

DOI:10.11918/j.issn.0367-6234.201811073

基于萤火虫算法的脑效应连接网络学习方法

纪子龙¹, 冀俊忠¹, 刘金铎¹, 杨翠翠¹

(1. 北京工业大学 信息学部计算机学院多媒体与智能软件技术北京市重点实验室, 北京 100124)

摘要: 脑效应连接网络学习是人脑连接组研究的一个重要研究课题, 准确识别脑效应连接网络对于脑疾病的早期诊断以及病理研究具有重要意义。本文将萤火虫算法与贝叶斯网相结合, 提出了一种带有繁殖机制的脑效应连接网络萤火虫学习方法。新方法使用 K2 评分作为目标函数来衡量萤火虫个体的绝对亮度, 利用萤火虫种群的寻优来完成脑效应连接网络的学习, 并利用繁殖机制对种群实施进一步的优化。首先将一种仅含少数边的脑效应连接网络表示成一个萤火虫个体, 并通过萤火虫个体的定向移动操作以及随机移动操作逐步构建脑效应连接网络; 然后每经过一定代数的寻优后, 萤火虫种群执行一次繁殖过程, 以优化效应连接网络的质量。最后, 当算法收敛时, 将萤火虫种群中绝对亮度最高个体所代表的网络结构作为学习到的最优脑效应连接网络。在多组模拟数据集上的实验结果验证了新算法中繁殖机制的有效性, 且与其它算法相比, 新算法具有明显优势。在真实数据上的实验也表明了算法的潜在实用性。

关键词: 脑网络; 脑效应连接网络; 贝叶斯网; 萤火虫算法; 繁殖机制

中图分类号: TP311; TP18

文献标志码: A

文章编号: 0367-6234(2019)05-0076-09

Learning effective connectivity network structure based on firefly algorithm

Ji Zilong¹, Ji Junzhong¹, Liu Jinduo¹, Yang Cuicui¹

(1. Beijing Municipal Key Laboratory of Multimedia and Intelligent Software Technology
Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China)

Abstract: Learning brain effective connectivity (EC) networks is an important topic within the community of human brain connectome. It is of great significance for the early diagnosis and pathological study of brain diseases to accurately identify the brain EC network structure. This paper combines the Firefly Algorithm (FA) with Bayesian network, and proposes a new method to learn brain EC networks by FA with a reproductive mechanism. The new method uses K2 score as the evaluation method of absolute brightness of fireflies, uses the optimization of firefly population to complete the learning of brain EC networks, and uses reproductive mechanism to further optimize the population. First, a firefly individual represented a brain EC network with a few edges, which was gradually constructed through the directional movements and random movements of the firefly individual. Then, a reproductive mechanism was employed to optimize the quality of networks after a certain number of evolution iterations. Finally, the network structure represented by the individuals with the highest absolute brightness in the population was used as the learning brain EC network. Experimental results on many simulated datasets verified the effectiveness of the reproductive mechanism, and the new algorithm has obvious advantages on the whole performance compared with other algorithms. Experimental results on real datasets also show the potential practicability of the new algorithm.

Keywords: brain network; brain effective connectivity network; Bayesian network; firefly algorithm (FA); reproductive mechanism

脑效应连接网络是一种由节点和有向边构成的图模型, 其中节点表示脑区, 有向边表示脑区之间的效应连接。准确地从影像数据中学习脑效应连接网络, 不仅能够加深研究人员对人脑功能复杂性的理

解, 而且有助于进行脑疾病的早期诊断以及病理研究^[1-2]。因此从影像数据中识别高质量的脑效应连接网络, 已经成为近年来人脑功能网络研究的一个重要课题。

功能磁共振成像技术 (functional magnetic resonance imaging, fMRI) 是一种非介入的活体人脑功能成像技术, 具有无创性、高时空分辨率等优势, 自诞生以来迅速成为脑科学研究领域最主要的技术手段。迄今为止, 研究人员已经提出了多种从 fMRI 数据中进行脑效应连接网络识别的方法。这些方法

收稿日期: 2018-12-28

基金项目: 国家自然科学基金项目 (61672065); 北京市博士后工作经费资助项目 (2017-ZZ-024)

作者简介: 纪子龙 (1994—), 男, 硕士生;
冀俊忠 (1969—), 男, 教授, 博导

通信作者: 冀俊忠, jjz01@bjut.edu.cn

大致可分为两类:基于模型驱动的方法和基于数据驱动的方法^[3]. 其中,基于模型驱动的方法是一种验证性方法,主要包括结构方程模型(structural equation modeling, SEM)^[4]和动态因果模型(dynamic causal modeling, DCM)^[5]. 已有研究表明,该类方法在缺少先验的情况下并不适用,而且仅能构建小规模的大脑效应连接网络^[3]. 基于数据驱动的方法不需先验知识和假设,能够直接从数据中得到脑区之间的因果关系,在神经影像技术不断发展和数据日益丰富的今天,该类方法逐渐成为主流方法. 常用的数据驱动方法包括格兰杰因果模型(Granger Causality, GC)^[6-7]、线性非高斯无环模型(Linear non-Gaussian acyclic model, LiNGAM)^[8-9]、广义同步(Generalised synchronization, GS)^[10]、Patel 条件独立性方法(Patel)^[11]和贝叶斯网方法^[12-13]等. 其中,格兰杰因果模型使用矢量自回归模型通常能够在广义平稳和零均值数据下实现脑区间脑效应连接的有效估计,已经被广泛地应用于脑效应连接网络的构建. 但是该类方法存在对数据噪声和下采样过程比较敏感的缺陷^[14]. 线性非高斯无环模型使用高阶分布统计和独立成分分析来估计网络的连接. 但该方法的有效性高度依赖于如下3个假设:数据产生过程是线性的、不存在隐变量、干扰变量服从非零方差的非高斯分布. 广义同步方法是通过分析脑区信号间的独立关系来评估脑区信号间的相互依赖性. 该方法采用3种相关的非线性独立度量(S^k 、 H^k 和 N^k). 虽然这些非线性独立度量是有方向的,但是不对称的方向判断有时会产生前后的矛盾^[3]. Patel 的条件独立性方法首先计算两个脑区信号之间的条件依赖关系,然后从条件依赖关系中得到两个脑区连接方向的度量 τ 和连接强度的度量 κ ,最后利用 τ 和 κ 分别进行连接方向的识别以及连接强度的判断. 该方法与已有方法相比,在连接方向识别上具有一定优势,但在连接强度的判断上具有灵敏度不高的缺陷. 总之,尽管上述4种数据驱动的方法各自具有一定的适用性,但普遍存在对数据噪声敏感、识别准确率不高等缺陷.

