

控制与决策

Control and Decision

基于邻域决策粗糙集的脑功能连接生物标记物识别

冀俊忠, 龙腾, 杨翠翠

引用本文:

冀俊忠, 龙腾, 杨翠翠. 基于邻域决策粗糙集的脑功能连接生物标记物识别[J]. *控制与决策*, 2023, 38(4): 1092–1100.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2021.1566>

您可能感兴趣的其他文章

Articles you may be interested in

基于脑电多特征融合的癫痫发作预测方法

A seizure prediction method based on EEG multi-feature fusion

控制与决策. 2023, 38(1): 161–170 <https://doi.org/10.13195/j.kzyjc.2021.0400>

基于增强深度特征和TSK模糊分类器的癫痫脑电信号识别

TSK fuzzy classifier based on enhanced deep feature for epilepsy EEG signal recognition

控制与决策. 2023, 38(1): 171–180 <https://doi.org/10.13195/j.kzyjc.2021.1031>

基于矩阵的混合型邻域决策粗糙集增量式更新算法

Incremental updating algorithms of neighborhood decision-theoretic rough set model for hybrid data based on matrix

控制与决策. 2022, 37(6): 1621–1631 <https://doi.org/10.13195/j.kzyjc.2020.1371>

基于时域扩张残差网络和双分支结构的人体行为识别

Human behavior recognition based on time domain extended residual network and dual branching structure

控制与决策. 2022, 37(11): 2993–3002 <https://doi.org/10.13195/j.kzyjc.2021.0648>

基于知识粒度特征的多目标粗糙集属性约简算法

Multi objective rough set attribute reduction algorithm based on characteristics of knowledge granularity

控制与决策. 2021, 36(1): 196–205 <https://doi.org/10.13195/j.kzyjc.2019.0490>

基于邻域决策粗糙集的脑功能连接生物标记物识别

冀俊忠^{1,2,3}, 龙腾^{1,2}, 杨翠翠^{1,2†}

(1. 北京工业大学 信息学部, 北京 100124; 2. 多媒体与智能软件技术北京市
重点实验室, 北京 100124; 3. 北京人工智能研究院, 北京 100124)

摘要: 脑功能连接判别特征可以作为神经精神类疾病诊断的一种生物标记物, 利用机器学习方法对其进行识别是脑科学研究中的一项重要课题. 已有脑功能连接生物标记物的识别方法大都忽视了脑功能连接数据高维、连续、多噪声的特性对识别性能的影响, 导致所得到的生物标记物的分类能力不强. 对此, 提出一种基于邻域决策粗糙集的脑功能连接生物标记物识别方法. 首先, 针对脑功能连接数据连续性和高噪声的特点, 引入能有效处理连续和高噪声数据的邻域决策粗糙集来识别分类能力更强的脑功能连接判别特征作为生物标记物; 其次, 针对脑功能连接数据的高维特点, 通过快速生成邻域和缩小特征搜索空间来保证邻域决策粗糙集识别脑功能连接生物标记物的效率. 在 ABIDE I 和 ADNI 数据集上的实验结果表明, 所提出方法能够准确快速地获得分类能力更强的脑功能连接判别特征, 有望为神经精神类疾病的诊断提供更准确的生物标记物.

关键词: 脑功能连接; 邻域决策粗糙集; 特征约简; 特征可分性度量; 哈希映射; 生物标记物

中图分类号: TP391.4

文献标志码: A

DOI: 10.13195/j.kzyjc.2021.1566

开放科学(资源服务)标识码(OSID):



引用格式: 冀俊忠, 龙腾, 杨翠翠. 基于邻域决策粗糙集的脑功能连接生物标记物识别[J]. 控制与决策, 2023, 38(4): 1092-1100.

