MetaRLEC: Meta-Reinforcement Learning for Discovery of Brain Effective Connectivity

Zuozhen Zhang, Junzhong Ji, Jinduo Liu *

Beijing Municipal Key Laboratory of Multimedia and Intelligent Software Technology, Beijing Institute of Artificial Intelligence, Faculty of Information Technology, Beijing University of Technology, Beijing, China
zzz3582@emails.bjut.edu.cn, jjz01@bjut.edu.cn, jinduo@bjut.edu.cn

Abstract

In recent years, the discovery of brain effective connectivity (EC) networks through computational analysis of functional magnetic resonance imaging (fMRI) data has gained prominence in neuroscience and neuroimaging. However, owing to the influence of diverse factors during data collection and processing, fMRI data typically exhibit high noise and limited sample characteristics, consequently leading to the suboptimal performance of current methods. In this paper, we propose a novel brain effective connectivity discovery method based on meta-reinforcement learning, called MetaRLEC. The method mainly consists of three modules: actor, critic, and meta-critic. MetaRLEC first employs an encoder-decoder framework: The encoder utilizing a transformer converts noisy fMRI data into a state embedding, and the decoder employing bidirectional LSTM discovers brain region dependencies from the state and generates actions (EC networks). Then, a critic network evaluates these actions, incentivizing the actor to learn higher-reward actions amidst the high-noise setting. Finally, a meta-critic framework facilitates online learning of historical state-action pairs, integrating an action-value neural network and supplementary training losses to enhance the model's adaptability to small-sample fMRI data. We conduct comprehensive experiments on both simulated and real-world data to demonstrate the efficacy of our proposed method.

Introduction

In recent years, the study of brain effective connectivity (EC) has gained prominence within neuroscience and brain imaging research. EC is defined as the causal influence that one brain region exerts over another (Friston 2011), and it plays a pivotal role in neural development and disease analysis (Ji et al. 2021b). With the ongoing advancement of neuroimaging data, particularly functional magnetic resonance imaging (fMRI), the pursuit of accurate and efficient methods to discover brain EC networks from such data has become a prominent focus in this research field.

Over the past few years, there has been a surge in the development of methods employing fMRI data to discover brain EC networks. These methods contribute to a deeper comprehension of the intricate interactions among distinct brain regions and their connection to cognitive processes. These methods can be broadly categorized into two groups: traditional machine learning-based (ML) methods and deep learning-based (DL) methods. ML methods (Liu et al. 2022; Pfarr et al. 2021; Mao et al. 2022; Jiang et al. 2023) typically employ straightforward models that offer interpretability, stability, and robustness owing to their dependence on prior knowledge and assumptions. Nonetheless, these approaches are vulnerable to noise and encounter challenges when handling temporal data and dynamic processes, as they do not explicitly account for temporal correlations. In contrast, DL methods have significant advantages in processing high-dimensional data and nonlinear information flow, which have been successfully applied in the study of brain EC, such as convolutional neural networks (CNNs) (Bagherzadeh, Shahabi, and Shalbaf 2022; Khan et al. 2023), recurrent neural network (RNN) (Ji et al. 2021a) and generative adversarial network (GAN) (Liu et al. 2020). These methods usually have higher accuracy and better scalability.

However, in clinical research, fMRI subjects may experience interference from their own physiological factors (e.g., heartbeat, breathing) and external environmental influences. As a consequence, fMRI data often exhibit substantial noise levels, significantly constraining the effectiveness of this approach. Additionally, owing to the substantial cost associated with fMRI data collection and processing, the available sample size of fMRI time series data remains limited. This scarcity of training data presents a formidable hurdle for deep learning models striving to accurately discover brain EC networks. Therefore, discovering brain EC networks from small-sample fMRI time series data with high noise remains an extremely challenging issue in this field.

Recently, there has been notable progress in the realm of reinforcement learning (Henderson et al. 2018). This technique uses an agent to explore the environment by trying different actions and continuously learns and optimizes strategies based on observed rewards or feedback from the environment, thus enabling it to adapt to different environmental noises and perform better. At the same time, the success of the meta-learning paradigm (Liang et al. 2023) on small sample problems in many application domains proves its effectiveness. On the one hand, meta-learning can make use of knowledge learned from other related tasks to assist in the learning of the task at hand. On the other hand, meta-

*Corresponding Author: Jinduo Liu.

learning can enhance the learning performance of reinforcement learning through optimization strategies (Gupta et al. 2018), loss functions (Huang et al. 2019), and intrinsic rewards (Zheng, Oh, and Singh 2018), enabling agents to use prior experience to make maximum learning progress to discover EC from small-sample fMRI data quickly and accurately.

In this paper, we propose a brain EC discovery method based on meta-reinforcement learning, called MetaRLEC. The method mainly consists of three modules: actor, critic, and meta-critic. Specifically, the approach is initiated by employing an encoder-decoder framework within the actor network. The encoder employs a transformer to encode the input high-noise fMRI data into a state embedding, while the decoder utilizes a bidirectional long short-term memory (LSTM) to discover brain region dependencies from the state and generate actions. Subsequently, the critic network assesses the actions, with the actor honing its ability to take higher-reward actions to adapt to the high-noise environment. Finally, the meta-critic framework is engaged for the online learning of historical state-action pairs, encompassing an action-value neural network and supplementary training losses to amplify the model's learning proficiency with small-sample fMRI data. The proposed method has undergone rigorous testing using both simulated and real-world fMRI data, with experimental results demonstrating certain performance advantages compared to existing state-of-the-art methods.

The principal contributions of this paper can be summarized as follows:

- To the best of our knowledge, this is the first work that integrates meta-reinforcement learning strategies to proficiently discover brain EC from small-sample fMRI data.
- We introduce a novel actor-critic framework for the discovery of brain EC, capable of effectively extracting features from high-noise fMRI data.
- We propose a meta-critic framework designed to facilitate actor-critic learning, thereby enabling the actor to achieve optimal learning efficiency and consequently enhancing the precision of brain EC discovery.
- Systematic experiments conducted on both simulated and real fMRI datasets demonstrate that the proposed method surpasses several state-of-the-art approaches in its performance on small-sample fMRI data.

