EC-GAN: Inferring Brain Effective Connectivity via Generative Adversarial Networks

Jinduo Liu,[1] **Junzhong Ji,**[1*] **Guangxu Xun,**[2] **Liuyi Yao,**[3] **Mengdi Huai,**[2] **Aidong Zhang**[2]
[1]Beijing Key Laboratory of Multimedia and Intelligent Software Technology, Beijing Artificial Intelligence Institute, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China
[2]Department of Computer Science, University of Virginia, Charlottesville, Virginia 22904, USA
[3]Department of Computer Science and Engineering, SUNY at Buffalo, Buffalo, New York 14260, USA
[1]liujinduo0607@emails.bjut.edu.cn, jjz01@bjut.edu.cn, [2]{gx5bt, mh6ck, aidong}@virginia.edu, [3]liuyiyao@buffalo.edu

Abstract

Inferring effective connectivity between different brain regions from functional magnetic resonance imaging (fMRI) data is an important advanced study in neuroinformatics in recent years. However, current methods have limited usage in effective connectivity studies due to the high noise and small sample size of fMRI data. In this paper, we propose a novel framework for inferring effective connectivity based on generative adversarial networks (GAN), named as EC-GAN. The proposed framework EC-GAN infers effective connectivity via an adversarial process, in which we simultaneously train two models: a generator and a discriminator. The generator consists of a set of effective connectivity generators based on structural equation models which can generate the fMRI time series of each brain region via effective connectivity. Meanwhile, the discriminator is employed to distinguish between the joint distributions of the real and generated fMRI time series. Experimental results on simulated data show that EC-GAN can better infer effective connectivity compared to other state-of-the-art methods. The real-world experiments indicate that EC-GAN can provide a new and reliable perspective analyzing the effective connectivity of fMRI data.

Introduction

Brain effective connectivity (EC), defined as the neural influence that one brain region exerts over another (Friston 1994), is important for the assessment of normal brain function (Hein et al. 2016), and its impairment is associated with neurodegenerative diseases, e.g., Alzheimer's disease (AD) (Rytsar et al. 2011). Naturally, inferring brain effective connectivity can be considered as a problem of searching or constructing a directed graph structure (causal discovery) in the human brain. Due to the critical impact on brain research and disease diagnosis, the study of brain EC networks has become a frontier subject.

Machine learning and data mining methods have enormous potential in effective connectivity network construction, because effective connectivity network construction is similar to graph construction. More precisely, inferring effective connectivity can be represented as a problem of constructing a directed graph structure from neuroimaging data, e.g., functional magnetic resonance imaging (fMRI) data. In other words, a brain effective connectivity network is a causal graph (directed graph) where nodes denote brain regions, the directed arcs denote effective connectivity between brain regions.

In the last decade, there has been a growing interest in the use of machine learning and data mining methods for brain effective connectivity network construction (Shimizu et al. 2006; Seth, Barrett, and Barnett 2015; Wang et al. 2017; Sanchez-Romero et al. 2019a; Liu et al. 2019). However, these methods have their own limitations and cannot accurately infer effective connectivity in some cases due to the characteristic of fMRI data (Smith et al. 2011). For instance, Linear non-Gaussian acyclic model (LiNGAM) uses temporal Independent component analysis (ICA) to infer effective connectivity from data. However, ICA requires a large number of data points, so it performs poorly when the fMRI data sample is small (Shimizu et al. 2006). Besides, LiNGAM also needs some prior assumptions on data generation and data disturbance. Granger causality (GC) methods infer effective connectivity in fMRI time series by the multiple regression of time-indexed variables on lagged values of variables and require the time series to be wide-sense stationary and have a zero mean (Seth, Barrett, and Barnett 2015). Bayesian network (BN) methods search for effective connectivity under the assumption that the true effective connectivity network forms a directed acyclic graph (DAG), which entails that there are no cycles in the networks. Because BN is represented by a DAG, it is not possible to model cyclic or bidirectional connections of the effective connectivity network with a BN method (Ramsey et al. 2010; Smith et al. 2011; Meek 2013; Liu et al. 2016; Zhou et al. 2016; Liu et al. 2019). Thus according to the related works, cyclic or bidirectional structure, non-stationary information, and small samples are the main factors that affect the performance of current methods. Therefore, it is necessary to develop novel methods for inferring effective connectivity networks from fMRI time series data, which can overcome the above problems.

Recently, generative adversarial networks (GANs) (Goodfellow et al. 2014) have demonstrated impres-

*Corresponding Author (jjz01@bjut.edu.cn).

sive performance for unsupervised learning tasks, and have been widely successful in several applications including image generation, image synthesis, image-to-image translation, time series imputation, and causal discovery (Kocaoglu et al. 2017; Kalainathan et al. 2018; Wang et al. 2018; Li et al. 2019). Motivated by the immense success of GANs in generating small sample and high noise simulated data that highly resembles real-world samples, we present a novel GAN-based framework to infer effective connectivity from fMRI data, named as EC-GAN. Similar to the original GANs, our framework is also based on a two-player game that involves a generator and a discriminator. Naturally, the generator we used consists of a lot of generated models (the number of the generated models is equal to the number of brain regions), and each generated model employs effective connectivity generators based on structural equation model (SEM) (Bühlmann, Peters, and Ernest 2014) to generate the fMRI time series of each brain region. The discriminator is employed to measure the difference between generated fMRI data and real fMRI data, and the generator aims to produce fMRI data indistinguishable by the discriminator from another given real fMRI data. When the generated fMRI data is similar to real fMRI data, then we can get the effective connectivity from the causal parameters of the effective connectivity generators. Thus the key difference between our framework and other GAN-based methods is that the goal of the framework is to infer the effective connectivity during the process of generating the fMRI data. The experimental results on both simulated fMRI data and real-world fMRI data show that our framework is more effective for inferring effective connectivity.

