The Sixteenth International Symposium on Artificial Life and Robotics 2011 (AROB 16th '11),
B-Con Plaza, Beppu,Oita, Japan, January 27-29, 2011

Multiple Granger causality tests for network structure estimation from time-series data

Hikaru Harima[1], Shigeyuki Oba[1,2], and Shin Ishii[1]
[1]*Graduate School of Informatics, Kyoto University*
[2]*PRESTO, Japan Science and Technology Agency*
(harima-h@sys.i.kyoto-u.ac.jp, oba@i.kyoto-u.ac.jp, ishii@i.kyoto-u.ac.jp)

Abstract: To identify network structures is a key for elucidating functions of various kinds of networks such as cortical local circuits. Granger causality (GC) test has been used for estimating directed network structure from time-course of neuronal activities. Although GC statistic for a pair of nodes can be substantially influenced by other nodes, ignoring such influence can degrade detection performance of multiple GC tests. To improve the multiple GC tests, and hence the estimation of large network structures, therefore, we propose an extension of GC by introducing optimal discovery procedure (ODP) that shows the best detection power in general multiple testing problems. Applying our proposed method to a benchmark dataset, we show the performance of estimating the network structure is improved over those by the exiting methods.

Keywords: Granger causality, network structure, time-series analysis, optimal discovery procedure

I. INTRODUCTION

To identify network structures is a key for elucidating functions of various kinds of networks; for example, a cortical local circuit would have its own function in the brain information processing, which is basically defined by the connectivity between constituent neurons. Since it is often difficult to directly observe anatomical structures especially in vivo, there have been numerous attempts to estimate network structures based on electrophysiological activities of neurons. For examples, multiple-electrode arrays [1] have been widely used, and recently, functional multi-neuron calcium imaging (fMCI) [2] comes to be available for observing neural activities with both temporally and spatially high resolutions.

Causal structure of a directed network can be uniquely represented by the set of directed causal links between pairs of network nodes; then testing whether a directed link exists or not between a single pair of nodes is the basis for estimating the whole network's causal structure. Granger causality (GC) [3] has become popularly used for estimating existence of each link; a directed causal influence from a node A to a node B is detected if time-series prediction accuracy for node B by means of an autoregressive (AR) model is significantly improved by incorporating time-series observation for node A into its explanatory variable. Although GC provides a useful tool for estimating network structure, there still remains the problem of '*multiplicity*'. As the number of elements (e.g., neurons) becomes large, the number of possible links grows huge, which makes the estimation of false positives and false negatives in the results obtained by GC tests quite difficult. A solid treatment of such multiplicity was previously proposed by Storey [4,5]; his optimal discovery procedure (ODP) is known to be the most powerful test in multiple simultaneous statistical hypothesis testing (MSHT), like the detection problem of causality links from huge number of possible links in neuronal networks.

In this study, we propose a new method to determine directed network structure based on application of ODP to GC tests. We demonstrate the performance of causal link detection of the proposed method is better than those by the existing method when applied to a benchmark datasets.

II. METHODS

1. Granger Causality

A GC test tries to identify causal relationship between two nodes by assuming an autoregressive (AR) model between them. Consider a test of a causal link from an input node Ii to an output node Oi based on their observations $\mathbf{x}^{(Ii)}$ and $\mathbf{x}^{(Oi)}$, where $\mathbf{x}^{*} = (x_1^{*},...,x_T^{*})$ denotes a time-course observed at a node *. The null hypothesis assumes the output node admits an individual AR process: then, if the null hypothesis holds, $x_t^{(Oi)}$ is predicted by

$$\hat{x}_t^{(Oi)}(\mathbf{x}^{(Oi)}) = \sum_{j=1}^{p} w_j^{(Oi:null)} x_{t-j}^{(Oi)} \quad ,$$

where p is the order of the AR process. On the other hand, the alternative hypothesis assumes causal influence from Ii to Oi. In this case, $x_t^{(Oi)}$ is predicted by a vector autoregressive (VAR) model which employs not only the past observations of Oi but also those of Ii as its explanatory variable:

$$\hat{x}_t^{(Oi)}(\mathbf{x}^{(Oi)}, \mathbf{x}^{(Ii)}) = \sum_{j=1}^{p} (w_j^{(Oi:alta)} x_{t-j}^{(Oi)} + w_j^{(Ii:alta)} x_{t-j}^{(Ii)}), \qquad (1)$$