Identifying brain functional connectivity biomarkers based on neighborhood decision rough set

Ji Jun-zhong^{1,2,3}, LONG Teng^{1,2}, YANG Cui-cui^{1,2†}

(1. Department of Information Science, Beijing University of Technology, Beijing 100124, China; 2. Beijing Municipal Key Laboratory of Multimedia and Intelligent Software Technology, Beijing 100124, China; 3. Beijing Artificial Intelligence Institute, Beijing 100124, China)

Abstract: Discriminant characteristics of brain functional connectivity can be used as a biomarker for the diagnosis of neuropsychiatric diseases. Using a machine learning method to identify is an important topic in brain science research. Most of the existing recognition methods of brain functional connectivity biomarkers ignore the impact of the characteristics of high-dimensional, continuous and multi noise of brain functional connectivity data on the recognition performance, resulting in the weak classification ability of the obtained biomarkers. This paper proposes a brain function connectivity biomarker recognition method based on neighborhood decision rough sets. Firstly, according to the characteristics of continuity and high noise of brain function connectivity data, a neighborhood decision rough set, which can effectively deal with continuous and high noise data, is introduced to identify brain functional connectivity discriminant features with stronger classification ability as biomarkers. Then, according to the high-dimensional characteristics of brain function connectivity data, the efficiency of identifying brain function connectivity biomarkers by the neighborhood decision rough set is guaranteed by quickly generating the neighborhood and reducing the feature search space. The experimental results on the ABIDE I and ADNI data sets show that the proposed method can accurately and quickly obtain the discriminant features of brain functional connectivity with stronger classification ability, and is expected to provide more accurate biomarkers for the diagnosis of neuropsychiatric diseases.

Keywords: brain function connectivity; neighborhood decision rough set; feature reduction; feature separability measurement; Hash mapping; biomarkers

收稿日期: 2021-09-07; 录用日期: 2022-02-25.

基金项目: 国家自然科学基金项目(61906010, 61672065); 北京市教委科技计划一般项目(KM202010005032).

责任编辑: 徐泽水.

†通讯作者. E-mail: yangcc@bjut.edu.cn.

0 引言

近期脑科学研究发现许多神经精神类疾病与脑功能连接的异常变化有关,表明脑功能连接可作为神经精神类疾病诊断的一种有效的生物标记物^[1-2]. 于是,近年来面向神经精神类疾病诊断任务,基于功能磁共振成像(functional magnetic resonance imaging, fMRI)技术的脑功能连接分类引起了研究人员的极大兴趣^[3-5],研究人员希望通过计算机辅助技术来提高神经精神类疾病诊断的准确性. 然而,分类方法的准确性很大程度上依赖于所使用的脑功能连接判别特征,因此,如何从高维的脑功能连接数据中识别出脑功能连接判别特征成为了脑科学领域研究的一项重要课题.

目前,脑功能连接判别特征的识别方法主要包括特征提取和特征选择两类方法. 其中,特征提取方法通过某种线性变换将脑功能连接数据从高维特征空间映射到低维特征空间,并把低维空间的特征作为判别特征. 例如,文献[6]采用主成分分析方法(principal component analysis, PCA)提取鉴别性特征,进而将中风患者分类到不同的治疗阶段. 文献[7]采用局部线性嵌入方法(locally linear embedding, LLE)获取低维特征,并通过自组织功能磁共振聚类来区分精神分裂症患者和正常被测试者. 特征提取方法对脑功能连接特征形式进行了转换,获得的判别特征不易理解,不利于医生进行疾病诊断和给出神经学解释. 特征选择方法依据度量指标从高维的脑功能连接特征中挑选若干个度量值高的特征作为判别特征,不会改变脑功能连接的原始特征形式. 例如,文献[8]使用非模型打分方法(*F-score*)衡量每个脑功能连接特征的分类判别能力,并选择评分较高的若干脑功能连接特征进行脑疾病的分类研究. 文献[9]使用最大相关最小冗余方法(max-relevance and min-redundancy, mRMR)选择重要的脑功能连接特征,并将其作为生物标记物. 尽管特征选择方法能够克服特征提取方法改变脑功能连接特征形式的不足,但它只关注单个特征的重要性,往往不考虑特征之间的相关性,制约了获得的脑功能连接判别特征的分类能力.