In a nutshell, the main contributions of our paper can be summarized as follows:

- To the best of our knowledge, the proposed EC-GAN is the first work that applies generative adversarial networks to the effective connectivity analysis of fMRI data, and can accurately infer brain effective connectivity from fMRI data.
- The framework of EC-GAN employs GANs with SEM to accurately infer the effective connectivity, which can model cyclic or bidirectional connections of the effective connectivity networks, and has no restrictive assumption on the underlying causal mechanisms and data distributions.
- Systematic experiments have been conducted to verify the proposed GAN-based framework (EC-GAN). The experimental results show that EC-GAN achieves better performance compared with other state-of-the-art methods.

Notation and Problem Formulation

We first give the notation in our paper, and then develop a mathematical definition of "effective connectivity".

In this paper, we employ capital letters, i.e., X_i, X_j to represent nodes (brain regions), and the bold letters $\mathbf{X}_i$ to indicate the time series of the brain region X_i. $PA(X_i)$ represents the parent nodes of brain region X_i. $X_i \perp\!\!\!\perp X_j$ and $X_i \not\perp\!\!\!\perp X_j$ represent independence and dependence between the two corresponding brain regions X_i and X_j, respectively.

According to the definition of effective connectivity (the neural influence that one brain region exerts over another), we can use a directed arc to model the effective connectivity and then employ a directed graph (causal graph) to model the effective connectivity network. Let $\mathcal{G}$ denote a directed graph, $\mathcal{D}$ denote the fMRI data set, and $\mathcal{P}$ denote the distribution of the data. Therefore, a brain effective connectivity network can be expressed as a directed graph $\mathcal{G} =< \mathbf{V}, \mathbf{E} >$, where $\mathbf{V}$ is a set of nodes with each node $X_i \in \mathbf{V}$ representing a brain region or region of interest (ROI); and $\mathbf{E}$ is a set of arcs with each arc $X_i \rightarrow X_j \in \mathbf{E}$ describing an effective connectivity from brain regions (ROIs) X_i to X_j.

Inferring brain effective connectivity via Generative Adversarial Networks

In this section, we present our adversarial framework, i.e., EC-GAN for inferring effective connectivity from fMRI data. Different from other GAN-based frameworks, our goal is to infer the effective connectivity during the process of generating fMRI data. We first give an overview of the proposed EC-GAN, then describe the details of the main components.

EC-GAN Architecture

Our proposed framework EC-GAN is made up of a generator (G) and a discriminator (D). The generator takes noise variable and real fMRI time series data as input, and generates the samples which are similar to the realistic fMRI time series data. The discriminator takes the real fMRI time series data and samples generated by the generator as inputs and tries to find a mapping that tells us the input data's probability of being real.

When we design the detailed structure of the EC-GAN, we first utilize a generator to generate time series data of all brain regions at once, however, it does not perform well. Therefore, we adopt a set of effective connectivity generators as a generator to generate samples (the number of effective connectivity generators is the same as the number of brain regions). Each effective connectivity generator is employed to generate the fMRI time series of one brain region based on the causal parameters between one brain region and another brain region. If the samples generated by the generator are very similar to the real input data, then we can get the effective connectivity from the causal parameters. The effective connectivity generator is designed based on the SEM model. We employ the SEM model because SEM has shown a strong estimation ability in identifying effective connectivity. The value of noise variable ε that we used as input is drawn from $\mathcal{N}(0, 1)$. The discriminators adopted are feed-forward fully connected neural networks with dropout, and all the effective connectivity generators share with one discriminator. The objective of the discriminators is to distinguish the real fMRI data and the generated data.

We use the following example to further illustrate the structure of the EC-GAN: suppose we want to infer the effective connectivity of five brain regions (ROIs) $X_i(i =$

