Granger causality analysis of task-related fMRI time series

Patawee Prakrankamanant ID 5630349021
Advisor: Assist. Prof. Jitkomut Songsiri

Department of Electrical Engineering Faculty of Engineering, Chulalongkorn University

May 18, 2017

Overview

- Project overview
- Method
  - Granger causal model
  - Autoregressive model with exogenous input
- Experiment

Project overview (1)

Problem statement

- we want to explore brain activity by using linear model that have input terms.

Objective

- to study how to apply linear model to explain brain activity
- to develop numerical algorithm to solve model estimation problem

Scope of work

- we focus only linear model when we know value of input signal.
- we consider autoregressive model with exogenous input and apply least square estimation.
- we test model from generated data and fMRI data sets.

Project overview (2)

Expected outcomes

- The estimation formulation for estimating parameter of autoregressive model with exogenous input
- MATLAB codes for learning brain connectivity.
  can solve estimating linear model problem of brain activity

Unfinished work

- we do not have experiments with real data set.

Cause of Unfinished work

- Raw brain image scans require many preprocessing steps.
- we have to optimize code and algorithm for solving large-scale problem.

Granger causal model (1)

Granger causality (GC) : How previous data can help to improve predicting data at present [M. Eichler,2015]
Autoregressive model (AR model) : One of linear model which is widely used to explain dynamic model of brain

$$y(t) = A_1 y(t-1) + A_2 y(t-2) + \cdots + A_p y(t-p) + e(t) \qquad (1)$$

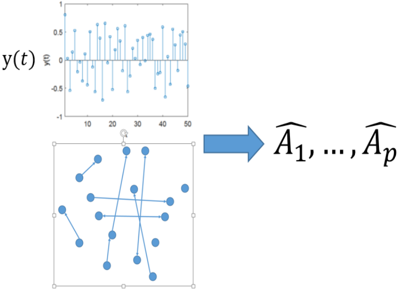
GC in AR model [M. Eichler,2015] : $y_j(t)$ is not Granger causal $y_i(t)$ if and only if

$$(A_k)_{ij} = 0 \quad k = 1, 2, \ldots, p \qquad (2)$$

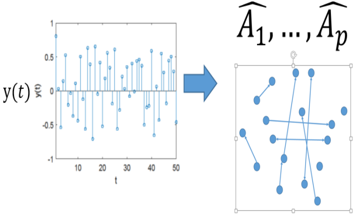
two important estimation formulations [J.Songsiri, 2013]

- estimated AR model with data given GC pattern
- explore GC pattern form times series data by using AR model

Granger causal model (2)



(a) estimated AR model with data given GC pattern



(b) explore GC pattern form times series data by using AR model

Contribution : Estimation of ARX model

AR model (1) does not have any input terms but real data sets are generated from external stimulus input. This project uses Autoregressive model with exogenous input.

$$y(t) = A_1 y(t-1) + \ldots + A_p y(t-p) + B_1 u(t-1) + \ldots + B_q u(t-q) + e(t)$$

property to use ARX model

- the ARX model is linear.
- the ARX model includes exogenous input terms.

two important estimation formulation

- given GC pattern estimated ARX model
- explore GC pattern form times series data by using ARX model

Least-square estimation of ARX model

Least squares : estimation parameter that the difference between real signal and signal generated from model is the least in term of norm-2

$$\underset{A,B}{\text{minimize}} \quad (1/2)\left(\sum_{t=p+1}^{N} \|y(t) - \sum_{j=1}^{p} A_j y(t-j) - \sum_{j=1}^{r} B_j u(t-j)\|_2^2 \right) \quad (3)$$

property of least squares

- equivalent to vector from : $\underset{x,z}{\text{minimize}} \quad (1/2)\|y_{vec} - Gx - Fz\|_2^2$ where $y(t) \rightarrow y_{vec}, A \rightarrow x, B \rightarrow z, y(t-j) \rightarrow G$ and $u(t-j) \rightarrow F$
- Solution has closed-form
- equivalent to solve linear system

Least-square estimation in ARX model with GC constraints (1)