Preliminary and Related Work

Notation and Problem Description

We introduce the notation and subsequently formulate a problem description for the task of discovering brain EC networks.

In this paper, we utilize uppercase letters, i.e., X_i to denote the variables of brain regions and use the lowercase letters x_i to represent the time series of the brain region X_i. The length of time series x_i is t. We use $\mathrm{Pa}\left(x_i\right)$ to denote the time series of parent nodes of brain region X_i. $X_a \rightarrow X_b$ represent the effects exerted by brain region X_a on brain region X_b, and $X_a \leftrightarrow X_b$ denotes that brain region X_a and brain region X_b have an effect on each other.

Based on the definition of brain EC, which refers to the neural influence exerted by one brain region over another, the EC among brain regions can be seen as directed edges within a causal graph (a directed graph) where the nodes represent different brain regions (Sanchez-Romero et al. 2019). Therefore, the task of learning brain EC can be transformed into a problem of discovering a causal graph from fMRI time series data. Let $\mathcal{G}$ denote a directed graph and $\mathcal{X}$ denote the fMRI dataset. Therefore, a brain EC 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 $X_a \rightarrow X_b \in \mathbf{E}$ describing an EC from brain regions (ROIs) X_a to X_b.

Meta-reinforcement Learning

Meta-reinforcement learning is a promising approach for tackling few-episode learning regimes. It aims to learn to adapt quickly to new tasks using the prior experience gained from multiple related tasks. It focuses on developing strategies for effective learning in complex environments, i.e., high noise. Two of the most popular are context-based methods (Sodhani, Zhang, and Pineau 2021; Kirsch et al. 2022; Yuan and Lu 2022) and optimization-based methods (Finn, Abbeel, and Levine 2017; Tang 2022; Bechtle et al. 2021). Another popular family of approaches uses a neural network model (NN) to extract some context that will be used to inform the policy (Wang and Van Hoof 2022). Context-based methods can use contextual information from the environment to optimize decisions but it may require large computational resources. Optimization-based methods allow knowledge transfer between tasks but it needs a large task family to train. Our meta-reinforcement learning framework is in the category of optimization-based methods, but unlike most of these methods, we are able to meta-learn the loss function online in parallel to learning a single extrinsic task without costly offline learning on a task family.

Method

In this section, we present the MetaRLEC method which is a meta-reinforcement learning based causal structure learning procedure to discover brain EC from fMRI time series data. We first give an overview of the proposed model, and then describe the details of the main components. Finally, we show the description of MetaRLEC.

MetaRLEC Architecture

The method primarily consists of three modules: actor, critic, and meta-critic. Specifically, the method begins by employing an encoder-decoder framework as the actor network. The encoder utilizes the transformer model to encode the input fMRI data into state embeddings, while the decoder employs bidirectional LSTM to discover the causal relationship among brain regions from these states, generating actions (brain EC network). Transformer-based encoder

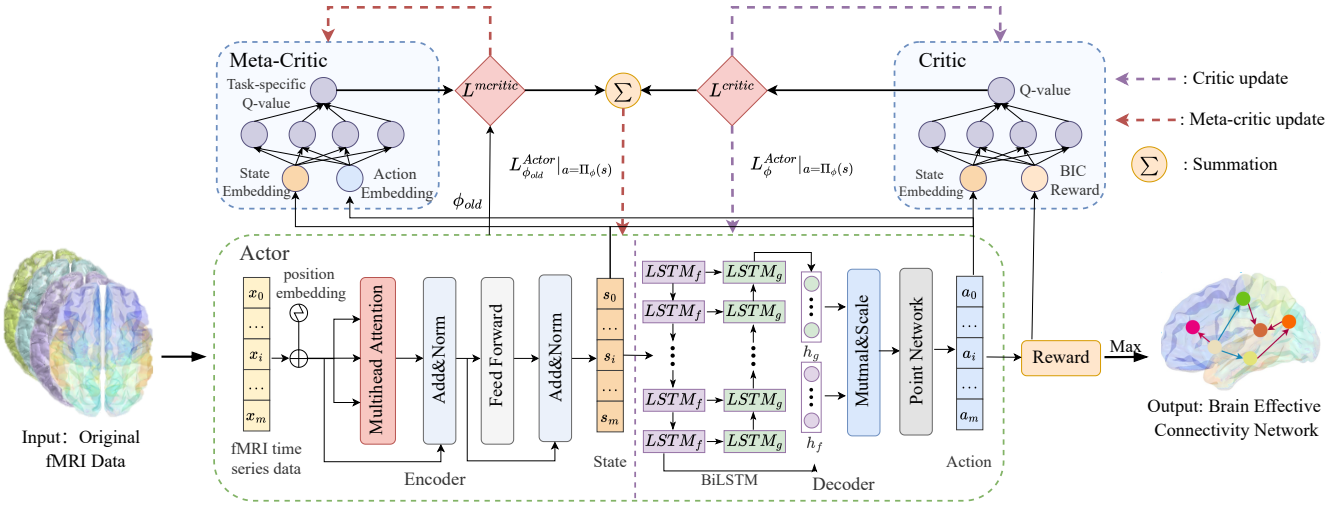

Figure 1: The architecture of MetaRLEC.

and BiLSTM-based decoder help capture long-term dependencies from high noise fMRI data. Then, the critic network evaluates the actions and trains the actor to adopt actions with higher returns. Finally, the meta-critic framework is employed to learn the historical state-action pairs online. This framework comprises an action-value neural network and additional training loss to enhance the learning ability of the model on small sample fMRI data. The structure of the proposed MetaRLEC is shown in Figure 1.