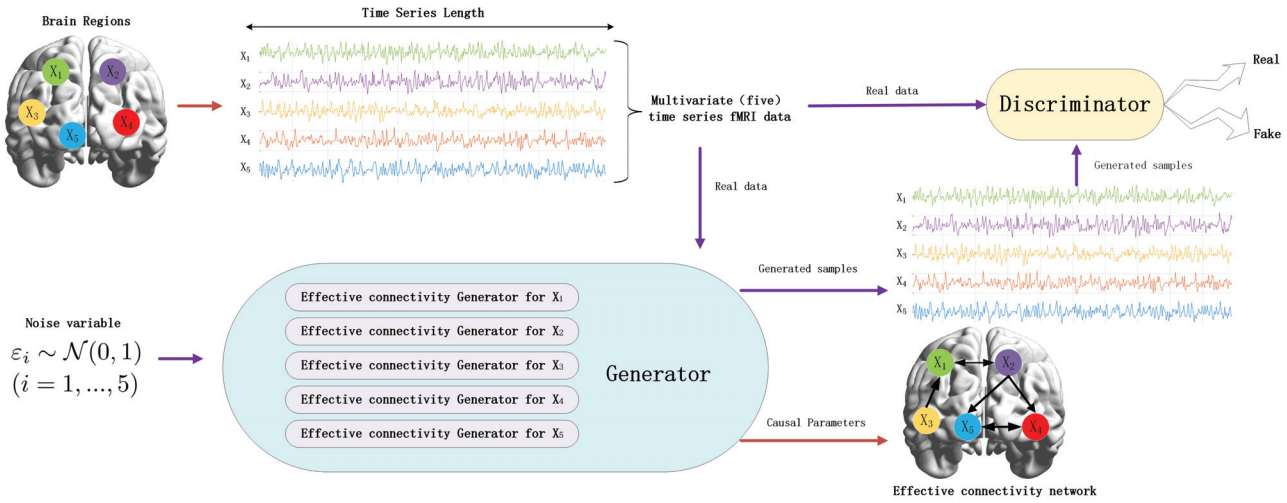

Figure 1: An example of inferring effective connectivity with EC-GAN. The input of the framework is the fMRI time series of the corresponding five brain regions, and the output of the framework is the effective connectivity networks (parameters in the effective connectivity generators). The generator consists of five effective connectivity generators, and all the effective connectivity generators share one discriminator.

$1, ..., 5$) from fMRI data. We first process the original fMRI data and get the time series of the corresponding five brain regions (ROIs). Then we need to employ five effective connectivity generators to generate each variable X_i. When the model is well-trained, and the generated samples $\mathbf{X}_{gen}$ of each brain region are very similar to real data $\mathbf{X}_{real}$. We can get the causal parameters from the effective connectivity generators. Finally, we can infer the effective connectivity network by the causal parameters. Figure. 1 shows the example of EC-GAN for inferring an effective connectivity network with five brain regions.

Effective Connectivity Generators based on Structural Equation Models

To infer effective connectivity and learn the distribution and characteristic of the original fMRI time series data, we develop a set of effective connectivity generators as a generator based on the SEM. Suppose we have n nodes (brain regions) X_i $(i = 1, ..., n)$. We can use a SEM model to present each node as:

$$X_i = f_i(PA(X_i), \varepsilon_i), \quad for\ i = 1, ..., n, \tag{1}$$

where $PA(X_i)$ denotes the set of parents for node X_i, f_i is a function from $\mathbb{R}^{|PA(X_i)|+1} \to \mathbb{R}$, and $\varepsilon_1, ..., \varepsilon_n$ are random noise in nodes X_i $(i = 1, ..., n)$, which are mutually independent. Thus a SEM is specified by a causal structure in terms of a directed graph $\mathcal{G}$, with the functions $f_i(\cdot)$ and the noise variable of the distribution of ε_i $(i = 1, ..., n)$. More specifically, if ε_i follows Gaussian distribution, we can rewrite the SEM model as:

$$X_i = \sum_{j \in PA(X_i)} f_{ij}(X_j) + \varepsilon_i, \quad for\ i = 1, ..., n, \tag{2}$$

where $\varepsilon_i \sim \mathcal{N}(0, \sigma_i^2)$, $\sigma_i^2 > 0$, and the function f_{ij} is the causal relationship between node X_i and its parents nodes (smooth function from $\mathbb{R} \to \mathbb{R}$). Thus $f_{ij}(\cdot) \neq 0$ if there is a directed arcs $X_j \to X_i$ in $\mathcal{G}$. For data $\mathcal{D}$ with n nodes, we can use a parameter set θ to represent the directed graph (causal relationships) of all nodes, that is,

$$\theta^{\mathcal{D}} = (f_{12}, f_{13}, ..., f_{1n}, f_{21}, ..., f_{2n}, ..., f_{n1}, ..., f_{n(n-1)}). \tag{3}$$

Furthermore, the n brain regions can be represented as a set of time series $\mathbf{X}_i = (x_{i1}, x_{i2}, ..., x_{it})$ $(i = 1, ..., n)$, where t is the length of the time series $\mathbf{X}_i$. Thus data $\mathcal{D}$ with n brain regions can be represented as:

$$\begin{aligned} \mathcal{D} =& \{\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n\} \\ =& (\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n)^\top \in \mathbb{R}^{t \times n}. \end{aligned} \tag{4}$$

Given data $\mathcal{D}$ and a SEM model, if we have all $f(\cdot)$ $(\theta^{\mathcal{D}})$ and the noise variable ε, we can estimate the time series of each brain region:

$$\begin{aligned} \hat{\mathbf{X}}_i =& \sum_{j \in PA(X_i)} \hat{f_{ij}}(\mathbf{X}_j) + \varepsilon_i, \\ =& \sum_{j=1}^{n} A_{ij} \mathbf{X}_j + \varepsilon_i, \end{aligned} \tag{5}$$

where A_{ij} is the causal parameter from brain region X_j to X_i. In particular, $A_{ij} = 0$ means $X_j \perp\!\!\!\perp X_i$ and there is no directed arcs from X_j to X_i (i.e., $X_j \nrightarrow X_i$). Besides, we do not consider the effective connectivity of a brain region itself, thus if $i = j$ $A_{ij} = 0$. Naturally, the effective connectivity between two brain regions, e.g., X_i and X_j, can be inferred by the decision of the causal parameters A_{ij} and A_{ji}.