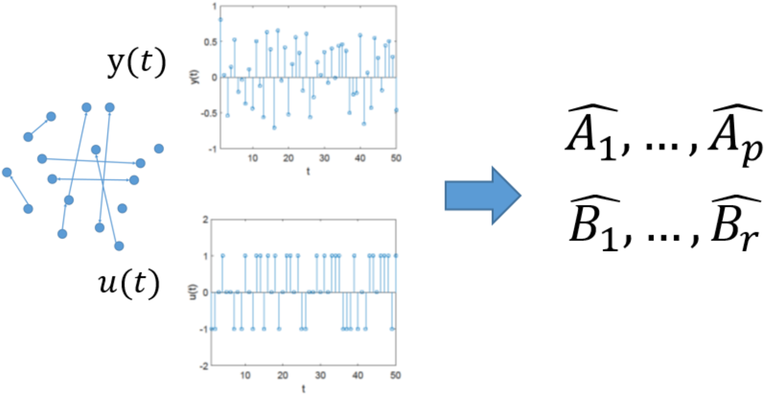


Figure: given GC pattern estimated ARX model

Least-square estimation in ARX model with GC constraints (2)

GC constraints : we know the pattern of $(A_k)_{ij} = 0, \; \forall k = 1, \ldots, p \; (i,j) \in \mathcal{J}$ or $x_k = 0 \; \forall k \in J$ in estimation problem.

$$\begin{array}{ll} \underset{x,z}{\text{minimize}} & (1/2)\|y_{vec} - Gx - Fz\|_2^2 \\ \text{subject to} & x_k = 0 \quad (k \in J) \end{array} \tag{4}$$

property of this estimation problem

- solving by eliminating constraints which reduces the problem to unconstrained Least squares problem
- still easy to solve

Learning GC pattern of ARX model (1)

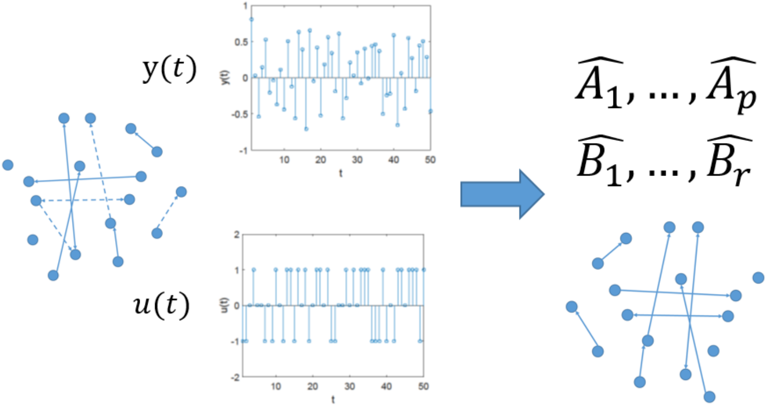


Figure: explore GC pattern form times series data by using ARX model

Learning GC pattern of ARX model (2)

Goal : find a good zero pattern in AR coefficients that best evaluate data.
we added $\|Px\|_{2,1}$ to the cost objective to promote group sparsity in X.

$$\begin{array}{ll} \underset{x,z}{\text{minimize}} & (1/2)\|y_{vec} - Gx - Fz\|_2^2 + \lambda\|Px\|_{2,1} \\ \text{subject to} & x_k = 0 \quad (k \in J) \end{array} \tag{5}$$

reason of adding $\|Px\|_{2,1}$ to estimation problem

- to force $x_k = 0 \; \exists k \notin J$ [A. Barbero, 2013]
- when λ is bigger, $\|Px\|_{2,1}$ will convert to zero.
- we have other's work that using this method [J.Songsiri, 2013]

Challenges in solving (5)

problem (5) is very hard to solve because

- the objective function is not differentiable.
- real data have a large number of variables ($\geq$ 400 millions) is very large (more than 400 millions variables).

we solved problem (5) by using Alternating Direction Method of Multipliers (ADMM) because

- ADMM use less memory storage (request only gradient).
- ADMM is fast converge algorithm ex. $q = 15, p = 2, r = 1$ and $N = 1000$ (number of variable = 465, number of data = 15000) , If we use Intel core-i7-6700HQ 2.60 GHz, Average time of ADMM is 1.041 s.
- suitable for problem (5)

we have provided an analysis of λ_c which is the minimum value of that correspond to zero solution.