where $w_j^{(Oi:null)}$, $w_j^{(Oi:alta)}$, and $w_j^{(Ii:alta)}$ are regression parameters. By simply assuming the prediction errors obey Gaussian distribution, the null and alternative likelihoods are given by

$$f_i(\mathbf{x}^{(Oi)}, \mathbf{x}^{(Ii)}) = \prod_{t=p+1}^{T} N(x_t^{(Oi)}; x_t^{(Oi)}(\mathbf{x}^{(Oi)}), \mu, \sigma_{i:null}^2) \quad ,$$

$$g_i(\mathbf{x}^{(Oi)}, \mathbf{x}^{(Ii)}) = \prod_{t=p+1}^{T} N(x_t^{(Oi)}; x_t^{(Oi)}(\mathbf{x}^{(Oi)}, \mathbf{x}^{(Ii)}), \mu, \sigma_{i:alta}^2) \quad ,$$

where $N(x; \mu, \sigma^2)$ denotes Gaussian probability density function of mean μ and variance σ^2. The regression parameters in these likelihood functions, $\hat{f}_i(\mathbf{x}^{(Oi)}, \mathbf{x}^{(Ii)})$ and $\hat{g}_i(\mathbf{x}^{(Oi)}, \mathbf{x}^{(Ii)})$, can be determined by maximum likelihood (ML) estimation, based on the observations. The variances in the null and alternative likelihood functions can also be determined by ML estimations, as:

$$\hat{\sigma}_{i:null}^2 = \frac{\Sigma_{t=p+1}^{T} (x_t^{(Oi)} - \hat{x}_t^{(Oi)}(\mathbf{x}^{(Oi)}))^2}{T-p},$$

$$\hat{\sigma}_{i:alta}^2 = \frac{\Sigma_{t=p+1}^{T} (x_t^{(Oi)} - \hat{x}_t^{(Oi)}(\mathbf{x}^{(Oi)}, \mathbf{x}^{(Ii)}))^2}{T-p}.$$

After obtaining the null and alternative likelihood functions, we have the log-likelihood ratio as

$$LLR_i = \frac{T-p}{2} \log \left(\frac{\hat{\sigma}_{i:null}^2}{\hat{\sigma}_{i:alta}^2} \right).$$

Since we have assumed the two likelihood functions both obey Gaussian, we can apply the standard F-test to the likelihood ratio; thus, the GC test can be used for identifying the causal relationship.

2. Optimal Discovery Procedure

The ODP defines the most powerful test in MSHT situations [4]. According to the theory of ODP, expected number of true positives (ETP) and expected number of false positives (EFP) in MSHT are considered, and a statistical test is said most powerful if any other test with smaller EFP cannot have larger ETP. Storey showed that such an ideal statistic, ODP, exists that setting any threshold on the ODP statistic becomes the most powerful test. More concretely, in MSHT, we consider m hypothesis tests at the same time, in each of which it is tested whether the null or alternative hypothesis is plausible after given observations of a single variable: $X_i, i = 1, ..., m$, based on the null and likelihood functions: $f_i(X_i)$ and $g_i(X_i)$. In this situation, the true ODP statistic (TODP) is defined by

$$S^{\mathrm{TODP}}(X_i) = \frac{\Sigma_{j=1}^{m} (1 - w_j) g_j(X_i)}{\Sigma_{j=1}^{m} w_j f_j(X_i)} \quad ,$$

where $w_i = 1$ (or $w_i = 0$) holds if the null (or alternative) hypothesis is true for the i-th test. Since $f_i(X_i)$, $g_i(X_i)$ and w_i are unknown in actual situations, however, the ODP statistic should be estimated based on observed data. Then, instead, we use estimated ODP (EODP):

$$S^{\mathrm{EODP}}(X_i) = \frac{\Sigma_{j=1}^{m} (1 - \hat{w}_j) \hat{g}_j(X_i)}{\Sigma_{j=1}^{m} \hat{w}_j \hat{f}_j(X_i)} \quad ,$$

where $\hat{w}_i$ is estimated by a conventional statistical test, and $\hat{f}_i(X_j)$ and $\hat{g}_i(X_j)$ are null and alternative likelihoods, respectively, estimated based on available observations. Let $\hat{F}_{ij} \equiv \hat{f}_i(X_j)$ and $\hat{G}_{ij} \equiv \hat{g}_i(X_j)$; we call these terms self-likelihood terms if $i = j$ and mutual-likelihood terms otherwise. The ODP statistic involves mutual-likelihood terms, and then, the detection power is improved over ordinary likelihood ratio tests which employ only the self-likelihood terms.