Then we develop a set of effective connectivity generators, which can generate the synthetic fMRI data by Eq.(5). In particular, the number of effective connectivity generators is the same as the number of brain regions. To generate

one brain region's synthetic fMRI time series data, the effective connectivity generators take three inputs, which are the causal parameter A, the other brain regions' fMRI time series data from real data $\mathcal{D}$, and the $\varepsilon \sim \mathcal{N}(0,1)$. As the SEM is a linear model, it may ignore the nonlinear characteristics of fMRI data. We add multilayer neural networks with tanh as the activation function into Eq.(5). To sum up, the effective connective generator (for brain region X_i) consists of causal parameter A_{ij} ($j \in PA(X_i)$) and ε_i followed by a fully connected multilayer neural network with a tanh activation.

When the effective connectivity generators are well-trained and the generated data is similar to the real fMRI data, the causal parameter A can reflect the causal relationships between the brain regions. In detail, if A_{ij} is close to zero ($A_{ij} \approx 0$), it means there is no effective connectivity from brain region X_j to X_i. If A_{ij} is larger than the threshold, which is determined by the maximum number of parents nodes $MaxP$, there should be an effective connectivity $X_j \rightarrow X_i$. Since the brain effective connectivity networks are not necessarily directed acyclic graphs (DAGs), we do not impose acyclic constraints on the effective connectivity generators. Therefore, if both A_{ij} and A_{ji} are larger than the threshold, we think that there should be a bidirectional connection (effective connectivity) in the network, i.e., $X_i \leftrightarrow X_j$.

EC-GAN Loss Function

To overcome overfitting and infer sparse effective connectivity networks, we present a new loss function by adding a sparsity penalty to the effective connectivity generators. First we define the sparsity penalty as:

$$L_p = \frac{\lambda}{2} \log t \, \|A\| \,, \tag{6}$$

where t is the length of time series, λ is the hyper-parameter that controls the sparsity, and $\|A\|$ denotes the network complexity, that is, the sum of causal parameters A for the effective connectivity network and is calculated as:

$$\|A\| = \sum_{i=1,j=1}^{n} A_{ij}, \tag{7}$$

where n is the number of nodes (brain regions). In particular, if $i = j$, we set $A_{ij} = 0$ (not considering the effective connectivity for one brain region itself), thus the trace of the matrix is zero, $tr(A) = 0$.

Finally, EC-GAN Loss Function is defined as:

$$\begin{aligned} \min_G \max_D V(G,D) = \; & \mathbb{E}_{\mathbf{X} \sim \mathcal{P}_{\mathcal{D}}(\mathbf{X})}[\log D(\mathbf{X})] \\ & + \sum_{i=1}^{n} \mathbb{E}_{\tilde{\mathbf{X}}_i \sim \mathcal{P}_{\mathcal{D}}(\tilde{\mathbf{X}}_i),\, \varepsilon \sim \mathcal{P}(\varepsilon)}[\log(1 - D(G_i(\tilde{\mathbf{X}}_i, \varepsilon)))] \\ & + L_p, \end{aligned} \tag{8}$$

where n is the number of brain regions, and $\tilde{\mathbf{X}}_i$ is the subset of real fMRI time series set $\mathbf{X}$ without $\mathbf{X}_i$ ($\tilde{\mathbf{X}}_i = \mathbf{X} \backslash \mathbf{X}_i$).

Inferring Effective Connectivity by EC-GAN

When using EC-GAN for inferring effective connectivity in practice, we first need to determine the hyperparameters of the network structure (the number of units for the hidden layer, the learning rate of generator and discriminator, and the regularization coefficient). Then we can use the fMRI time series data with known ground-truth of the effective connectivity network to select the hyperparameters with the best performance.

After the hyperparameters are determined, we can use the EC-GAN to infer effective connectivity from fMRI time series data. Algorithm 1 shows the full details of the proposed EC-GAN.

Algorithm 1: EC-GAN: Inferring effective connectivity with Generative Adversarial Networks

Input: fMRI time series data.
Output: Effective connectivity network.
Initialization:
Initialize Generator G and Discriminator D;
Initialize causal parameters A;
Set Maximum number of parent nodes $MaxP$;
Model Training:
for *number of training iterations* **do**
  for *each brain region* X_i **do**
    Training Discriminator:
    Use the current G_i to generate k negative samples of brain region X_i;
    Get k positive samples from input data;
    Train the Discriminator D using the k pairs of samples;
    Update the discriminator by ascending its stochastic gradient.
    Training Generator:
    Generate a sample of brain region X_i by G_i based on causal parameters A.
    Update the generator by descending its stochastic gradient.
  end
end
Return: Effective connectivity network (causal parameters A).

Experimental Setup

To assess the performance of EC-GAN, we first use a common evaluation method, which is to test the proposed method and other comparison methods on some simulated fMRI datasets generated from known ground-truth networks. And then, to illustrate the application potential of EC-GAN, we apply it to two real fMRI datasets, i.e., resting-state fMRI dataset and task fMRI dataset.

Data Description

Benchmark Simulation Dataset The benchmark simulation datasets we used are generated by (Smith et al. 2011) and (Sanchez-Romero et al. 2019a), which are widely used

for detecting methods' performance on inferring effective connectivity. In our experiments, we aim to test the different abilities of methods, i.e., the performance of methods on the data with small samples data (run on every single subject), or with non-stationary connection strengths, or with bidirectional structural networks.

The Smith simulated fMRI data are generated based on the dynamic causal modeling (DCM), where the regions of interests (ROIs) are nodes embedded in a directed network. In our experiments, we employ the Simulation 22 in the Smith datasets. We select this simulation dataset mainly because this dataset contains non-stationary connection strengths, and almost all methods perform poorly in this dataset (Ramsey, Sanchez-Romero, and Glymour 2014). Smith datasets were obtained from (Smith et al. 2011).