近年来,基于贝叶斯网(Bayesian network, BN)的学习方法采用无监督的数据驱动方法构建脑效应连接网络,具有良好的灵活性和适应性,已经成为一个新的研究热点^[12]. 文献[3]中对多种基于BN的脑效应连接网络学习方法进行了对比测试,发现贪婪等价类搜索(greedy equivalence search, GES)的性能相对较好,但是该方法采用贪婪搜索技术,容易陷入局部最优. 为了克服上述缺陷,已有学者将全局随机搜索的群智能方法与贝叶斯网模型学习相结合来

完成脑效应连接网络的学习. 例如,文献[15]提出一种从fMRI数据中构建脑效应连接网络的人工免疫方法(AIAEC),该方法能够学习到较高精度的效应连接网络. 文献[16]提出了一种基于蚁群算法的脑效应连接网络学习方法(ACOEC),该方法不仅能够准确识别网络的连接和方向,而且能够有效量化网络的连接强度. 综上所述,脑效应连接网络识别的研究进行得如火如荼. 虽然群智能算法已经在脑效应连接网络学习中进行了有效尝试,但探索脑效应连接网络学习新方法仍是该领域极具挑战性的研究问题.

萤火虫算法(Firefly algorithm, FA)作为一种新型的群智能算法在多个领域已经表现出了较好的效果^[17-19],在脑效应连接网络学习问题上可能具有潜在的应用价值,故本文提出了一种带有繁殖机制的脑效应连接网络萤火虫学习方法(Firefly algorithm with reproductive mechanism for learning brain effective connectivity network, FAR-EC),新方法主要是通过萤火虫种群的迭代寻优机制,搜索最佳的脑效应连接网络结构. 具体来说,新方法在种群的每次迭代中,每个个体通过定向移动和随机移动操作逐步完成脑效应连接网络的构建. 在经过一定代数的寻优后,种群执行一次繁殖操作以进一步提高整体寻优的质量. 最后,将最优个体所代表的网络结构作为学习到的脑效应连接网络. 在模拟数据集上的实验证明了繁殖机制的有效性,而且与多个算法相比,新算法在多组数据上都能够得到质量更高的脑效应连接网络. 在真实数据集上的实验结果也表明,新算法具有较好的实用性.

1 萤火虫算法

萤火虫算法的核心思想是:绝对亮度低的萤火虫会被绝对亮度高的萤火虫吸引,根据位置更新公式更新原有位置,向高亮度位置移动,移动距离由吸引度大小决定,吸引度大小与相对亮度成比例. 通常,把待求解问题的目标函数值作为萤火虫的绝对亮度.

萤火虫 i 在萤火虫 j 所在位置处的光强度为萤火虫 i 对萤火虫 j 的相对亮度,记作 I_{ij} . 考虑到光传播过程中光的强度会随着距离的增加以及空气的吸收而减弱,故将其定义如下:

$$I_{ij} = I_i e^{-\gamma r_{ij}^2}. \quad (1)$$

式中: I_i 为萤火虫 i 的绝对亮度, γ 为光强度吸收系数, r_{ij} 为萤火虫 i 到 j 的距离.

$$r_{ij} = d(x_i - x_j) = \sqrt{\sum_{k=1}^d (x_{i,k} - x_{j,k})^2}. \quad (2)$$

假设萤火虫 i 被萤火虫 j 吸引且向 j 所处位置

移动. 吸引度的大小与相对亮度 I_{ij} 成比例, 相对亮度越大则吸引度越大, 吸引度 β_{ij} 的公式为

$$\beta_{ij} = \beta_0 e^{-\gamma r_{ij}^2}. \quad (3)$$

式中 β_0 为最大吸引度. 萤火虫 i 的位置更新公式为

$$x_i = x_i + \beta_{ij}(x_j - x_i) + \alpha \varepsilon_i. \quad (4)$$

式中: α 为步长因子, ε_i 为服从高斯分布或者均匀分布的随机数.

2 带有繁殖机制的脑效应连接网络萤火虫学习算法

带有繁殖机制的脑效应连接网络萤火虫学习算法 (FAR-EC) 的大致思路如下: 首先, 将每个萤火虫个体初始化为仅含少量边的脑效应连接网络结构, 并且使用 K2 评分作为目标函数来衡量萤火虫个体的绝对亮度大小; 然后, 每个个体通过定向移动以及随机移动完成脑效应连接网络的构建; 接着, 每隔一定代数, 执行一次繁殖过程来对种群进行优化. 最后, 得到满足要求的脑效应连接网络.