3. ODP for GC

Here, we apply the idea of ODP to multiple GC tests. When a directed network consists of N nodes, the number of its possible connections is as many as $M = N(N-1)$. Although the conventional application of GC tests to network structure estimation has been ignored this multiplicity, introduction of the idea of ODP would improve the detection power, because of possible correlations between constituent nodes in the network. In multiple GC test situations, the i-th test would be performed based on a set of observations of the corresponding pair of nodes. Then, the mutual-likelihood functions of null and alternative hypotheses in ODP employ observations of the two nodes as:

$$\hat{F}_{ji} = f_j(\mathbf{x}^{(Oi)}, \mathbf{x}^{(Ii)}) \quad ,$$

$$\hat{G}_{ji} = g_j(\mathbf{x}^{(Oi)}, \mathbf{x}^{(Ii)}) \quad .$$

To obtain the estimate EODP statistic, the weight parameters w_i in equation (1) is necessary; in this case, we set $\hat{w}_i = 1$ (or $\hat{w}_i = 0$) if the null (or alternative) hypothesis is accepted in the individual i-th GC test. It should be noted that this individual test is just for

defining EODP statistic, and actual MSHT is performed based on the EODP statistic, not by the individual GC test. Consequently, we obtain ODPGC (EODPGC) statistic as:

$$EODPGC_i(\mathbf{x}^{(Oi)}, \mathbf{x}^{(Ii)}) = \frac{\Sigma_j (1-\hat{w}_j)\widehat{G_{ji}}}{\Sigma_j \hat{w}_j \widehat{F_{ji}}} \quad .$$

III. SIMULATION

1. Setting

Computer simulation was performed to examine the proposal method. By assuming that true network structure consists of 5 nodes and 5 directed causal links (Fig. 1), simulated time-series were obtained by means of a VAR model of the true order being p = 3; this benchmark setting was also used in a previous study [6].

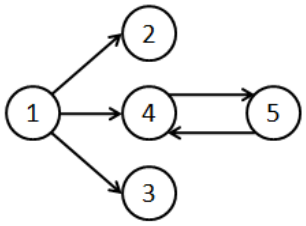

Fig.1. A five node network

In addition, we prepared larger but redundant networks which were constructed by copying the 5-node graph (Fig. 1) $m = 1,2,...,14$ times; then we have in total 14 networks with $M = 5,10,...,70$ nodes. For example, the network with 15 nodes has three identical sub-networks of five nodes and no causal link between the sub-networks. For each of the 14 networks, time-series of length T = 50 was for every node generated. This length T = 50 was set to examine the applicability of statistical tests to short time-series.

Detection accuracy of causal links was examined for the proposed method (EODPGC), ODP with true weights (TODPGC), and simple individual GC tests. The general performance of statistics was examined in terms of ROC curves. Since an ROC curve represents the behavior of the statistic within a two-dimensional plane of the false positive rate (horizontal axis) and the true positive rate (vertical axis), for each threshold setting. Because detection accuracy of the statistic depends on this threshold setting, we also examined area under the ROC curve (AUC) as a general criterion; a larger AUC means higher detection accuracy regardless of the threshold setting.

2. Result

Figure 2 shows ROC curves for the three methods in the case of p = 3 and M = 50.

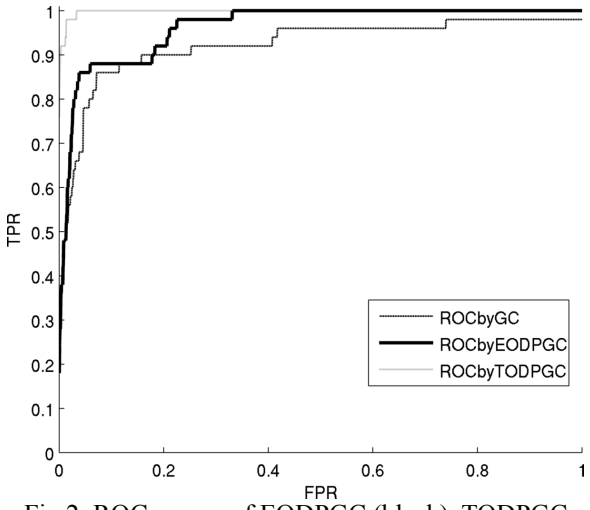

Fig.2. ROC curves of EODPGC (black), TODPGC (light gray) and GC test (gray): $(p, T, M) = (3, 50, 50)$

In this case, ODPGC with the true weight values (w) (TODPGC) achieved almost 100% accuracy with appropriate thresholds, though the true weight parameter is not available in actual applications. EODPGC, which employs the estimated weight values performed better than the GC tests; this is also confirmed by AUC scores (EODPGC: ODPGC with the estimated weight values ($\hat{w}$), GC: Granger causality).