We chose Simulation 1 in Sanchez simulated dataset because we want to detect the performance of different methods on inferring the bidirectional structure of effective connectivity networks. The ground-truth networks of the Sanchez simulated fMRI data contain different bidirectional structures with different degrees of complexity (Sanchez-Romero et al. 2019a). The detailed description of the two datasets we used is shown in Table 1.

Table 1: Description of the benchmark simulation data.

Dataset	Nodes	TR (s)	Data points	Subjects
Smith	5	3.00	200	50
Sanchez	5	1.20	500	60

Real Resting-state fMRI Dataset The real resting-state fMRI dataset used in this paper are obtained from (Shah et al. 2017). The resting-state fMRI data are acquired at TR (Repetition Time) = 1s, 7 min fMRI sessions for each subject, the number of data points is 421, and the number of subjects is 23. We consider the following seven ROIs from the medial temporal lobe, which is referred to (Sanchez-Romero et al. 2019a). The detail information of ROIs is shown in Table 2.

Table 2: The ROIs of the real resting-state fMRI dataset.

NO.	ROIs	Detailed descriptions
1	CA1	Cornu Ammonis 1
2	CA23DG	CA2, CA3 and Dentate Gyrus
3	SUB	Subiculum
4	ERC	Entorhinal Cortex
5	BA35	Brodmann Areas 35
6	BA36	Brodmann Areas 36
7	PHC	Parahippocampal Cortex

Real Task fMRI Dataset We also use real task fMRI dataset by (Ramsey et al. 2010) to test the performance of the EC-GAN. The real task fMRI data are acquired with a 3T scanner, TR = 2s, and the number of data points is 160, and the number of subjects is 9. For our analysis and comparison, we consider 8 ROIs which include: left and right occipital cortex (LOCC, ROCC), left and right anterior cingulate cortex (LACC, RACC), left and right inferior frontal gyrus (LIFG, RIFG), and left and right inferior parietal (LIPL, RIPL).

Comparison Methods for Evaluation

To intuitively show the competitiveness of the EC-GAN, we compare EC-GAN with the other seven methods, some of them are classical methods, and some of them are state-of-the-art methods. In particular, all baseline methods are proposed for inferring effective connectivity (or haven used widely for inferring effective connectivity). These methods include: Peter and Clark (PC) (Meek 2013), Greedy equivalence search (GES) (Ramsey et al. 2010), Linear non-Gaussian acyclic model (LiNGAM) (Shimizu et al. 2006), Granger causality (GC) (Seth, Barrett, and Barnett 2015), Group iterative multiple model estimation (GIMME) (Gates and Molenaar 2012), Patel's conditional dependence measure (Patel) (Wang et al. 2017), and Ant colony optimization combining with voxel activation information (VACOEC) (Liu et al. 2019), respectively.

The parameters of the baseline methods under comparison are selected according to the existed literature, and all codes are from the authors (literature). (Smith et al. 2011; Gates and Molenaar 2012; Liu et al. 2019). The default parameter configurations of the corresponding methods are as follows. PC runs with $Alpha = 0.05$. The parameters of GES is set as $PenaltyDiscount = 1.0$. LiNGAM uses the parameters where $Prune\ Factor = 1.0$. GC is set as $max\_lag \in [1, 30]$, $Alpha = 0.05$. GIMME is performed with $groupcutoff = 0.8$, $subcutoff = 0.5$. Patel runs with $bin = 0.75$. VACOEC uses the parameters where $\alpha = 1$, $\beta = 2$, $\rho = 0.2$, $q_0 = 0.8$, $n = 10$, $p = 0.6$, $K = 0.2$.

Evaluation metrics

We compared the learned result to ground-truths on the four most common graph metrics (Cai et al. 2018; Chikahara and Fujino 2018; Huang et al. 2018): 1) Precision, 2) Recall, 3) F1-measure (F1), and 4) Structural Hamming distance (SHD). In detail, Precision and Recall range from 0 to 1, and F1 value combines the effects of Precision and Recall. If F1 is equal to 1, that indicates the method correctly infers all arcs. SHD is the total number of edge additions (extra arcs), and deletions (missing arcs) needed to convert the learned effective connectivity network into the ground truth network. If SHD is equal to 0, which means that the method correctly infers all arcs.

Model Configuration

Before conducting the comparative experiments, we first generated some simulated data with known ground-truth to select the hyper-parameters of the EC-GAN. The generation model of simulations is referenced to the method in (Smith et al. 2011), which used the dynamic causal modeling (DCM) to generate the fMRI time series data. In our experiments, the EC-GAN employs n effective connectivity

Table 3: The mean and the standard deviation results of 8 methods on benchmark simulation dataset.