2.1 绝对亮度评价方法

K2 评分是一种常用的贝叶斯网结构评分度量标准, 它不仅能够评价候选网络结构与数据集的匹配程度, 而且具有可分解性, 即一个节点结构发生变化时, 仅需重新计算该节点评分. 由于本文是基于贝

叶斯网络的学习方法, 故使用 K2 评分作为目标函数来衡量萤火虫个体的绝对亮度, K2 评分值越大, 绝对亮度越高. 萤火虫 i 的绝对亮度计算公式如下

$$I_i = f(G_i, D_{\text{data}}) = \sum_{i=1}^n \sum_{j=1}^{q_i} \left(\log \left(\frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \right) + \sum_{k=1}^{r_i} \log(N_{ijk}) \right). \quad (5)$$

式中: G_i 为萤火虫 i 所代表的脑效应连接网络, D_{data} 为 fMRI 数据集, n 为网络中含有的脑区个数, q_i 是节点 B_i 的父节点可能取值的组合数, r_i 是节点 B_i 可能取值的数目, N_{ij} 是数据集中节点 B_i 取第 j 个值的实例数量, 例如节点 B_i 可能取值为 $\{0, 1\}$, 则 N_{i1} 代表数据集中节点 B_i 取值为第 1 个值 (取值为 0) 的实例数量, N_{ijk} 是数据集中节点 B_i 取第 j 个值, B_i 的父节点取第 k 个组合值时的实例数量.

2.2 萤火虫种群的初始化

萤火虫个体初始化过程见图 1, 图中 $G_i(m)$ 为萤火虫 i 所代表的网络结构中含有 m 条边. 萤火虫 i 从空图 $G_i(0)$ 开始, 在保证有向无环图的条件下不断加入使得图 K2 评分增大的边, 直到边数达到预先设定的边数为止, 得到萤火虫 i 的初始个体.

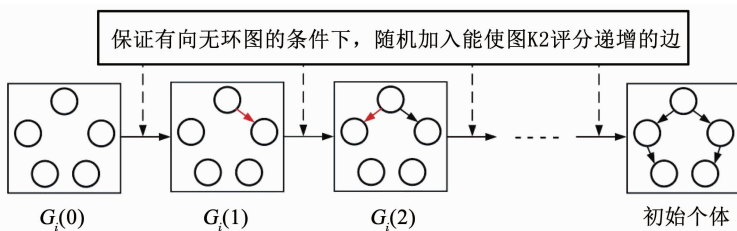


图 1 萤火虫个体 i 初始化实例

Fig. 1 Example of initialization of the firefly individual i

2.3 萤火虫个体的位置更新方式

假设萤火虫 i 向萤火虫 j 所处位置移动, 则萤火虫的位置更新方式具体可以描述为: 首先, 计算 i 与 j 之间的距离 r_{ij} , 即 i 与 j 所代表的脑效应连接网络中边不一致的数量; 然后, 计算 i 与 j 之间的吸引度 β_{ij} ; 最后, 将 β_{ij} 与移动阈值 p_m 进行比较, 若大于 p_m , i 进行定向移动操作, 反之, i 进行随机移动操作. 下面将具体介绍这两个操作.

定向移动操作: 见图 2, 假设萤火虫 i 向 j 移动, 在满足有向无环图的条件下, 萤火虫 i 随机选择图 G_i 中一条边 (虚线所标识边), 将其变为与 G_j 一致, 得到 i 的新位置 $G_{i,\text{new}}$.

环图的条件下, 随机进行减边、加边和反向边操作, 分别得到图结构 $G_{i,\text{del}}$ 、 $G_{i,\text{add}}$ 和 $G_{i,\text{rev}}$; 然后从中选择使萤火虫绝对亮度增加最多的操作, 即加边操作; 将执行加边操作所得到的图结构 $G_{i,\text{add}}$ 作为 i 的新位置 $G_{i,\text{new}}$.

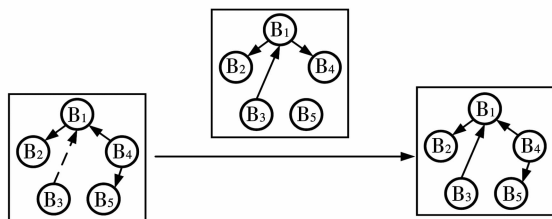


图 2 萤火虫个体 i 定向移动实例

Fig. 2 Example of directional movement of the firefly individual i

随机移动操作: 见图 3, 萤火虫 i 在满足有向无

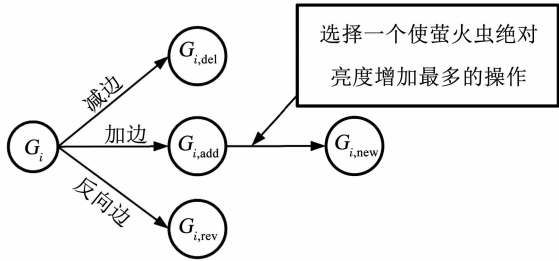


图 3 萤火虫个体 i 随机移动实例

Fig. 3 Example of random movement of the firefly individual i

2.4 繁殖机制

萤火虫成虫交配产生的幼虫只有成长为成虫后,才发出性别交流信号,进行下一代繁殖. 本文受上述现象启发,提出了一种繁殖机制,以此加强萤火虫之间的信息交流,以及提高种群的多样性. 繁殖机制主要由配偶选择、子代产生、子代成长和种群更新 4 个操作组成,下面将具体介绍这 4 个操作.

配偶选择:首先规定结为配偶的两萤火虫不再与种群中其余个体结对. 从当前绝对亮度最大个体 u 开始,将距离 u 最远个体 v 作为 u 的配偶.

子代产生:结为配偶的萤火虫分别向对方所处位置进行一次定向移动,将移动后所处的新位置,作为产生的两个子代的初生位置,此时子代为初生子代.

子代成长:初生子代进行一定次数的随机移动后加入到萤火虫种群中.

种群更新:设萤火虫种群规模为 a . 首先,去除种群中网络结构重复的个体;然后,判断去重后剩余个体数量,若大于 a ,则取绝对亮度最大的前 a 个个体作为下一代种群. 反之,则按照初始化过程,重复加入新个体到种群中,直到种群规模达到 a .

2.5 算法描述

FAR-EC 算法的核心思想是利用萤火虫个体的定向移动、随机移动和繁殖操作完成种群的进化. 具体的伪代码见算法 1.