Actor

Transformer-based Encoder Given the fMRI time series with n brain regions X_i $(i = 1, ..., n)$ and t length, the input training data $\mathcal{X}$ can be represented as:

$$\mathcal{X} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n)^\top \in \mathbb{R}^{n \times t}. \tag{1}$$

A common way is to use $\mathcal{X}$ as input to the network directly. However, the high noise characteristic of fMRI data presents a great challenge for general feed-forward neural networks to capture the underlying causal relationships directly using $\mathcal{X}$ as states. Consequently, incorporating an encoder module to preprocess the fMRI data proves beneficial in extracting useful information and finding better brain EC networks. Therefore, we sample $\mathcal{X}$ m times and use an encoder to extract the state embedding each X^m to state $S_{enc} := \{s_0, s_1, ..., s_m\}$.

For the model design of the encoder, we utilize the transformer model, which involves first embedding the inputs via a linear layer, followed by processing them through multiple identical encoder blocks comprising a multihead self-attention layer and a feed-forward layer; we posit that multihead self-attention is well suited for extracting temporal features from fMRI time series data, as it reduces reliance on external information and better captures internal correlations within the fMRI data. The encoder block operations are as follows:

$$\begin{aligned} X' &= Linear(X^m) \in \mathbb{R}^{m \times n \times h}, \\ Q_l &= W^Q X'^l + \epsilon^Q, \\ K_l &= W^K X'^l + \epsilon^K, \\ V_l &= W^V X'^l + \epsilon^V, \end{aligned} \tag{2}$$

where X^m denotes the embedded input, $Linear$ is a fully connected linear layer that provides a linear transformation of the input X^m, h denotes the number of hidden layer nodes of the fully connected linear layer, and X'^l expresses the l-th input after dividing the embedding X' into the L head self-attention layer. W^Q, W^K and W^V denote the network parameters for the self-attention layer, and ϵ^Q, ϵ^K and ϵ^V are the bias vectors. Then, we can obtain Q_l, K_l and V_l, which denote the query, key, and value vector of the self-attention layer, respectively. Therefore, the self-attention can be calculated as follows:

$$\begin{aligned} \text{Attn}_l &= \text{softmax}\left(\frac{Q_l K_l^T}{\sqrt{d_{K_l}}}\right) V_l, \\ S_{enc} &= \text{Concat}\left(Attn_1, Attn_2, ..., Attn_L\right), \end{aligned} \tag{3}$$

where d denotes the number of elements in the last dimension of the query, key, and value vector Q_l, K_l, and V_l, Attn_l describes the l-th head attention vector. Then, we can collect all L heads of the self-attention vectors. And obtain the embedded state S_{enc}.

BiLSTM-based Decoder Based on the state S_{enc} learned by the encoder, we use a decoder to discover the underlying causal relationships among brain regions. We choose bidirectional long short-term memory (Bi-LSTM) (Zhang et al. 2015) with an attention layer for the decoder since it not only has a strong ability to capture the dependencies between time series but can also solve the problems of gradient disappearance and gradient explosion in the process of

long sequence training. The Bi-LSTM model is made up of multiple LSTM cells.

The process of the Bi-LSTM model consists of two layers: the forward layer and the backward layer. Specifically, the forward layer, from time step 1 to t, updates the long-term memory and stores the hidden state. Given the encoder output S_{enc} of the j-th time step s_j, the hidden state can be represented as follows:

$$\overrightarrow{H_j} = f\left(s_j W_1^{(f)} + \overrightarrow{H_{j-1}} W_2^{(f)} + \text{ bias }^{(f)}\right), \quad (4)$$

where $W_1^{(f)}, W_2^{(f)}$ and bias $^{(f)}$ are parameters of the forward layer and function f denotes the LSTM model. The process of the backward layer is the same as the process of the forward layer, except the time step is from t to 1:

$$\overleftarrow{H_j} = f\left(s_j W_1^{(g)} + \overleftarrow{H_{j-1}} W_2^{(g)} + \text{ bias }^{(g)}\right), \quad (5)$$

where $W_1^{(g)}, W_2^{(g)}$ and bias $^{(g)}$ are parameters of the backward layer. After concatenating the hidden state of both layers, we obtain the hidden state of the Bi-LSTM. The output can be represented by the following formula:

$$O_j = \left[\overrightarrow{H_j}, \overleftarrow{H_j}\right] W + \text{bias}. \quad (6)$$

We then use a pointer network to generate actions (brain EC networks) from O_j. We restrict each node to be selected only once by masking the selected nodes to generate a sparse brain EC network $\mathcal{G}$.

Critic

For the critic, we use a 3-layer feed-forward neural network with a ReLU activation function. The input of the critic network is the decoder output (actions) and the rewards. To better assess the value of the learning brain EC network, we use the Bayesian information criterion score (BIC) as the reward function as follows:

$$\mathrm{S}_{\mathrm{BIC}}(\mathcal{G}) = \sum_{i=1}^{n}\left[\sum_{j=1}^{t} \log p\left(x_{ij} \mid \mathrm{Pa}\left(x_{ij}\right); \theta_j\right) - \frac{|\theta_j|}{2}\log t\right], \quad (7)$$

where θ_j is the parameter associated with each likelihood, and $|\theta_j|$ denotes the dimension of the parameter. x_{ij} denotes the data point of brain region X_i at time j. The BIC score enables us to identify the optimal brain EC that best aligns with the fMRI time series data, which leads to improved performance of the EC network. Therefore, the reward can be described as:

$$\text{reward}(\mathcal{G}) = -[\mathrm{S}_{\mathrm{BIC}}(\mathcal{G}) + \lambda A(\mathcal{G})], \quad (8)$$

where $\lambda \geq 0$ is a parameter that controls the sparsity of brain EC networks and $A(\mathcal{G})$ is the sparse penalty function as $A(\mathcal{G}) = \|\mathcal{G}\|_1$. $\mathrm{S}_{\mathrm{BIC}}$ denotes the score for the action (brain EC network). By utilizing three fully connected layers, the critic network can effectively capture the intricate relationship between actions and rewards. At the same time, the output of the critic network provides a loss L^{critic} for the actor that trains actor network to produce more highly rewarded actions (brain EC networks).