Next, the performances were compared by varying the order of AR model p. The number of nodes M was set at 50. Both GC and EODPGC performed the best when p = 3. This result is reasonable because we used a VAR-model of the order of 3 to generate the simulation time-series data.

Next, we examined how the AUC behaves as the network size increased: $M = 5,10,...,70$ (Fig. 3). The order of AR model p was fixed at 3.

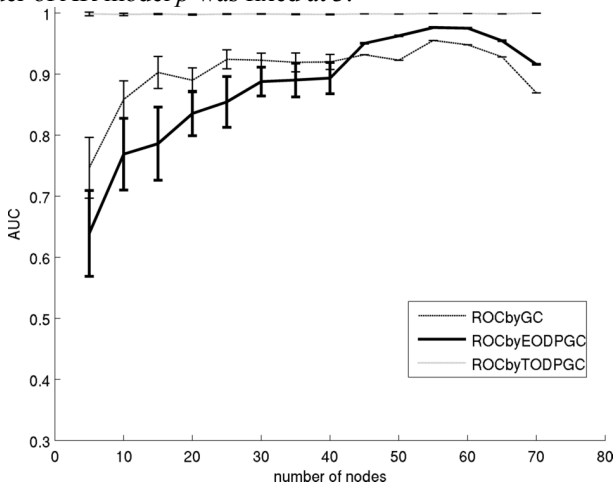

Fig.3.AUC for a various numbers of network nodes, where $(p, T) = (3, 50)$

In this figure, the mean AUC value and half of standard deviation over 20 trials are shown for $M = 5,...,40$,

and a single AUC value for $M = 45,...,70$. As the multiplicity became large due to the increase in the number of network nodes, such as $M > 40$, the performance of EODPGC was likely better than that of GC. Although individual GC tests do not consider the correlation between the tests, the correlation does exist effectively, because the connectivity between a pair of nodes is not independent of that of another pair, due to the eventual effective correlation between the nodes in the network. The statistics based on ODP show good performance especially when such effective correlation exists between the multiple tests to be performed at the same time.

IV. CONCLUSION

We proposed a powerful method to identify a large causal network structure. This method is an application of the ODP theory to multiple GC tests. Comparing our method with the standard GC tests, we showed a good performance for identifying the network structure especially when the network size is large. There would be some remaining issues; comparison with recent methods, for example, conditional granger causality (CGC) [7] and partial granger causality (PGC) [6] should be done in the near future. Application to electrophysiological data from real neuronal networks is also an important future work.

REFERENCES

[1] R. Q. Quiroga, S. Panzeri (2009), "Extracting information from neuronal populations: information theory and decoding approaches", Nature Reviews Neuroscience, 10:173-185.
[2] T. Sasaki, N. Matsuki, Y. Ikegaya (2007), "Metastability of active CA3 networks", The Journal of Neuroscience, 27:517-528.
[3] C. Granger (1969), "Investigating causal relations by econometric models and cross-spectral methods", Econometrica, 37:424-438.
[4] J. D. Storey (2007), "The optimal discovery procedure: a new approach to simultaneous significance testing", Journal of the Royal Statistical Society. Series B, Statistical methodology, 69: 347-368.
[5] J. D. Storey (2006), "The optimal discovery procedure for large-scale significance testing, with applications to comparative microarray experiments", Biostatistics (Oxford, England), 8: 414-439.
[6] S. Guo (2008), A. K. Seth, K. M. Kendrick, C. Zhou, J. Feng, "Partial Granger causality–Eliminating exogenous inputs and latent variables", Journal of Neuroscience Methods, 172: 79-93.
[7] J. Geweke (1982), "Measurement of linear dependence and feedback between multiple time series", Journal of the American Statistical Association, 77: 304-313.