算法 1: FAR-EC

Input: fMRI 数据集

Output: 脑效应连接网络

- 1) 初始化参数: a (种群规模)、 β_0 (最大吸引度)、 γ (光传播系数)、 l_{step} (隔代执行繁殖机制代数)、 p_m (移动阈值)
- 2) 产生初始萤火虫种群: 每个萤火虫个体被初始化为仅含有少数条边的脑效应连接网络结构
- 3) **While** 不满足终止条件
- 4) **for** $i = 1 : a$
- 5) **for** $j = 1 : a$
- 6) **if** $I_i < I_j$
- 7) 计算萤火虫 i, j 之间的距离 r_{ij} 以及吸引度 β_{ij}
- 8) **if** $\beta_{ij} > p_m$

- 9) 萤火虫 i 执行定向移动操作
- 10) **else**
- 11) 萤火虫 i 执行随机移动操作
- 12) **end**
- 13) **end**
- 14) 计算萤火虫 i 新位置的绝对亮度 I_i
- 15) **end**
- 16) **end**
- 17) **if** 当前迭代次数整除 l_{step}
- 18) 执行繁殖机制
- 19) **end**
- 20) **end**
- 21) 输出最亮个体所代表的脑效应连接网络

基于算法 1 的描述,下面对 FAR-EC 算法的时间复杂度进行分析. 种群初始化阶段时间复杂度为 $O(a \cdot n \cdot M)$, n 为网络中的脑区个数, a 为种群规模, M 为预先设定的边数. 萤火虫位置移动时间复杂度为 $O(n^2 \cdot C)$. 繁殖机制时间复杂度为 $O(a \cdot n^2 \cdot C \cdot L)$, L 为子代成长时所进行随机运动的次数. 则种群优化阶段的时间复杂度为 $O(L \cdot a \cdot n^2 \cdot C \cdot R + T \cdot a^2 \cdot n^2 \cdot C)$, T 为算法迭代次数, R 为执行繁殖机制次数. 故 FAR-EC 算法的时间复杂度为 $O(a \cdot n \cdot M + L \cdot a \cdot n^2 \cdot C \cdot R + T \cdot a^2 \cdot n^2 \cdot C)$. 由于 $T \cdot a^2 \cdot n^2 \cdot C > L \cdot a \cdot n^2 \cdot C \cdot R$ 且 $a \cdot n \cdot M < T \cdot a^2 \cdot n^2 \cdot C$, 故 FAR-EC 算法最终时间复杂度近似为 $O(T \cdot a^2 \cdot n^2 \cdot C)$.

3 试验结果及分析

3.1 数据集

3.1.1 模拟数据集

在本文中使用通用的由 Smith 等人在文献[3]中所给出的模拟 fMRI 数据集 (www.fmrib.ox.ac.uk/datasets/netstim/index.html). 该数据集每组数据均包含 50 个被试,其具体设置见表 1.

表 1 模拟数据集的具体参数

Tab. 1 Specific parameters of the simulated dataset

Sim	nodes	Session duration/min	TR/s	Noise/%	HRF, std. dev./s	Other factors
1	5	2.5	3.00	1.0	0.5	
2	5	5	3.00	1.0	0.5	
3	5	10	3.00	1.0	0.5	
4	5	60	3.00	1.0	0.5	
5	10	10	3.00	1.0	0.5	
6	15	10	3.00	1.0	0.5	
7	50	10	3.00	1.0	0.5	
8	5	10	3.00	1.0	0.5	2-group test
9	5	10	3.00	1.0	0.5	shared inputs
10	5	250	3.00	1.0	0.5	shared inputs
11	5	10	3.00	1.0	0.5	backwards connections
12	5	10	3.00	1.0	0.5	more connections

表 1 中, nodes 为节点数(脑区数); Session duration 为扫描时间; TR 为扫描重复时间; Noise 为噪声占比; HRF. std. dev. 为血氧动力学响应函数延迟时间标准差; Other factors 为其他因素, 主要包含: 2-group test(两组测试); shared inputs(外部输入); backwards connections(反向连接); more connections(稠密连接).

3.1.2 真实数据集

本文使用的真实 fMRI 数据来自于阿尔兹海默症神经影像学联盟(Alzheimer’s Disease neuroimaging Initiative, ADNI)数据集(www.loni.ucla.edu). 主要选取了两组被试, 分别是阿尔茨海默病(Alzheimer’s disease, AD)患者组和正常组, 被试的基本信息见表 2.

表 2 被试的基本信息
Tab. 2 Information of the subjects

指标	被试数量/个	男	女	年龄范围	平均年龄
正常组	50	27	23	70 ~ 85	77.6
AD 患者组	50	20	30	69 ~ 83	74.8

本文使用解剖学标记模板(anatomical automatic labeling template, AAL)将每幅图像分割成 116 个感兴趣区域(region of interest, ROI)然后根据已有文献选择被认为与 AD 病较为相关的 16 个 ROI, 其主要分布在额叶、顶叶、颞叶. 具体选择的 ROI 见表 3.

表 3 选择的感兴趣区域
Tab. 3 Selected regions of interest

叶区	感兴趣区域	编号
额叶	Frontal_Mid_L(左额中回)	1
	Frontal_Mid_R(右额中回)	2
	Frontal_Sup_L(左背外侧额上回)	3
	Frontal_Sup_R(右背外侧额上回)	4
顶叶	Cingulum_Post_L(左后扣带回)	5
	Cingulum_Post_R(右后扣带回)	6
	Precuneus_L(左楔前叶)	7
	Precuneus_R(右楔前叶)	8
顶叶	Parietal_Inf_L(左顶下缘角回)	9
	Parietal_Inf_R(右顶下缘角回)	10
颞叶	Temporal_Inf_L(左颞下回)	11
	Temporal_Inf_R(右颞下回)	12
	Hippocampus_L(左海马)	13
	Hippocampus_R(右海马)	14
	Angular_L(左角回)	15
	Angular_R(右角回)	16

由于 FAR-EC 算法不能直接使用连续的 fMRI 数据, 故需要对数据进行离散化的处理. 对于被试每个脑区的时间序列, 使用等频离散化方法将体素时

间序列离散成若干部分, 其中每个部分包含相同的数量的体素值.