Meta-critic

In this section, we introduce a meta-critic framework to enhance the basic EC learning methods. The framework provides an additional loss to guide the learning process of the primary actor and critic networks and a meta-critic network that acts as a higher-level agent that observes the policy and outcomes of actor and critic models. Based on this observation, the meta-critic network provides feedback and guides the algorithm to obtain maximum learning efficiency in high noise and small sample fMRI data.

In contrast to the primary critic, the meta-critic is trained in a meta-learning way to expedite the learning process rather than solely calculating the action-value function. Generally, the actor is trained by both critic and meta-critic-provided losses, The critic is trained as usual, and crucially, the meta-critic is trained online to generate maximum learning progress in the actor. It is worth noting that the meta-critic can be effectively learned online within a single task, which differs from the prevailing meta-learning paradigm.

The input of the meta-critic network is the actor encoder output (state) and decoder output (action). The basic idea is similar to the feature extractor in supervised learning, the actor needs to analyze and extract information from fMRI time series data for decision-making. We divide this process into two steps: feature extraction and decision-making (i.e., the actor encoder and decoder). The meta-critic network provides a loss $L^{mcritic}$ to evaluate the output of both processes simultaneously as follows.

$$L^{mcritic}\left(d_{trn}; \phi\right) = \frac{1}{m}\sum_{i=1}^{m} f_\omega\left(s_i, a_i\right), \quad (9)$$

where d_{trn} is a batch m sampled from fMRI time series data $\mathcal{X}$, ϕ denotes the parameters of the actor network, and s_i and a_i denote the state and action, respectively. To minimize unnecessary computations, we use a 3-layer feed-forward neural network with a ReLU activation function as f_ω, which is similar to the critic network. The higher $L^{mcritic}$ for the state and action is, the better the actor is learned.

Then, we give the meta-loss definition as follows, which aims to measure whether the meta-critic improves the performance of the actor compared to the primary critic.

$$\begin{aligned}
\phi_{old} =& \phi - \eta\frac{\partial L^{critic}\left(d_{trn}\right)}{\partial \phi}, \\
\phi_{new} =& \phi - \eta\frac{\partial L^{critic}\left(d_{trn}\right)}{\partial \phi} - \eta\frac{\partial L_\omega^{\text{mcritic}}\left(d_{trn}\right)}{\partial \phi}, \\
L^{meta} =& \tanh\left(L^{critic}\left(d_{val}; \phi_{new}\right) - L^{critic}\left(d_{val}; \phi_{old}\right)\right),
\end{aligned} \quad (10)$$

where η denotes the learning rate. d_{val} are the different batches sampled from fMRI time series data $\mathcal{X}$, d_{trn} is for training and d_{val} is for validation of online learning. L^{critic} denotes the loss provided by the primary critic. The $\tanh$ function ensures that the meta-loss range is always distributed in $(-1, 1)$ and caps the magnitude of the meta-gradient. ϕ_{old}, ϕ_{new} denotes the parameters of the actor optimized only by L^{critic} and both two losses, respectively. If

meta-critic provided a beneficial source of loss, ϕ_{new} should be a better parameter than ϕ and, in particular, a better parameter than ϕ_{old}.

Estimating EC by MetaRLEC

Above all, the MetaRLEC algorithm mainly consists of three phases: the initialization phase, meta-training phase, and meta-online testing phase. The algorithm is formally stated in Algorithm 1.

In the initialization phase, MetaRLEC performs actor, critic and meta-critic model initialization and sets some basic parameters. The optimization process of MetaRLEC can be seen as a bi-level optimization problem. The meta-training phase corresponds to inner-level optimization, and the meta-online testing phase corresponds to outer-level optimization. Specifically, in the meta-training phase, for R outer loops, MetaRLEC first samples a mini-batch d_{trn} from $\mathcal{X}$ and employs the actor network to generate the state and action. Then, the algorithm utilizes the critic network to observe the reward of the action. Finally, MetaRLEC calculates the basic critic loss and the meta-critic loss from the state, action and reward and performs the first update on the actor and critic networks. In the meta-online testing phase, this algorithm first resamples a batch of fMRI data d_{val} and obtains the corresponding state and action for calculating the meta-loss. Then, MetaRLEC provides a second update to the actor network to ensure that the actor network can train faster and learn a more accurate brain EC network. Finally, the algorithm outputs the brain EC network with the highest reward and post processes it by thresholding.

Algorithm 1: MetaRLEC

Input: Original fMRI time-series data.
Output: Brain EC network.

Initialization:
Initialize actor, critic, meta-critic network
Parameters of actor encoder and decoder, critic, meta-critic: ϕ_{en}, ϕ_{de}, θ and ω.
fMRI time-series data $\mathcal{X}$.
for R iterations **do**
  Meta training:
  Sample training batch d_{trn} from $\mathcal{X}$;
  Get state by ϕ_{en} and select action by ϕ_{de};
  Observe the reward by eq.8;
  Calculate the basic critic loss L^{critic} and meta-critic loss $L^{mcritic}$ by eq.9;
  Update $actor_{old}$ and critic according to L^{critic} only as $\phi_{old} = \phi - \eta \nabla_{\phi} L^{critic}$;
  Update $actor_{new}$ by L^{critic} and $L^{mcritic}$ as $\phi_{new} = \phi_{old} - \eta \nabla_{\phi} L^{mcritic}$;
  Online meta testing:
  Sample testing batch d_{val} from $\mathcal{X}$;
  Get state and action by ϕ_{old} and ϕ_{new}, respectively;
  Calculate meta loss by eq.10;
  Update actor and meta-critic parameters;
end for
Obtain brain EC network with the highest reward;
Post-process;
Return: Brain EC network $\mathcal{G}$.

Experiments

To assess the effectiveness of MetaRLEC, we first conduct a comparative experiment with other state-of-the-art methods using simulated fMRI data from known ground-truth networks. Then, we demonstrate the practical application of our proposed method by applying MetaRLEC to publicly available real fMRI data. The code is available at https://github.com/layzoom/MetaRLEC.