粗糙集是一种处理不精确、不一致和不确定知识的数学工具,它能够在不改变知识系统原有特征形式且充分考虑特征相关性的前提下发现关键的特征. 并且,粗糙集理论经过多年的发展已经包含了处理各种数据类型的模型,如处理连续型数据的邻域粗糙集^[10],处理噪声数据的决策粗糙集^[11],处理连续噪声数据的邻域决策粗糙集^[12]等. 其中邻域决策粗糙

集由于结合了邻域粗糙集处理连续型数据和决策粗糙集处理噪声数据的优势,近年来逐渐获得了一些学者的关注. 文献[13]以减小边界区域、降低总体决策代价以及增大邻域半径为原则,结合多目标优化的方法对邻域决策粗糙集中的邻域半径和概率阈值进行了求解;文献[14]提出混合型信息系统下的邻域决策粗糙集模型,并给出了该模型的增量式更新算法;文献[15]将局部策略引入邻域决策粗糙集,以识别异常数据,并利用粒子群优化算法为模型提供最优的邻域粒度. 这些研究工作显示了邻域决策粗糙集在特征约简、增量学习以及异常数据检测等数据挖掘领域的优异性能.

鉴于脑功能连接是体现脑区之间功能关联程度的连续型数据,且因各种主客观因素的影响含有许多噪声,本文提出一种基于邻域决策粗糙集的脑功能连接生物标记物识别方法(Identifying Brain Functional Connectivity Biomarkers Based on Neighborhood Decision Rough Set, BNDRS). 具体来说,本文根据邻域决策粗糙集的正域大小来衡量脑功能连接特征子集的分类能力,采用添加-删除操作^[16]搜索脑功能连接判别特征. 为了保证搜索效率,本文首先利用哈希映射^[17]和邻域关系的对称性^[18]快速得到被试的邻域;其次基于特征可分性度量^[19]对脑功能连接特征进行分组来产生粗粒度特征搜索空间. 在ABIDE I和ADNI数据集上与多个方法的实验结果表明,本文所提出方法能够高效地获得分类能力更强的脑功能连接判别特征,有望为神经精神类疾病的诊断和治疗提供更加准确的生物标记物.

本文剩余部分的组织结构如下:第1节介绍邻域决策粗糙集的基本知识;第2节详细阐述本文所提出的基于邻域决策粗糙集的脑功能连接生物标记物识别方法;第3节通过实验验证所提出方法的有效性;第4节对本文工作进行总结并指出下一步的研究方向.

1 邻域决策粗糙集

在邻域决策粗糙集中,决策表由一个四元组 $S = \langle U, A_t = C \cup D, \{V_a | a \in A_t\}, \{I_a | a \in A_t\} \rangle$ 表示. 其中: U 是非空有限的样本集合; A_t 是非空有限的特征集合; C 是条件特征的集合; D 是决策特征的集合; V_a 是特征 $a \in A_t$ 的特征值的集合; $I_a: U \rightarrow V_a$ 是映射函数,表示样本与其特征值之间的映射关系. 样本 x 的等价类用其邻域 $\delta_B(x)$ 表示,邻域 $\delta_B(x)$ 的计算公式如下:

$$\delta_B(x) = \{x_i | x_i \in U, d_B(x, x_i) \leq \delta\}. \quad (1)$$

其中: B 是条件特征子集, 满足 $B \subseteq C$, δ 表示邻域半径, $d_B(x, x_i)$ 表示样本 x 与 x_i 关于条件特征子集 B 的距离, $d_B(x, x_i)$ 通常使用如下欧氏距离进行计算:

$$d_B(x, x_i) = \left(\sum_{j=1}^m |I_{a_j}(x) - I_{a_j}(x_i)|^2 \right)^{\frac{1}{2}}. \quad (2)$$

m 表示条件特征子集 B 中包含的特征数量. 对于样本集 $X \subseteq U$, 条件特征子集 $B \subseteq C$, 任意样本 x 在条件特征子集 B 下属于样本集 X 的概率为

$$P(X|\delta_B(x)) = \frac{|\delta_B(x) \cap X|}{|\delta_B(x)|}, \quad (3)$$

其中 $|\cdot|$ 表示样本集 “ \cdot ” 中包含的样本数量.