3.2 评价标准

本文中使用精度(Precision)、召回率(Recall)、 F 度量(Fmeasure)来评价学习到的脑效应连接网络的连接和方向. 用 LN 表示学习到的网络, GN 表示真实网络. 连接的精度、召回率和 F 度量分别定义如下:

$$P_c = \frac{C_s}{C_s + C_a}, \tag{6}$$

$$R_c = \frac{C_s}{T_{tc}}, \tag{7}$$

$$F_c = \frac{2 \times P_c \times R_c}{P_c + R_c}. \tag{8}$$

式中: P_c 为连接的精度, R_c 为连接的召回率, F_c 为连接的 F 度量, C_s 为 LN 与 GN 中含有相同连接的数量, C_a 为存在于 LN 但不存在于 GN 的连接数量, T_{tc} 为 GN 含有的连接总数.

方向的精度、召回率和 F 度量分别定义如下:

$$P_d = \frac{D_s}{D_s + D_w + D_a}, \tag{9}$$

$$R_d = \frac{D_s}{T_{td}}, \tag{10}$$

$$F_d = \frac{2 \times P_d \times R_d}{P_d + R_d}. \tag{11}$$

式中: P_d 为方向的精度, R_d 为方向的召回率, F_d 为方向的 F 度量, D_s 为 LN 与 GN 具有相同方向连接的数量, D_w 为 LN 与 GN 中连接相同但方向不同的连接数量, D_a 为存在于 LN 但不存在于 GN 的方向连接的数量, T_{td} 为 GN 含有的方向总数.

3.3 参数对算法影响

本文在 9 个模拟数据集上主要测试了最大吸引度 β_0 、光吸收系数 γ 、移动阈值 p_m 和执行繁殖机制的隔代代数 l_{step} 对 FAR-EC 算法的影响. 在测试过程中保持其余参数不变, 种群规模设为 20, 仅变动待测试参数, 测试结果见图 4.

总体来看, 随着参数取值的变化 FAR-EC 算法的性能会发生相应的波动. 究其原因, 参数 β_0 、 γ 、 p_m 的变动会影响算法全局和局部搜索的平衡, 参数 l_{step} 的变动会影响算法进行繁殖机制的次数. 具体来看, 当 $\beta_0 = 0.5$ 、 $\gamma = 0.01$ 、 $p_m = 0.04$ 、 $l_{step} = 2$ 或 10 时算法对于方向的学习能力达到最强. 但值得注意的是, 参数 l_{step} 控制进行繁殖操作的频率, 而进行过多次繁殖操作会降低算法的效率.

综上可得, FAR-EC 算法最优参数设置如下: $a = 20$, $\beta_0 = 0.5$, $\gamma = 0.01$, $p_m = 0.04$, $l_{step} = 10$.

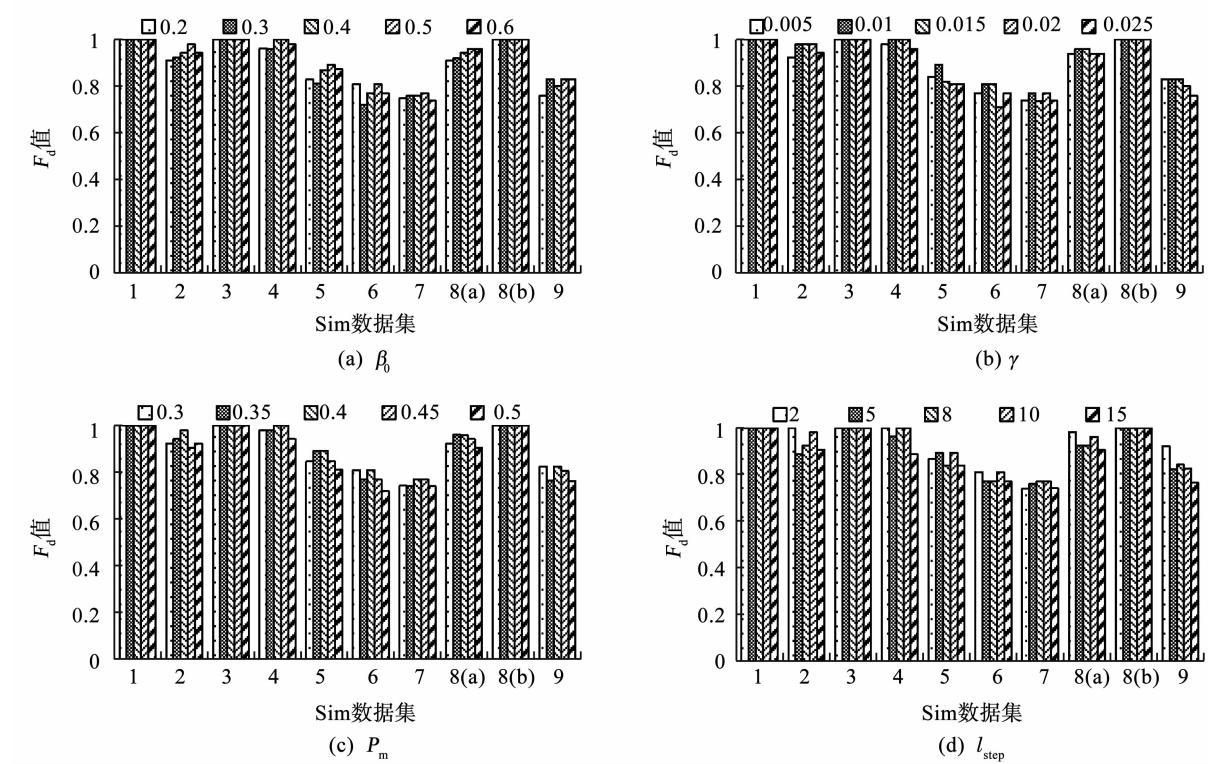


图4 参数发生变化时 FAR-EC 算法在 Sim1 ~ 9 上的 F_d 值比较
Fig. 4 Comparison of F_d by FAR-EC on Sim1 ~ 9 when parameters change