Experiment

Goal : illustrate the benefit of proposed formulations

- experiment 1 : To show the ARX model outperforms AR model
  - more accuracy GC patterns
  - less model error in estimated AR parameter
- experiment 2 : To show the estimated effect form input of estimation from inputs. Dense input is more suitable than sparse input due to the principal of persistent excitation.
  - less model error in estimated parameter
  - more accuracy GC patterns
  - more accuracy selected model order

Experiment : AR model VS ARX model

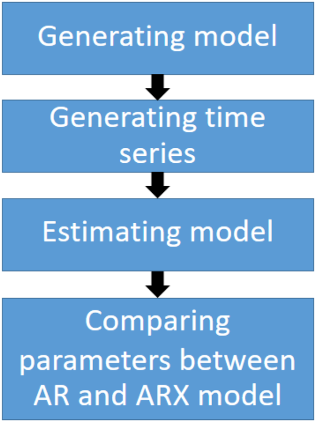
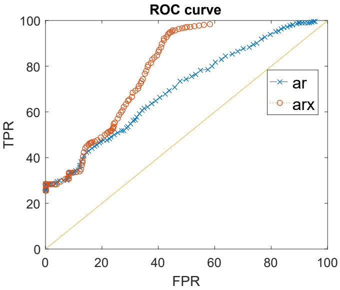
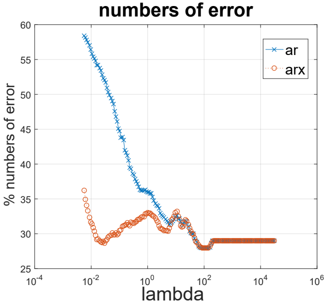


Figure: Process in "AR model vs ARX model" experiment

- granger causality pattern
  - Indicator : Granger causality pattern in AR and ARX model By using (5) with $\lambda \in [0, \lambda_c]$
- model error : auto regressive part
  - Indicator : $\|A - \hat{A}\|_F^2$ in AR and ARX model By using (5) with $\lambda \in [0, \lambda_c]$



(a) ROC curve of GC pattern of AR and ARX model



(b) numbers of fault GC pattern in AR and ARX model

In ROC curve, the upper left corner of ARX model are overtop the upper left corner of AR model and minimum numbers of fault GC pattern which is in ARX model is less than AR model. So, ARX model can detect GC pattern better than AR model.

Experiment : AR VS ARX (model error : auto regressive part)

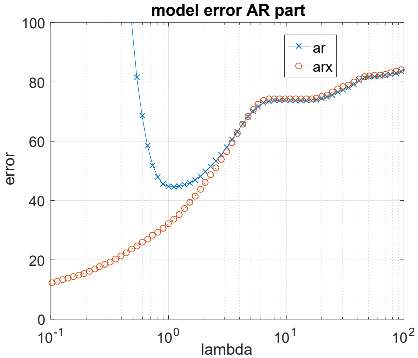
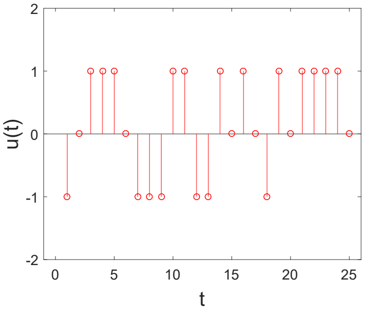


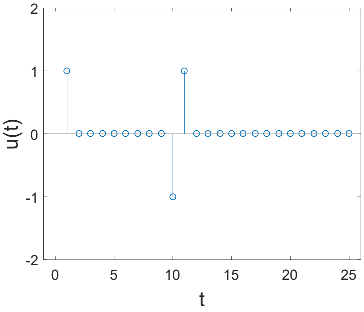
Figure: estimation error of AR parameters in AR and ARX model by using (5)

minimum of estimation error in ARX model is less than AR model.

dense input : persistent excitation is high
sparse input : persistent excitation is low