邻域决策粗糙集中任意样本集 X 可能存在与条件特征子集 B 相关的 3 个域: 正域 $\text{POS}_B(X)$ 、边界域 $\text{BND}_B(X)$ 和负域 $\text{NEG}_B(X)$, 任意样本 x 对应于这 3 个域, 可构造 3 个决策动作, a_P 、 a_B 和 a_N . 不同决策动作将会产生不同的决策代价, 通常使用 λ_{PP} 、 λ_{BP} 和 λ_{NP} 分别表示当样本 x 属于样本集 X 时, 采取决策动作 a_P 、 a_B 和 a_N 产生的决策代价; λ_{PN} 、 λ_{BN} 和 λ_{NN} 分别表示当样本 x 不属于样本集 X 时, 采取决策动作 a_P 、 a_B 和 a_N 产生的决策代价. 基于最小决策代价的原则, 邻域决策粗糙集使用以下决策规则判定样本 x 所属的域:

- 1) 若 $P(X|\delta_B(x)) \geq \alpha$, 则判定 $x \in \text{POS}_B(X)$;
- 2) 若 $\beta < P(X|\delta_B(x)) < \alpha$, 则判定 $x \in \text{BND}_B(X)$;
- 3) 若 $P(X|\delta_B(x)) \leq \beta$, 则判定 $x \in \text{NEG}_B(X)$.

其中: 概率阈值 α 、 β 的计算公式如下:

$$\alpha = \frac{\lambda_{PN} - \lambda_{BN}}{(\lambda_{PN} - \lambda_{BN}) + (\lambda_{BP} - \lambda_{PP})}, \quad (4)$$

$$\beta = \frac{\lambda_{BN} - \lambda_{NN}}{(\lambda_{BN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{BP})}. \quad (5)$$

决策代价满足以下条件^[11]:

$$\frac{\lambda_{NP} - \lambda_{BP}}{\lambda_{BN} - \lambda_{NN}} > \frac{\lambda_{BP} - \lambda_{PP}}{\lambda_{PN} - \lambda_{BN}}. \quad (6)$$

若用 $\pi_D = \{D_1, D_2, \dots, D_m\}$ 表示根据决策特征 D 得到的决策类的集合, m 表示决策类的数量, 则决策类的集合 π_D 关于特征子集 B 的正域、边界域和负域分别为

$$\text{POS}_B(\pi_D) = \bigcup_{1 \leq i \leq m} \text{POS}_B(D_i), \quad (7)$$

$$\text{BND}_B(\pi_D) = \bigcup_{1 \leq i \leq m} \text{BND}_B(D_i), \quad (8)$$

$$\text{NEG}_B(\pi_D) = U - \text{POS}_B(\pi_D) \cup \text{BND}_B(\pi_D). \quad (9)$$

邻域决策粗糙集可基于正域衡量相应特征子集的判别能力, 并利用添加-删除操作来寻找分类能力强的特征子集. 令 R 为当前特征子集, 则向当前特征子集 R 添加特征时, 特征子集 $C-R$ 中任意特征 c 的重要度为

$$\text{SIG}(c) = |\text{POS}_{R \cup \{c\}}(\pi_D)| - |\text{POS}_R(\pi_D)|. \quad (10)$$

从当前特征子集 R 中删除特征时, 当前特征子集 R 中任意特征 r 的重要度为

$$\text{SIG}(r) = |\text{POS}_R(\pi_D)| - |\text{POS}_{R - \{r\}}(\pi_D)|. \quad (11)$$

上述特征约简过程的时间复杂度为 $O(|C||U|^2)$, 可见, 邻域决策粗糙集进行特征约简的时间复杂度较高. 因此, 若利用邻域决策粗糙集识别与神经精神疾病相关的脑功能连接生物标记物, 则尽管有望提升所得脑功能连接生物标记物的准确率, 但识别效率较低. 为此, 本文所提出方法采取了有效的策略, 同时保证利用邻域决策粗糙集识别脑功能连接生物标记物的准确率和效率.

2 基于邻域决策粗糙集的脑功能连接生物标记物识别方法

本节将详细介绍本文所提出的基于邻域决策粗糙集的脑功能连接生物标记物识别方法 (BNDRS). BNDRS 根据邻域决策粗糙集正域的大小衡量相应脑功能连接特征子集的分类能力, 采用添加-删除操作搜索脑功能连接判别特征, 并通过快速生成邻域和缩小特征搜索空间来保证其时间性能. BNDRS 的流程如图 1 所示, 接下来将对 BNDRS 包含的主要步骤和机制进行详细介绍.