Data	Metrics	Methods							
		PC	GES	LiNGAM	GC	GIMME	Patel	VACOEC	EC-GAN
Smith	Precision	0.31±0.02	0.61±0.03	0.48±0.04	0.48±0.04	0.48±0.02	0.56±0.02	0.65±0.02	**0.64±0.01**
	Recall	0.26±0.03	0.46±0.02	0.26±0.03	0.28±0.02	0.46±0.01	0.54±0.01	0.48±0.02	**0.76±0.02**
	F1	0.28±0.02	0.52±0.03	0.34±0.03	0.35±0.03	0.46±0.02	0.55±0.01	0.55±0.02	**0.70±0.01**
	SHD	6.50±0.18	3.90±0.20	5.40±0.21	5.70±0.39	5.20±0.15	4.50±0.07	4.30±0.14	**3.40±0.11**
Sanchez	Precision	0.48±0.01	0.58±0.02	0.52±0.02	0.52±0.01	0.53±0.01	0.57±0.01	0.67±0.02	**0.72±0.01**
	Recall	0.49±0.01	0.67±0.02	0.54±0.02	0.48±0.01	0.70±0.01	0.55±0.02	0.75±0.02	**0.84±0.01**
	F1	0.48±0.01	0.62±0.02	0.53±0.02	0.49±0.01	0.60±0.01	0.55±0.01	0.70±0.02	**0.77±0.01**
	SHD	4.70±0.10	3.30±0.15	4.70±0.14	5.70±0.12	4.20±0.08	4.70±0.18	3.40±0.21	**2.20±0.11**

generators, and each of them has one hidden layer of m neurons with tanh as the activation function. The discriminator has two hidden layers of m Sigmod units on each layer. For a five nodes fMRI time series data, the hyper-parameters of EC-GAN are set as: the learning rate of generator and discriminator are 0.1, the number of units m is 100, sparsity parameter λ is 5. The threshold of causal parameters (A) is determined by the maximum number of parents $MaxP$, and it is the same for all nodes. The selection of the threshold depends on the number of nodes and how sparse networks we want to get. For the two simulated fMRI datasets (five nodes effective connectivity networks), we set $MaxP = 2$. And for the two real fMRI datasets (eight and nine nodes effective connectivity networks), all hyper-parameters are the same with the simulated datasets but set $MaxP = 5$.

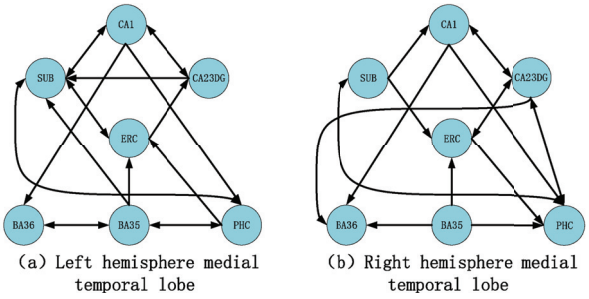

Figure 2: The effective connectivity networks inferred by EC-GAN from the left hemisphere medial temporal lobe (a) and the right hemisphere medial temporal lobe (b).

Experimental Results

Results on Benchmark Simulated fMRI Dataset

In the experiments, we randomly select 20 subjects from Smith dataset and Sanchez dataset, and show the mean and the standard deviation results of 8 methods using single subjects. The results on Smith dataset and Sanchez dataset are shown in Table 3 (Each method runs on single subjects, so we show the mean μ and the standard deviation σ results over all subjects). In our experiments, we use single subjects (with 200 or 500 data points), mainly because we want to test the performance of different methods on a small sample size of fMRI data. In particular, an algorithm performs well when it gets higher values of precision, recall, F1, and lower value of SHD.

The Smith dataset we used is generated with non-stationarity connection strength between brain regions, and the connection strength is modulated over time by additional random processes. Besides, the data points of each subject in this dataset are really small, only 200 data points per subject. In this situation, current methods always perform worse and unable to accurately identify effective connectivity. From Table 3 we can find that most baseline methods perform worse, however, our proposed framework EC-GAN achieves the best performance in Precision, Recall, F1 and SHD. It is worth to note that the Recall and F1 of EC-GAN are much higher than other methods, which indicates that EC-GAN can infer more reliable effective connectivity than other compared methods. Besides, the SHD value of EC is smaller than all other methods, which means that EC has fewer error arcs compared to other methods. Results on Smith simulated dataset show that EC-GAN has better performance under the situations of non-stationarity connection strength and small sample size of fMRI data.

As described in the experimental setup section, the ground-truth of Sanchez simulated dataset is a five nodes graph with one bidirectional arc. Thus, this dataset is employed to test whether a method can infer the bidirectional structure of effective connectivity network. From Table 3 we can find that the EC-GAN achieves the best performance on Precision, Recall, F1 and SHD. Therefore, EC-GAN performs better than the seven comparison methods on the simulated datasets. We will show its performance in real fMRI data in the following section.

Results on Real Resting-state fMRI Dataset

Different from the simulated data, we do not have a fully defined ground-truth to exactly assess the performance of different methods from real fMRI (Sanchez-Romero et al. 2019a). Instead, we have partial knowledge about the presence of structural connections between brain regions on the medial temporal lobe from some current works. Thus we evaluate the performance of EC-GAN on the medial temporal lobe data from the left and right hemispheres of the brain. In the real-world data experiments, we run our proposed framework EC-GAN on one repetition of all 23 subjects concatenated (23 subjects $\times$ 421 time points = 9683