Data Description

Benchmark Simulation Dataset The benchmark simulation datasets [1] we used are supported by Smith et al. (Smith et al. 2011), which are generated by dynamic causal models (DCM). We selected 4 kinds of typical simulation cases to test the performance of the MetaRLEC algorithm, including Sim1 (5 nodes, 5 arcs), Sim2 (5 nodes, 7 arcs), Sim3 (10 nodes, 11 arcs), Sim4 (15 nodes, 19 arcs). Each simulation case consists of 50 subjects, with each session lasting for 600 seconds, closely resembling real-world scenarios. The time repetition (TR) is set at 3.0 seconds, resulting in a preprocessed time series length of 200 data points, which can be considered a relatively small sample size.

Real fMRI Dataset The real fMRI time-series dataset [2] used in this paper is resting-state fMRI data. The resting-state fMRI data for 23 human subjects are acquired at a TR is 1.0 seconds, 7 min fMRI sessions for each subject, resulting in 421 data points of the fMRI time series. We consider the following seven regions of interest (ROIs) from the medial temporal lobe, which is referred to in (Shah et al. 2018), including CA1 (Cornu Ammonis1), CA23DG (Cornu Ammonis2,3 and Dentate Gyrus), SUB (Subiculum), ERC (Entorhinal Cortex), BA35 (Brodmann Areas 35), BA36 (Brodmann Areas 36) and PHC (Parahippocampal Cortex). We use the numerical sequence 1 to 7 to represent them respectively.

[1] https://www.fmrib.ox.ac.uk/datasets/netsim/index.html
[2] https://github.com/shahpreya/MTlnet

Baseline Methods

To test and verify the competitiveness of MetaRLEC, we compare MetaRLEC with the other 8 brain EC learning methods, including classical machine learning methods and state-of-the-art deep learning methods which are Patel (Patel, Bowman, and Rilling 2006), pwLiNGAM (Hyvarinen 2010), lsGC (DSouza et al. 2017), Two-Step (Sanchez-Romero et al. 2019), EC-RGAN (Ji et al. 2021a), RL-EC (Lu et al. 2022), CR-VAE (Li, Yu, and Principe 2023), and DiffAN (Sanchez et al. 2023). The parameters of the algorithms under comparison are selected according to the existing literature and we fine-tune 10 subjects to select the optimal parameters. We use the most common graph metrics (Zhang et al. 2021) to evaluate the performance of those methods, including Precision, Recall, Structural Hamming Distance (SHD), and F1 score (F1).

Data	Metrics	Methods								
		Patel	pwLiNGAM	lsGC	Two-Step	EC-RGAN	RL-EC	CR-VAE	DiffAN	MetaRLEC
Sim1	Precision	0.40±0.09	0.34±0.10	0.36±0.20	0.56±0.16	0.32±0.11	0.35±0.36	0.22±0.09	0.40±0.25	**0.62±0.16**
	Recall	0.60±0.20	**0.77±0.21**	0.37±0.21	0.73±0.23	0.60±0.27	0.30±0.25	0.41±0.22	0.35±0.18	0.65±0.19
	SHD	2.92±1.30	5.35±1.30	4.94±1.65	**1.60±1.27**	4.86±1.41	3.58±1.30	6.32±1.26	3.40±1.00	2.28±1.03
	F1	0.46±0.10	0.47±0.13	0.35±0.17	0.62±0.17	0.40±0.15	0.35±0.28	0.28±0.12	0.40±0.20	**0.63±0.17**
Sim2	Precision	0.36±0.08	0.24±0.14	0.23±0.18	0.40±0.18	0.41±0.11	0.44±0.30	0.26±0.12	0.46±0.28	**0.59±0.24**
	Recall	0.57±0.19	0.63±0.21	0.37±0.21	0.50±0.21	0.48±0.25	0.22±0.16	**0.73±0.18**	0.29±0.18	0.55±0.28
	SHD	4.04±1.54	5.29±1.74	5.88±1.44	3.74±1.56	5.20±1.39	5.58±1.29	4.68±1.24	5.30±1.55	**3.70±2.13**
	F1	0.46±0.10	0.45±0.14	0.35±0.17	0.49±0.17	0.41±0.17	0.29±0.21	0.45±0.09	0.35±0.22	**0.57±0.26**
Sim3	Precision	0.28±0.08	0.18±0.03	0.24±0.11	0.44±0.11	0.14±0.02	0.50±0.10	0.12±0.02	0.52±0.18	**0.54±0.14**
	Recall	0.51±0.16	0.72±0.13	0.34±0.16	0.76±0.15	**0.82±0.13**	0.37±0.12	0.80±0.15	0.36±0.12	0.58±0.16
	SHD	10.38±2.76	28.49±3.43	14.78±3.17	6.62±3.06	32.44±2.21	7.02±1.40	33.74±1.92	7.40±1.51	**6.32±2.20**
	F1	0.35±0.07	0.29±0.05	0.27±0.10	0.55±0.12	0.24±0.03	0.42±0.11	0.21±0.03	0.42±0.14	**0.56±0.15**
Sim4	Precision	0.23±0.07	0.17±0.03	0.21±0.08	0.41±0.07	0.11±0.01	0.50±0.19	0.09±0.01	0.49±0.14	**0.51±0.14**
	Recall	0.40±0.11	0.74±0.10	0.29±0.12	0.68±0.12	**0.91±0.05**	0.22±0.11	0.80±0.10	0.31±0.10	0.51±0.14
	SHD	21.31±4.50	58.10±7.53	26.90±5.66	14.58±3.67	79.86±3.07	15.12±2.13	86.04±2.59	14.34±2.01	**13.44±3.41**
	F1	0.28±0.06	0.27±0.04	0.23±0.07	**0.51±0.08**	0.20±0.01	0.30±0.13	0.16±0.02	0.38±0.11	0.51±0.14

Table 1: The mean and the standard deviation results of 9 methods on Smith simulated dataset using single subject data.