3.4 繁殖机制有效性检验

为评价繁殖机制对 FAR-EC 算法性能的影响, 本文在 9 组模拟数据上, 将 FAR-EC 与 FA-EC (不带繁殖机制的脑效应连接网络萤火虫学习方法) 进行了对比, 结果见图 5。

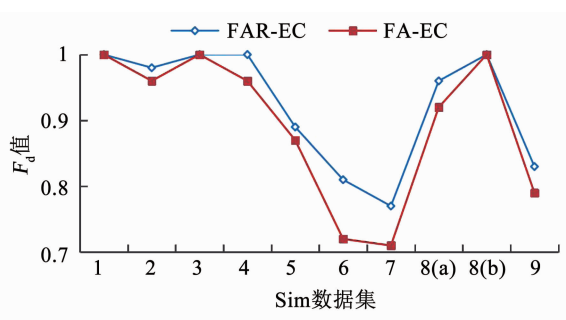


图5 两种算法在 Sim1 ~ 9 上运行结果的 F_d 值比较
Fig. 5 Comparison of F_d by two algorithms on Sim1 ~ 9

由于两种算法均能准确地识别连接, 故本文仅比较它们对于方向的识别能力. 总的来看, FAR-EC 算法在各组数据上的 F_d 值均不低于 FA-EC 算法, 说明繁殖机制对网络质量的提升具有一定作用, 特别是在脑区较多的 Sim6、7 上, 两种算法的 F_d 值相差较大, 说明随着脑区的增多, 繁殖机制的提升效果愈发显著。

3.5 与其它算法的性能比较

为展示 FAR-EC 算法的性能, 本文将其与 7 个

脑效应连接网络学习算法在 12 组模拟数据上进行了比较. 下面, 将具体分析不同实验条件下, FAR-EC 算法的性能。

样本数量: Sim1 ~ 4 中分别包含 2 500、5 000、10 000、60 000 条数据. 在 Sim1 ~ 4 上, FAR-EC 算法的 F_d 的值均保持相对稳定, 说明算法具有较好的稳定性. 在 Sim1 上, FAR-EC 算法的 F_d 值高于其余 7 种算法, 说明具有较强的小样本学习能力. 在 Sim4 上, FAR-EC 算法也能得到最优的结构, 说明具有较强的大样本学习能力。

节点数量: Sim5 ~ 7 分别包含 10、15、50 个节点. 随着节点数量的增加, 尽管 FAR-EC 算法对方向的识别能力有所下降, 但其依然能够准确的识别连接. 在 Sim5、7 上, FAR-EC 算法的 F_d 值均取得了前 3 位的成绩. 在 Sim6 上, 由于在本组数据上出现了多个评分相同的候选网络结构, FAR-EC 算法输出选择具有一定的随机性, 故性能有所下降, 但性能依然优于 GES、GC 算法。

连接强度: Sim8 将 50 个被试分为两组. 前 25 个为 Sim8(a), 后 25 个为 Sim8(b). Sim8(b) 将一条连接的连接强度值减半. FAR-EC 算法在 Sim8 上 F_c 、 F_d 值均取得了第 1 位的成绩, 说明在连接强度发生变化时 FAR-EC 算法能够准确的识别连接以及方向。

表 4 8 种算法在 Sim1 ~ 12 上的结果
Tab. 4 Results of eight algorithms on Sim1 ~ 12

Sim	F 度量	GES	LiNGAM	GC	GS	Patel	AIAEC	ACOECE	FAR-EC
1	F_c	1	1	0.83	1	1	1	1	1
	F_d	0.8	0.6	0.33	0.6	0.8	0.96	0.96	1
2	F_c	1	0.91	1	1	1	1	1	1
	F_d	1	0.73	0.6	0.8	0.98	0.98	0.98	0.98
3	F_c	1	0.91	1	1	1	1	1	1
	F_d	0.4	0.91	1	0.6	0.8	1	1	1
4	F_c	1	1	1	1	1	1	1	1
	F_d	0.8	1	0.6	1	1	1	1	1
5	F_c	1	0.96	1	1	1	1	1	1
	F_d	0.64	0.87	0.64	0.82	0.82	0.91	0.89	0.89
6	F_c	1	0.95	0.97	1	1	1	1	1
	F_d	0.5	0.84	0.65	0.89	0.84	0.84	0.86	0.81
7	F_c	1	0.98	0.95	1	1	1	1	1
	F_d	0.59	0.68	0.58	0.79	0.77	0.75	0.8	0.77
8(a)	F_c	1	1	0.89	1	1	1	1	1
	F_d	0.6	1	0.89	0.6	0.6	0.88	0.92	0.96
8(b)	F_c	1	1	1	1	1	1	1	1
	F_d	1	1	0.8	0.6	0.8	0.92	0.88	1
9	F_c	0.91	1	0.91	0.83	0.83	1	1	1
	F_d	0.55	0.8	0.36	0.5	0.67	0.76	0.80	0.83
10	F_c	0.77	1	0.91	0.77	0.83	1	1	1
	F_d	0.46	0.8	0.55	0.62	0.83	0.84	0.82	0.88
11	F_c	0.8	0.6	0.67	0.75	0.75	1	1	1
	F_d	0.8	0.4	0.22	0.25	0.5	0.64	0.64	0.64
12	F_c	1	0.93	0.93	0.93	0.93	1	1	1
	F_d	0.43	0.67	0.4	0.67	0.93	0.60	0.60	0.72

神经噪声:Sim9、10 将一些可看作神经噪声的外部输入加入到网络中,这使得所有算法对于连接以及方向的识别能力下降.但在此情况下,FAR-EC 算法性能优于其余算法,说明其具有较好的抗噪能力.