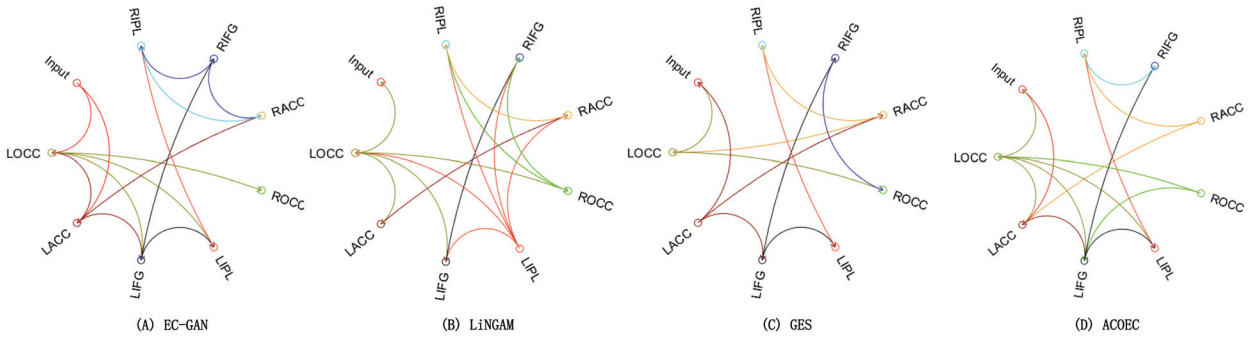

Figure 3: The brain effective connectivity networks inferred by EC-GAN, LiNGAM, GES, and ACOEC on the task fMRI data. The outermost rings represent the brain regions and the center is a representation of brain effective connectivity. The color of the arrows is the same as the parent nodes.

data points) from Shah resting-state fMRI data (Shah et al. 2017). The data contains 7 ROIs from the medial temporal lobe (left hemisphere and right hemisphere are analyzed separately). Figure. 2 illustrates the effective connectivity networks inferred by EC-GAN from the left hemisphere medial temporal lobe and right hemisphere medial temporal lobe.

From Figure. 2 we can see that the effective connectivity network of the left hemisphere medial temporal lobe (Figure. 2 (a)) is closely similar to that of the right hemisphere medial temporal lobe (Figure. 2 (b)), and has some differences. These differences are mainly caused by the connections of CA23DG, effective connectivity $CA23DG \rightarrow SUB$ is in the left hemisphere while $CA23DG \leftrightarrow PHC$ and $CA23DG \rightarrow BA36$ are in the right hemisphere. Compared with the previous studies (Sanchez-Romero et al. 2019a), the effective connectivity network of the left hemisphere medial temporal lobe in Figure. 2 (a) is consistent with the effective connectivity networks estimated by (Sanchez-Romero et al. 2019a). Besides, as is suggested by (Lavenex and Amaral 2000), there are direct two-way connections between PHC and BA35 ($PHC \leftrightarrow BA35$), which is missing in (Sanchez-Romero et al. 2019a), however, this connection is found by EC-GAN. In particular, EC-GAN infers the effective connectivity $ERC \rightarrow CA23DG$, and this effective connectivity is the main pathway connecting the medial temporal lobe cortices with the hippocampus. However, the two methods proposed in (Sanchez-Romero et al. 2019a) fail to detect the effective connectivity ERC→CA23DG, which is inferred by EC-GAN. Therefore, the new proposed method EC-GAN can provide a reliable perspective for the analysis of brain effective connectivity networks.

Results on Real Task fMRI Dataset

In this section, we run EC-GAN on one repetition of 9 subjects concatenated (9 subjects $\times$ 160 time points = 1440 data points) from (Sanchez-Romero et al. 2019b) task fMRI data.

As we do not have a fully defined ground-truth to exactly assess the performance of different methods from real fMRI data (Sanchez-Romero et al. 2019a). We evaluate the performance of LiNGAM, GES and ACOEC (these methods perform well on two simulated datasets) on the task fMRI dataset based on the partial known knowledge. (Sanchez-Romero et al. 2019a) suggested that we can model the dynamics of the task with an input variable for which er expect feedforward edges into the regions of interest and not vice versa. Therefore, we employ EC-GAN and other three compared methods to infer the brain effective connectivity networks with 9 ROIs (include input variable) on the task fMRI data, and the results are graphically rendered in a circular diagram format in Figure. 3

From Figure. 3 we can find that only EC-GAN correctly inferred the effective connectivity $Input \rightarrow LOCC$ and $Input \rightarrow LACC$, which indicates that the proposed framework EC-GAN can correctly infer the feedforward connections from the Input variable to the brain regions. This result is consistent with the rhyming task in (Ramsey et al. 2010) and (Sanchez-Romero et al. 2019a). From Figure. 3 (A), we also find that the left hemisphere of brain regions always activated earlier than the right hemisphere of brain regions under this task, as the information flow is following the chain of Input-LOCC/LIPG-ROCC/RIPG ($Input \rightarrow LOCC$, $LOCC \rightarrow ROCC$, $Input \rightarrow LACC$, $LACC \rightarrow RACC$).

In a word, the new framework EC-GAN can provide a reliable perspective for the analysis of effective connectivity on both resting-state fMRI data and task fMRI data.

Conclusion

In this paper, we proposed a new framework for inferring brain effective connectivity from fMRI data based on generative adversarial networks (GAN), named as EC-GAN. The proposed framework infers effective connectivity via a generator and a discriminator. In detail, the generator is composed of several effective connectivity generators which can generate the fMRI time series of each brain region based on effective connectivity, and the discriminator is employed to distinguish between the joint distributions of real and generated fMRI time series. Experimental results on both simulated and real-world data demonstrate the efficacy of our proposed framework.

Acknowledgments

We thank reviewers for their helpful advice. This work is partly supported by NSFC Research Program (61672065).

References

Bühlmann, P.; Peters, J.; and Ernest, J. 2014. Cam: Causal additive models, high-dimensional order search and penalized regression. *The Annals of Statistics* 2526–2556.