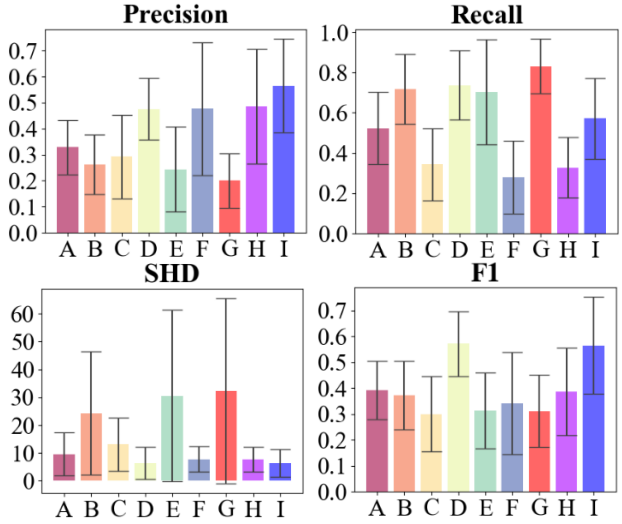

Figure 2: Averaged performance of 9 algorithms on the Smith dataset. A: Patel; B: pwLiNGAM; C: lsGC; D: Two-Step; E: EC-RGAN; F: RL-EC; G: CR-VAE; H: DiffAN; I: MetaRLEC

Methods	Precision	Recall	SHD	F1
Patel	$1.58E-42$	$1.35E-02$	$2.07E-06$	$8.04E-25$
pwLiNGAM	$1.15E-58$	$1.74E-14$	$4.48E-23$	$3.94E-27$
lsGC	$6.10E-44$	$2.56E-27$	$1.04E-16$	$1.08E-42$
Two-Step	$2.04E-08$	$1.95E-17$	$8.65E-01$	$6.04E-01$
EC-RGAN	$1.10E-55$	$1.03E-08$	$1.82E-22$	$7.29E-40$
RL-EC	$1.79E-04$	$1.20E-40$	$4.88E-03$	$3.19E-26$
CR-VAE	$4.25E-76$	$1.16E-40$	$4.24E-22$	$2.68E-41$
DiffAN	$1.74E-04$	$5.58E-34$	$1.47E-02$	$1.59E-20$

Table 2: T test of MetaRLEC and other 8 methods on Precision, Recall, SHD, and F1.

Ablation Study

We conduct ablation experiments on transformer-based encoder (noTE), BiLSTM-based decoder (noBD), and meta-critic network (noMC) on the Sim1 dataset. We briefly demonstrate the results of ablation on F1 metrics (noTE: 0.55 ± 0.13, noBD:0.47 ± 0.21, noMC: 0.44 ± 0.15, MetaRLEC: 0.63 ± 0.17). The meta-critic network has the greatest impact on the results.

Experimental Results

Results and Discussion

To test the performance of each method on a small sample size of fMRI data, we run the 8 baseline methods and MetaRLEC on 4 Smith benchmark simulation datasets in the experiments. We perform individual analyses on every subject and present the mean μ and the standard deviation σ across all subjects. We evaluate these 9 learning methods on Precision, Recall, SHD and F1. In particular, an algorithm performs well when it obtains higher values of Precision, Recall, and F1 and a lower SHD. The results on simulation datasets are shown in Table 1 and Figure 2. The height of each bar indicates the mean value and the error bar denotes the standard deviation. We can see that MetaRLEC achieves optimal or near-optimal results.

To clearly show the significant differences between these algorithms, we use the Friedman test and T test to attest to the significant difference between these algorithms. If the p-value obtained from the test is less than 0.05, we consider that a significant difference exists in the corresponding experimental results. In detail, we first perform the Friedman test on the results of each method for each subject on simulated data. The Friedman test indicates a significant difference between the nine algorithms (p-value < 0.05). Furthermore, we perform the T test on the results of MetaRLEC and other methods, which are described in Table 2. From Table 1 and 2, we can find the MetaRLEC has significant difference (better performance) compared to most other methods.

Generally, deep learning methods are capable of extracting deep features from fMRI data, allowing for more ac-

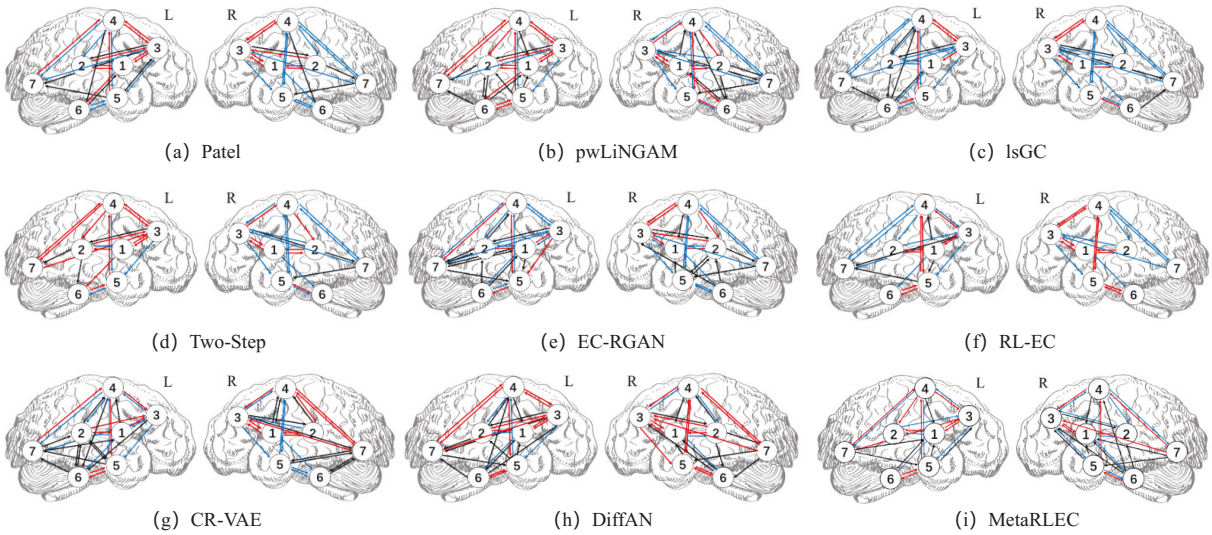

Figure 3: The EC network inferred by 9 methods on real fMRI data. Red lines: important connections between brain regions; Black lines: incorrect or spurious connections; Blue lines: missing connections.

curate and precise results. In contrast, traditional machine learning methods often struggle to overcome the inherent limitations and defects of fMRI data, leading to poorer performance. Overall, the findings suggest that MetaRLEC has significant potential for applications in small sample and high noise fMRI data analysis and may provide more reliable and accurate insights into brain EC.