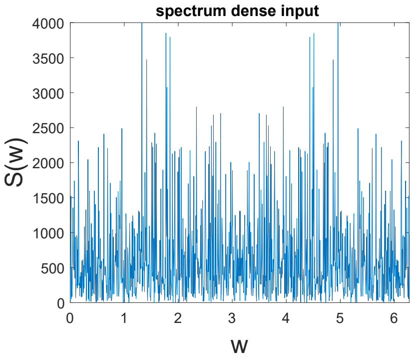
(a) dense input



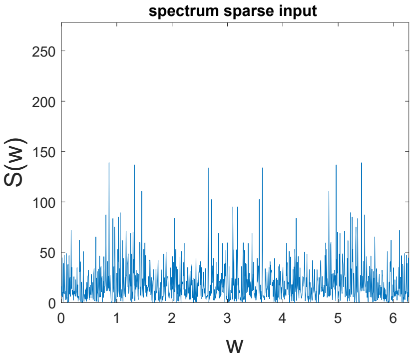
(b) sparse input

Figure: example of dense input and sparse input

Experiment : Sparse input vs Dense input (2)



(a) spectrum of dense input



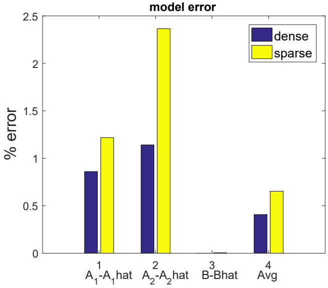
(b) spectrum of dense input

Figure: spectrum of input : Dense input have excitation more frequencies than sparse input

Experiment : Sparse input vs Dense input (3)

- model error
  - Indicator : $\|A - \hat{A}\|_F^2$ and $\|B - \hat{B}\|_F^2$ in ARX model By using (4)
- Granger causality pattern
  - Indicator : Granger causality pattern in ARX model by using (5) with $\lambda \in [0, \lambda_c]$
- model selection
  - Indicator : BIC score by using (4)

Experiment : Sparse input vs Dense input (model error)



(a) estimation error in AR part

(b) estimation error in B part

If data is generated form dense input, parameter estimation is better than data generated form sparse input.

Experiment : Sparse input vs Dense input (granger causality pattern)

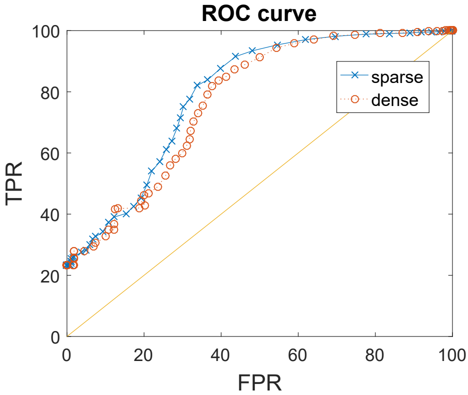


Figure: ROC curve of data generated form dense input and sparse input

Data generated from sparse input and dense input can be used to estimate GC pattern nearly

Experiment : Sparse input vs Dense input (granger causality pattern) (2)

$$y(t) = A_1 y(t-1) + \ldots + A_p y(t-p) + B_1 u(t-1) + \ldots + B_q u(t-q) + e(t) \quad (6)$$

From (6), the generated A_k parameter are sparse (density 40%). Dynamic in $y(t)$ could not change very much. But dense input affects on the fluctuation in $y(t)$ very much which effect on A_k estimation that $\hat{A}_k$ could be more density. The GC pattern of data generated by dense input may be more incorrect than data generated by sparse input.

Experiment : Sparse input vs Dense input (model selection)

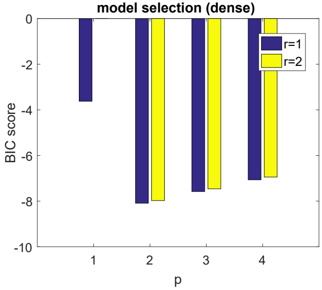


Figure: example of BIC score in experiment (p is lagging order of AR and r is lagging order of input.)

True model order : lagging order of AR is 2 and lagging order of input is 1 ($p = 2, r = 1$).

In 30 times, the accuracy of $\hat{p}, \hat{r}$ from data is generated form dense input is 100 %.