Cai, R.; Qiao, J.; Zhang, Z.; and Hao, Z. 2018. Structural equational likelihood framework for causal discovery. In *Proc. of AAAI'18*.

Chikahara, Y., and Fujino, A. 2018. Causal inference in time series via supervised learning. In *Proc. of IJCAI'18*.

Friston, K. J. 1994. Functional and effective connectivity in neuroimaging: a synthesis. *Human Brain Mapping* 56–78.

Gates, K. M., and Molenaar, P. C. 2012. Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples. *NeuroImage* 310–319.

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative adversarial nets. In *Proc. of NeurIPS'14*, 2672–2680.

Hein, G.; Morishima, Y.; Leiberg, S.; Sul, S.; and Fehr, E. 2016. The brain's functional network architecture reveals human motives. *Science* 1074–1078.

Huang, B.; Zhang, K.; Lin, Y.; Scholkopf, B.; and Glymour, C. 2018. Generalized score functions for causal discovery. In *Proc. of KDD'18*, 1551–1560.

Kalainathan, D.; Goudet, O.; Guyon, I.; Lopez-Paz, D.; and Sebag, M. 2018. Sam: Structural agnostic model, causal discovery and penalized adversarial learning. *arXiv preprint arXiv:1803.04929*.

Kocaoglu, M.; Snyder, C.; Dimakis, A. G.; and Vishwanath, S. 2017. Causalgan: Learning causal implicit generative models with adversarial training. *arXiv preprint arXiv:1709.02023*.

Lavenex, P., and Amaral, D. G. 2000. Hippocampal-neocortical interaction: A hierarchy of associativity. *Hippocampus* 420–430.

Li, Z.; Zhang, T.; Wan, P.; and Zhang, D. 2019. Segan: Structure-enhanced generative adversarial network for compressed sensing mri reconstruction. In *Proc. of AAAI'19*, 1012–1019.

Liu, J.; Ji, J.; Zhang, A.; and Liang, P. 2016. An ant colony optimization algorithm for learning brain effective connectivity network from fmri data. In *Proc. of BIBM'16*, 360–367. IEEE.

Liu, J.; Ji, J.; Jia, X.; and Zhang, A. 2019. Learning brain effective connectivity network structure using ant colony optimization combining with voxel activation information. *IEEE journal of biomedical and health informatics*.

Meek, C. 2013. Causal inference and causal explanation with background knowledge. *arXiv preprint arXiv:1302.4972*.

Ramsey, J. D.; Hanson, S. J.; C., H.; Halchenkob, Y. O.; Poldrackc, R. A.; and Glymour, C. 2010. Six problems for causal inference from fmri. *Magnetic resonance in medicine* 1545–1558.

Ramsey, J. D.; Sanchez-Romero, R.; and Glymour, C. 2014. Non-gaussian methods and high-pass filters in the estimation of effective connections. *Neuroimage* 986–1006.

Rytsar, R.; Fornari, E.; S., F. R.; Ghika, J. A.; and Knyazeva, M. G. 2011. Inhibition in early alzheimers disease: An fmri-based study of effective connectivity. *Neuroimage* 1131–1139.

Sanchez-Romero, R.; Ramsey, J. D.; Zhang, K.; Glymour, M. R.; Huang, B.; and Glymour, C. 2019a. Estimating feedforward and feedback effective connections from fmri time series: Assessments of statistical methods. *Network Neuroscience* 274–306.

Sanchez-Romero, R.; Ramsey, J. D.; Zhang, K.; Glymour, M. R.; Huang, B.; and Glymour, C. 2019b. Supporting information for estimating feedforward and feedback effective connections from fmri time series: Assessments of statistical methods. *Network Neuroscience* 274–306.

Seth, A. K.; Barrett, A. B.; and Barnett, L. 2015. Granger causality analysis in neuroscience and neuroimaging. *Journal of Neuroscience* 3293–3297.

Shah, P.; Bassett, D. S.; Wisse, L. E.; Detre, J. A.; Stein, J. M.; Yushkevich, P. A.; and Das, S. R. 2017. Mapping the structural and functional network architecture of the medial temporal lobe using 7t mri. *Human Brain Mapping* 851–865.

Shimizu, S.; Hoyer, P. O.; Hyvärinen, A.; and Kerminen, A. 2006. A linear non-gaussian acyclic model for causal discovery. *The Journal of Machine Learning Research* 2003–2030.

Smith, S. M.; Miller, K. L.; Salimi-Khorshidi, G.; Webster, M.; Beckmann, C. F.; Nichols, T. E.; Ramsey, J. D.; and Woolrich, M. W. 2011. Inhibition in early alzheimers disease: An fmri-based study of effective connectivity. *Neuroimage* 875–891.

Wang, Y.; David, O.; Hu, X.; and Deshpande, G. 2017. Can patel's τ accurately estimate directionality of connections in brain networks from fmri. *Magnetic resonance in medicine* 2003–2010.

Wang, H.; Wang, J.; Wang, j.; Zhao, M.; Zhang, W.; Zhang, F.; Xie, X.; and Guo, M. 2018. Graphgan: Graph representation learning with generative adversarial nets. In *Proc. of KDD'18*.

Zhou, L.; Wang, L.; Liu, L.; Ogunbona, P.; and Shen, D. 2016. Learning discriminative bayesian networks from high-dimensional continuous neuroimaging data. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2269–2283.