Results on Real fMRI Dataset

Different from the simulated data, we do not have the fully defined ground-truth to exactly assess the performance of different causal search algorithm methods from real fMRI. Instead, we have partial knowledge about the presence of EC between brain regions on the medial temporal lobe from current work (Sanchez-Romero et al. 2019).

For the real fMRI data, we run MetaRLEC on every individual subject for the seven medial temporal lobe ROIs of the left and right hemispheres separately. We define the EC between two brain regions as actually discovered when we consider edges that appear in 35% of the 23 individual subjects (more than 8 subjects). Figure 3 illustrates the EC networks inferred by 9 methods from the left and right lobe. In Figure 3 (i), we can see that the EC networks of the left hemisphere medial temporal lobe are closely similar to the right and have some differences. These differences are mainly caused by the connections of $CA1$, $CA23DG$, SUB, ERC, and $BA35, 36$.

Compared with previous studies (Sanchez-Romero et al. 2019), overall, the EC network of the left hemisphere medial temporal lobe learned by MetaRLEC is similar to the EC networks estimated by Sanchez-Romero et al. (Sanchez-Romero et al. 2019) and has some differences. We only find one-way connections between $CA1$ and $CA23DG$ ($CA23DG \leftrightarrow CA1$), SUB and ERC ($ERC \leftrightarrow SUB$), ERC and PHC ($PHC \leftrightarrow ERC$), and ERC and $BA35$ ($BA35 \leftrightarrow ERC$). One possible explanation is that the BIC reward motivates the actor to generate a directed acyclic graph, which may limit the model's ability to discover cyclic dependencies between brain regions.

In addition, as suggested by Lavenex and Amaral (Lavenex and Amaral 2000), the flow of information from the medial temporal lobe cortices ($BA35, BA36, PHC$) directly into the entorhinal cortex (ERC) and travel to $CA23DG$ to $CA1$, this ($ERC \rightarrow CA23DG$) is the main pathway connecting the medial temporal lobe cortices with the hippocampus. It is worth noting that MetaRLEC infers the EC $ERC \rightarrow CA23DG$ in the left hemisphere. However, we also discovered reversal of some important brain ECs, such as the one-way connection $CA23DG \leftrightarrow SUB$, $SUB \leftrightarrow BA35$ and $CA1 \leftrightarrow BA36$. In contrast to other brain EC estimating methods, MetaRLEC estimates all brain EC, although some connections were one-way or reversed. Therefore, the new MetaRLEC method can provide a reliable perspective for the analysis of brain EC networks.

Conclusion

Estimating brain EC from high-noise small-sample fMRI time-series data is a challenging problem in the study of the brain connectome. In this paper, we propose a novel EC discovery method based on meta-reinforcement learning, called MetaRLEC. MetaRLEC first employs an actor to extract the features of brain regions and discover the brain EC network. Then, it utilizes a critic to evaluate the brain EC network and provide feedback. Finally, it leverages meta-critic to guide the actor to obtain maximum learning efficiency in high noise and small sample fMRI data. Experimental results on both synthetic and real datasets show that MetaRLEC performs well compared to the state-of-the-art methods, which shows that the meta-reinforcement learning approach has great development potential in brain EC discovery. In the future, we consider extending this work to learning large-scale brain EC networks.

Acknowledgments

This work was partly supported by National Natural Science Foundation of China Research Program (62106009, 62276010), in part by R&D Program of Beijing Municipal Education Commission (KM202210005030, KZ202210005009).

References

Bagherzadeh, S.; Shahabi, M. S.; and Shalbaf, A. 2022. Detection of schizophrenia using hybrid of deep learning and brain effective connectivity image from electroencephalogram signal. *Computers in Biology and Medicine*, 146: 105570.

Bechtle, S.; Molchanov, A.; Chebotar, Y.; Grefenstette, E.; Righetti, L.; Sukhatme, G.; and Meier, F. 2021. Meta learning via learned loss. In *2020 25th International Conference on Pattern Recognition (ICPR)*, 4161–4168. IEEE.

DSouza, A. M.; Abidin, A. Z.; Leistritz, L.; and Wismüller, A. 2017. Exploring connectivity with large-scale Granger causality on resting-state functional MRI. *Journal of neuroscience methods*, 287: 68–79.

Finn, C.; Abbeel, P.; and Levine, S. 2017. Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning*, 1126–1135. PMLR.

Friston, K. J. 2011. Functional and effective connectivity: a review. *Brain connectivity*, 1(1): 13–36.

Gupta, A.; Mendonca, R.; Liu, Y.; Abbeel, P.; and Levine, S. 2018. Meta-reinforcement learning of structured exploration strategies. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, 5307–5316.

Henderson, P.; Islam, R.; Bachman, P.; Pineau, J.; Precup, D.; and Meger, D. 2018. Deep reinforcement learning that matters. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32.

Huang, C.; Zhai, S.; Talbott, W.; Martin, M. B.; Sun, S.-Y.; Guestrin, C.; and Susskind, J. 2019. Addressing the loss-metric mismatch with adaptive loss alignment. In *International conference on machine learning*, 2891–2900. PMLR.