2.1 基于哈希映射和邻域对称性的快速邻域生成

为了减小生成被试邻域的时间复杂度, BNDRS 对被试进行哈希映射, 将每个被试映射到以数组作为存储结构的哈希表中. 对于脑功能连接特征子集 B , 任意被试 x 映射的数组在哈希表中的位置为

$$H_B(x) = \left\lceil \frac{d_B(x, x_0)}{\delta} \right\rceil. \quad (12)$$

其中: $d_B(x, x_0)$ 表示被试 x 与参照被试 x_0 关于脑功能连接特征子集 B 的欧氏距离, $\lceil d_B(x, x_0)/\delta \rceil$ 表示对 $d_B(x, x_0)/\delta$ 向上取整. 参照被试 x_0 是由脑功能连接特征集 B 中每一个脑功能连接特征的最小值组成的被试, 满足对于 $\forall b \in B, x_i \in U$, 有 $I_b(x_0) = \min[I_b(x_i)]$ 成立. 其中 $\min[I_b(x_i)]$ 表示脑功能连接特征 b 的最小值. 映射到同一个数组的被试根据映射的先后次序依次进行存储, 从而避免哈希冲突. 所有被试均存储至相应数组后, 哈希表中第 k 个位置的数组包含的被试为

$$A_k = \{x_i | x_i \in U, k = H_B(x_i)\}. \quad (13)$$

基于上面的哈希映射, 可以证明任意被试的邻域至多存在于该被试所在的数组以及与该被试相邻的数组. 下面给出上述结论的证明过程.