The accuracy of $\hat{p}, \hat{r}$ from data is generated form dense input is 100 %.

Both of data can select order of model well.

conclusion

- If signal is stimulated by external input,using ARX model to learn signal is more suitable than AR model
- If persistent excitation input is high, the model estimation performance is better.
- The estimation problem can be solved in large scale by the selected ADMM algorithm.

Q & A

Back up (Alternating Direction Method of Multipliers (1))

$$\begin{array}{ll}\underset{x,y}{\text{minimize}} & (1/2)\|y - Gx - Fz\|_2^2 + \lambda\|Px\|_{2,1} \\ \text{subject to} & x_k = 0 \quad (k \in J)\end{array}$$

we can create the augmented Lagrangian which is

$$\mathcal{L}_\rho = f(x_1, x_2) + h(x_3) + u^T\left(Px_1 - x_3\right) + \frac{\rho}{2}\left\|Px_1 - x_3\right\|_2^2$$

we define

$$\begin{aligned} f(x,z) &= (1/2)\|y - Gx - Fz\|_2^2 = (1/2)\left\|y - \begin{bmatrix} G & F \end{bmatrix}\begin{bmatrix} x \\ z \end{bmatrix}\right\|_2^2 \\ h(x) &= \lambda\|x\|_{2,1} \end{aligned}$$

Back up (Alternating Direction Method of Multipliers (2))

$$
\begin{array}{ll}
\underset{x_1, x_2}{\text{minimize}} & f(x_1, x_2) + h(x_3) \\
\text{subject to} & Px_1 - x_3 = 0
\end{array}
$$

$$
\begin{aligned}
(x_1^+, x_2^+) &= \underset{x_1, x_2}{\operatorname{argmin}}(f(x_1, x_2) + \mu^T(Px_1 - x_3) + \frac{\rho}{2}\|Px_1 - x_3\|_2^2) \\
x_3^+ &= \underset{x_3}{\operatorname{argmin}}\left(h(x_3) + \mu^T(Px_1^+ - x_3) + \frac{\rho}{2}\|Px_1^+ - x_3\|_2^2\right) \\
u^+ &= u + \rho\left(Px_1^+ - x_3^+\right)
\end{aligned}
$$

Back up (Alternating Direction Method of Multipliers (3))

$$\|r^k\|_2 \leq \epsilon^{\texttt{pri}} \qquad \|s^k\|_2 \leq \epsilon^{\texttt{dual}}$$

$$\left(\begin{bmatrix} G^T G & G^T F \\ F^T G & F^T F \end{bmatrix} + \rho \begin{bmatrix} P^T P & 0 \\ 0 & 0 \end{bmatrix} \right) \begin{bmatrix} x_1^+ \\ x_2^+ \end{bmatrix} = \begin{bmatrix} G^T \\ F^T \end{bmatrix} y + \begin{bmatrix} P^T \\ 0 \end{bmatrix} (\rho x_3 - u) \quad (7)$$

Back up (Alternating Direction Method of Multipliers (4))

$$
\begin{aligned}
x_3^+ &= \underset{x_3}{\operatorname{argmin}}\Big(\lambda\|x_3\|_{2,1} + u^T(Px_1^+ - x_3) + \frac{\rho}{2}\|Px_1^+ - x_3\|_2^2\Big) \\
&= \underset{x_3}{\operatorname{argmin}}\Big(\lambda\|x_3\|_{2,1} + \frac{\rho}{2}\|Px_1^+ - x_3 + u/\rho\|_2^2\Big) \\
&= \underset{x_3}{\operatorname{argmin}}\Big(\frac{\lambda}{\rho}\|x_3\|_{2,1} + \frac{1}{2}\|Px_1^+ - x_3 + u/\rho\|_2^2\Big)
\end{aligned} \tag{8}
$$

we can separate problem (8) to

$$
\underset{x_i}{\operatorname{minimize}} \quad \gamma\|x_i\|_2 + (1/2)\|x_i - w_i\|_2^2 \tag{9}
$$

The close-form solution of problem (9) is

$$
x_i^+ = \max\left\{1 - \frac{\gamma}{\|w_i\|_2}, 0\right\} w_i
$$