表 5 FAR-EC 与对比算法在模拟数据集上关于 F_d 值的 Friedman 检验结果

Tab. 5 Friedman test results of F_d by FAR-EC and contrast algorithm on the Smith dataset

	GES	LiNGAM	GC	GS	Patel	AIAEC	ACOECE
P 值	0.0348	0.048	0.001	0.035	0.020	0.157	0.257
显著性	yes	yes	yes	yes	yes	no	no

负向连接:Sim11 随机使一条连接存在两个方向,此时 8 种算法的性能均有所下降,但 FAR-EC 算法对连接和连接方向的识别能力依然处于前两位,依然具有较好的性能.

稠密连接:Sim12 在 Sim3 的基础上多加入了两条连接,使得网络到达一个连接稠密的状态.在 Sim12 上 FAR-EC 算法能够准确识别连接,对于方

向的识别能力仅次于 Patel.

由于多数算法对于边的识别能力较好,故本文仅给出了在 5% 显著水平下 FAR-EC 算法与对比算法关于 F_d 值的 Friedman 检验结果,具体见表 5. 其中, yes 表示存在显著性差异, no 表示不存在显著性差异. 由表 5 得, FAR-EC 算法对于方向的识别能力显著优于除 AIAEC、ACOECE 外的其余 5 种算法. 以上可见, 对比其它 7 种算法, FAR-EC 算法具有优良的性能, 是一种很具有竞争力的算法.

3.6 真实数据集上的结果与分析

使用 FAR-EC 算法在真实数据集上学习到的脑效应连接网络见图 6, 图 6(a) 为正常人脑效应连接网络, 图 6(b) 为 AD 患者脑效应连接网络.

首先统计了两个脑效应连接网络的连接总数, 得到 AD 患者连接数为 38, 正常人连接数为 44, 显然 AD 患者存在的连接数要小于正常人, 这种失连接的现象与已有文献[13、20]得到的结果相似, 可看作 AD 的病理生物学标记. 进一步, 已知认知以及记忆功能减弱是 AD 的症状之一, 而海马区(图 6 中

13、14 节点)主要负责信息的临时存储,同时还可将短时的程序性记忆转化为长时记忆,故海马对 AD 病情的发展有重大的影响. 通过对比图 6 中两个脑效应连接网络可得,AD 患者的脑效应连接网络中海马与其它 ROI 之间的连接明显少于正常被试的数量,存在明显的失连接. 另外,还可发现 AD 患者脑效应连接网络中后扣带回(图 6 中 5、6 节点)与

其它 ROI 与之间的连接数明显小于正常被试,而在静息态下后扣带回是新陈代谢最活跃的脑区,它参与情景记忆的提取,是 AD 最早发生新陈代谢降低的区域. 上述研究结果表明,FAR-EC 算法在真实数据集上可以学习到合理的脑效应连接网络,再次反映了算法的有效性.

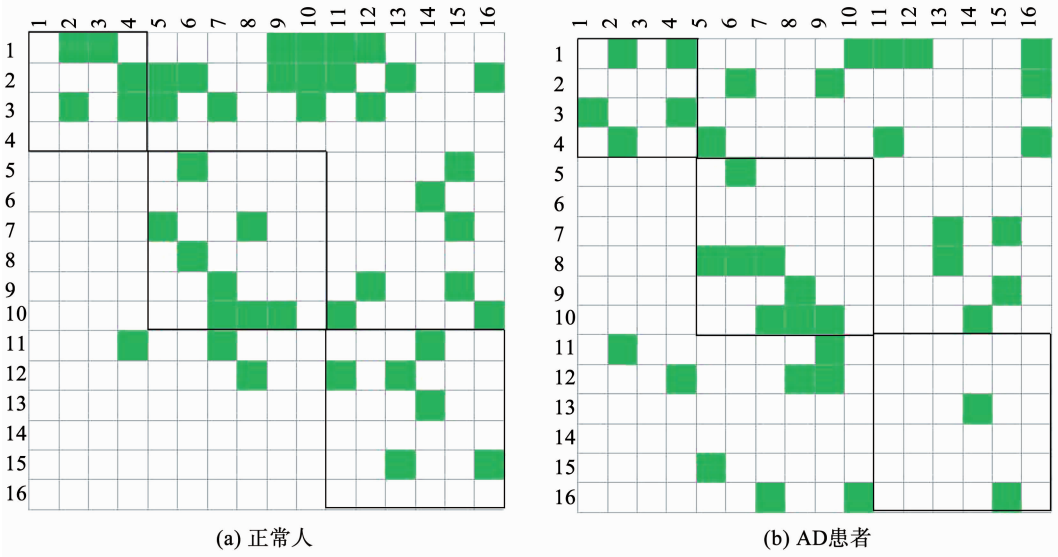


图 6 FAR-EC 在真实数据集上构建的脑效应连接网络
Fig.6 EC network constructed by FAR-EC on real dataset

4 结 论

针对脑效应连接网络学习问题,本文基于萤火虫算法及其繁殖机制提出了一种新方法,即 FAR-EC. 该方法使用 K2 评分衡量萤火虫的绝对亮度,通过萤火虫的吸引移动操作,逐步构建脑效应连接网络. 同时,本文通过模拟萤火虫的繁殖行为提出了一种繁殖机制,以此来进一步优化种群的质量. 在 Simth 数据集上的实验证实了繁殖机制的有效性,且与同类算法相比,FAR-EC 算法具有一定优势. 而在 ADNI 数据集上的实验表明,FAR-EC 算法在 AD 病理研究中具有潜在的应用价值.

本文方法是基于群智能的脑效应连接网络学习的一个新探索,为构建高质量的脑效应连接网络提供了一个有效的方法,在多个模拟数据集的实验表明,本文方法求得的脑效应连接网络质量较其它算法有所提升,但当脑区数量增多时,求解精度有所下降,因此如何提高大规模脑效应连接网络求解精度,将是下一步研究的主要方向.