Hyvarinen, A. 2010. Pairwise measures of causal direction in linear non-gaussian acyclic models. In *Proceedings of 2nd Asian Conference on Machine Learning*, 1–16. JMLR Workshop and Conference Proceedings.

Ji, J.; Liu, J.; Han, L.; and Wang, F. 2021a. Estimating Effective Connectivity by Recurrent Generative Adversarial Networks. *IEEE Transactions on Medical Imaging*, 40(12): 3326–3336.

Ji, J.; Zou, A.; Liu, J.; Yang, C.; Zhang, X.; and Song, Y. 2021b. A Survey on Brain Effective Connectivity Network Learning. *IEEE Transactions on Neural Networks and Learning Systems*.

Jiang, Y.; Zhang, Y.; Nie, L.; Liu, H.; and Zheng, J. 2023. Identification and effective connections of core networks in patients with temporal lobe epilepsy and cognitive impairment: Granger causality analysis and multivariate pattern analysis. *International Journal of Neuroscience*, 133(9): 935–946.

Khan, D. M.; Yahya, N.; Kamel, N.; and Faye, I. 2023. A novel method for efficient estimation of brain effective connectivity in EEG. *Computer Methods and Programs in Biomedicine*, 228: 107242.

Kirsch, L.; Flennerhag, S.; van Hasselt, H.; Friesen, A.; Oh, J.; and Chen, Y. 2022. Introducing symmetries to black box meta reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 7202–7210.

Lavenex, P.; and Amaral, D. G. 2000. Hippocampal-neocortical interaction: A hierarchy of associativity. *Hippocampus*, 10(4): 420–430.

Li, H.; Yu, S.; and Principe, J. 2023. Causal Recurrent Variational Autoencoder for Medical Time Series Generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 8562–8570.

Liang, B.; Li, X.; Gui, L.; Fu, Y.; He, Y.; Yang, M.; and Xu, R. 2023. Few-shot aspect category sentiment analysis via meta-learning. *ACM Transactions on Information Systems*, 41(1): 1–31.

Liu, J.; Ji, J.; Xun, G.; Yao, L.; Huai, M.; and Zhang, A. 2020. EC-GAN: Inferring brain effective connectivity via generative adversarial networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 4852–4859.

Liu, J.; Ji, J.; Xun, G.; and Zhang, A. 2022. Inferring Effective Connectivity Networks From fMRI Time Series With a Temporal Entropy-Score. *IEEE Transactions on Neural Networks and Learning Systems*, 33(10): 5993–6006.

Lu, Y.; Liu, J.; Ji, J.; Lv, H.; and Huai, M. 2022. Brain Effective Connectivity Learning with Deep Reinforcement Learning. In *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 1664–1667. IEEE.

Mao, C. P.; Yang, H. J.; Zhang, Q. J.; Yang, Q. X.; and Li, X. H. 2022. Altered effective connectivity within the cingulo-frontal-parietal cognitive attention networks in chronic low back pain: A dynamic causal modeling study. *Brain Imaging and Behavior*, 16(4): 1516–1527.

Patel, R. S.; Bowman, F. D.; and Rilling, J. K. 2006. A Bayesian approach to determining connectivity of the human brain. *Human brain mapping*, 27(3): 267–276.

Pfarr, J.-K.; Brosch, K.; Meller, T.; Ringwald, K. G.; Schmitt, S.; Stein, F.; Meinert, S.; Grotegerd, D.; Thiel, K.; Lemke, H.; et al. 2021. Brain structural connectivity, anhedonia, and phenotypes of major depressive disorder: A structural equation model approach. *Human Brain Mapping*, 42(15): 5063–5074.

Sanchez, P.; Liu, X.; O'Neil, A. Q.; and Tsaftaris, S. A. 2023. Diffusion Models for Causal Discovery via Topological Ordering. In *The Eleventh International Conference on Learning Representations*.

Sanchez-Romero, R.; Ramsey, J. D.; Zhang, K.; Glymour, M. R.; Huang, B.; and Glymour, C. 2019. Estimating feedforward and feedback effective connections from fMRI time

series: Assessments of statistical methods. *Network Neuroscience*, 3(2): 274–306.

Shah, P.; Bassett, D. S.; Wisse, L. E.; Detre, J. A.; Stein, J. M.; Yushkevich, P. A.; Shinohara, R. T.; Pluta, J. B.; Valenciano, E.; Daffner, M.; et al. 2018. Mapping the structural and functional network architecture of the medial temporal lobe using 7T MRI. *Human Brain Mapping*, 39(2): 851–865.

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. Network modelling methods for FMRI. *Neuroimage*, 54(2): 875–891.

Sodhani, S.; Zhang, A.; and Pineau, J. 2021. Multi-task reinforcement learning with context-based representations. In *International Conference on Machine Learning*, 9767–9779. PMLR.

Tang, Y. 2022. Biased Gradient Estimate with Drastic Variance Reduction for Meta Reinforcement Learning. In *International Conference on Machine Learning*, 21050–21075. PMLR.

Wang, Q.; and Van Hoof, H. 2022. Model-based meta reinforcement learning using graph structured surrogate models and amortized policy search. In *International Conference on Machine Learning*, 23055–23077. PMLR.

Yuan, H.; and Lu, Z. 2022. Robust task representations for offline meta-reinforcement learning via contrastive learning. In *International Conference on Machine Learning*, 25747–25759. PMLR.

Zhang, K.; Zhu, S.; Kalander, M.; Ng, I.; Ye, J.; Chen, Z.; and Pan, L. 2021. gCastle: A Python Toolbox for Causal Discovery. arXiv:2111.15155.

Zhang, S.; Zheng, D.; Hu, X.; and Yang, M. 2015. Bidirectional long short-term memory networks for relation classification. In *Proceedings of the 29th Pacific Asia conference on language, information and computation*, 73–78.

Zheng, Z.; Oh, J.; and Singh, S. 2018. On learning intrinsic rewards for policy gradient methods. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, 4649–4659.