(14), 144103 (2008)
5. Zhao, W., Serpedin, E., Dougherty, E.R.: Inferring gene regulatory networks from time series data using the minimum description length principle. Bioinformatics **22**(17), 2129–2135 (2006)
6. Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. Journal of the royal statistical society: series B (statistical methodology) **67**(2), 301–320 (2005)