参考文献

[1] HEIN G, MORISHIMA Y, LEIBERG S, et al. The brain's functional network architecture reveals human motives [J].

Science, 2016, 351 (6277): 1074. DOI: 10. 1126/science. aac7992
[2]左西年,张喆,贺永,等. 人脑功能连接组:方法学、发展轨线
和行为关联[J]. 科学通报, 2012, 57 (35): 3399. DOI:10. 1360/972012 - 702
ZUOXinian, ZHANG Zhe, HE Yong, et al. The human functional connectome: Its methodology, developmental trajectory and behavioral association (in Chinese) [J]. Chinese Science Bullitin (Chinese Version), 2012, 57 (35): 3399. DOI:10. 1360/972012 - 702
[3]SMITH S M, MILLER K L, SALIMI K G, et al. Network modeling methods for fMRI[J]. Neuroimage, 2011, 54 (2): 875. DOI:10. 1016/j. neuroimage. 2010. 08. 063
[4]SAMDI S B, TING C M, OMBAO H, et al. A unified estimation framework for state-related changes in effective brain connectivity [J]. IEEE Transactions on Bbio-medical Engineering, 2016, 64 (4): 844. DOI:10. 1109/TBME. 2016. 2580738
[5]FRASSLE S, LOMAKINA E I, RAZI A, et al. Regression DCM for fMRI[J]. NeuroImage, 2017, 155: 406. DOI:10. 1016/j. neuroimage. 2017. 02. 090
[6]FAROKHZADI M, HOSSEIN Z G A, SOLTANIAN Z H. Nonlinear effective connectivity measure based on adaptive Neuro Fuzzy Inference System and Granger Causality [J]. NeuroImage, 2018, 181: 382. DOI:10. 1016/j. neuroimage. 2018. 07. 024
[7]SETH A K. A MATLAB toolbox for Granger causal connectivity analysis[J]. Journal of Neuroscience Methods, 2010, 186 (2): 262. DOI:10. 1016/j. jneumeth. 2009. 11. 020
[8]XU L, FAN T, WU X, et al. A pooling-LiNGAM algorithm for

- effective connectivity analysis of fMRI data [J]. *Frontiers in Computational Neuroscience*, 2014, 8 (8): 125. DOI:10.3389/fncom.2014.00125
- [9] LIU Y, WU X, ZHANG J, et al. Altered effective connectivity model in the default mode network between bipolar and unipolar depression based on resting-state fMRI [J]. *Journal of Affective Disorders*, 2015, 182: 8. DOI:10.1016/j.jad.2015.04.009
- [10] DAUWELS J, VIALATTE F, MUSHA T, et al. A comparative study of synchrony measures for the early diagnosis of Alzheimer's disease based on EEG [J]. *Neuroimage*, 2010, 49 (1): 668. DOI:10.1016/j.neuroimage.2009.06.056
- [11] PATEL R S, BOWMAN D B, RILLING J K. A Bayesian approach to determining connectivity of the human brain [J]. *Human Brain Mapping*, 2010, 27 (3): 267. DOI:10.1002/hbm.20182
- [12] IDE J S, ZHANG S, LI C S R. Bayesian network models in brain functional connectivity analysis [J]. *International Journal of Approximate Reasoning Official Publication of the North American Fuzzy Information Processing Society*, 2014, 55 (1): 23. DOI:10.1016/j.ijar.2013.03.013
- [13] HUANG S, LI J, YE J, et al. A sparse structure learning algorithm for Gaussian Bayesian Network identification from high-dimensional data [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2013, 35 (6): 1328. DOI:10.1109/tpami.2012.129
- [14] GOEBEL R, ROEBROECK A, KIM D S, et al. Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping [J]. *Magnetic Resonance Imaging*, 2003, 21 (10): 1251. DOI:10.1016/j.mri.2003.08.026
- [15] JI J, LIU J, LIANG P, et al. Learning effective connectivity network structure from fMRI data based on artificial immune algorithm [J]. *PLOS ONE*, 2016, 11 (4): e0152600. DOI:10.1371/journal.pone.0152600
- [16] LIU J, JI J, ZHANG A, et al. An ant colony optimization algorithm for learning brain effective connectivity network from fMRI data [C]//*IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. Shenzhen: IEEE Press, 2016: 360. DII: 10.1109/BIBM.2016.7822546
- [17] 谢承旺, 肖驰, 丁立新, 等. HMOFA: 一种混合型多目标萤火虫算法 [J]. *软件学报*, 2018, 29 (4): 1143. DOI: 10.13328/j.cnki.jos.005275
- XIE Chengwang, XIAO Chi, DING Lixin, et al. HMOFA: A hybrid multi-objective firefly algorithm [J]. *Journal of Software*, 2018, 29 (4): 1143. DOI:10.13328/j.cnki.jos.005275
- [18] LEI X, WANG F, WU F X, et al. Protein complex identification through Markov clustering with firefly algorithm on dynamic protein-protein interaction networks [J]. *Information Sciences*, 2016, 329 (6): 303. DOI:10.1016/j.ins.2015.09.028
- [19] 田梦楚, 薄煜明, 陈志敏, 等. 萤火虫算法智能优化粒子滤波 [J]. *自动化学报*, 2016, 42 (1): 89. DOI: 10.16383/j.aas.2016.c150221
- TIAN Mengchu, BO Yuming, CHEN Zhimin, et al. Firefly algorithm intelligence optimized particle filter [J]. *Acta Automatica Sinica*, 2016, 42 (1): 89. DOI:10.16383/j.aas.2016.c150221
- [20] WU X, LI R, FLEISHER A S, et al. Altered default mode network connectivity in Alzheimer's disease-A resting functional MRI and Bayesian network study [J]. *Human Brain Mapping*, 2011, 32 (11): 1868. DOI:10.1002/hbm.21153

(编辑 苗秀芝)