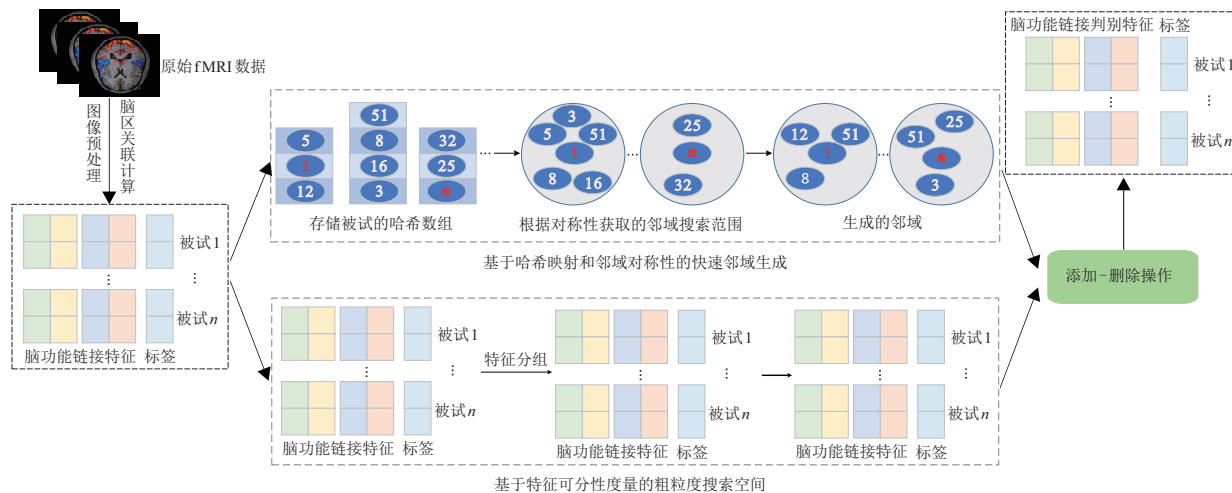


图1 本文所提BNDRS方法的示意图

证明 对于任意被试 $x \in A_k$, 有

$$k\delta \geq d_B(x, x_0) > (k-1)\delta. \quad (14)$$

对于任意被试 $x_i \in A_{k+2}$, 有

$$(k+2)\delta \geq d_B(x_i, x_0) > (k+1)\delta. \quad (15)$$

结合式(11)和(12)可得

$$d_B(x_i, x_0) - d_B(x, x_0) > \delta. \quad (16)$$

若被试 x, x_0, x_i 的位置构成三角形, 则有

$$d_B(x, x_i) > d_B(x_i, x_0) - d_B(x, x_0) > \delta; \quad (17)$$

若被试 x, x_0, x_i 的位置在一条直线上, 则有

$$d_B(x, x_i) \geq d_B(x_i, x_0) - d_B(x, x_0) > \delta. \quad (18)$$

由式(14)和(15)可知, 对于任意被试 $x \in A_k$, 任意被试 $x_i \in A_{k+2}$, 有 $d_B(x, x_i) > \delta$ 成立. 同样, 对于数组 $A_{k-2}, A_{k\pm l}$ 中的任意被试 $x_j (l = 3, 4, \dots)$, 也可得出 $d_B(x, x_j) > \delta$. 因此, 使用哈希映射后, 任意被试的邻域至多只存在于该被试所在的数组以及与该被试相邻的数组, 证明成立. \square

邻域关系具有对称性, 即若被试 x 的邻域中包含被试 x' , 则被试 x' 的邻域中也包含被试 x . 为了进一步提高生成被试邻域的效率, BNDRS 在哈希数组存储结构的基础上充分利用邻域关系的对称性. 在生成被试 x 的邻域的过程中, 一旦确定被试 x 与 x' 构成邻域关系, 则在将被试 x' 添加至被试 x 的邻域的同时, 也将被试 x 添加至被试 x' 的邻域. 按照这种方式生成邻域, 任意两个被试只需要进行一次欧氏距离的计算即可确认两者是否具有邻域关系.

综上, 将被试映射到哈希数组后, 结合邻域关系的对称性, 对于数组 A_k 中的任意被试 x , 生成它的邻域至多只需要与数组 A_k 中下标大于它的被试以及数组 A_{k+1} 中的被试计算欧式距离. 而根据式(1)生成

被试 x 的邻域则需要与被试集中剩余所有被试计算欧式距离, 因此, BNDRS 可以提高生成被试邻域的效率.

2.2 基于特征可分性度量的粗粒度搜索空间

特征可分性度量是一种根据特征在同类样本的相似程度和不同类样本的差异程度来衡量其分类能力的度量指标, 计算公式如下:

$$ASM(c) = \frac{|\mu_+(c) - \mu_-(c)|}{\sigma_+(c) + \sigma_-(c)}. \quad (19)$$

其中: c 表示某个脑功能连接特征, $\mu_+(c)$ 表示所有患病被试关于脑功能连接特征 c 的平均值, $\mu_-(c)$ 表示所有正常被试关于脑功能连接特征 c 的平均值, $\sigma_+(c)$ 表示所有患病被试关于脑功能连接特征 c 的标准差, $\sigma_-(c)$ 表示所有正常被试关于脑功能连接特征 c 的标准差, $|\mu_+(c) - \mu_-(c)|$ 反映了不同类别被试在脑功能连接特征 c 上的差异程度, $\sigma_+(c) + \sigma_-(c)$ 反映了相同类别被试在脑功能连接特征 c 上的相似程度.

由式(16)可知, 特征可分性度量值越大, 表明相应脑功能连接特征对不同类别被试的区分度越大, 对相同类别被试的区分度越小, 可见特征可分性度量可以反映脑功能连接特征的分类能力. 若将特征可分性度量值较大的若干特征同时加入到当前脑功能连接特征子集中, 则有望快速增加该特征子集的分类能力. 然而, 具有较大特征可分性度量值的若干脑功能连接特征有可能与当前脑功能连接特征子集中已有的特征存在冗余. 因此, 在对当前脑功能连接特征子集进行添加操作时, 需要计算各个特征组合的重要度, 然后再将重要度最大的特征组合中的特征添加至当前脑功能连接特征子集. 特征组合重要度的计算公式如下:

$$\text{SIG}(f) = \text{POS}_{|R \cup f|}(\pi_D) - \text{POS}_R(\pi_D). \quad (20)$$

其中: f 表示某个特征组合, R 表示当前脑功能连接特征子集, π_D 表示决策类的集合.

根据上面的描述, 基于特征可分性度量的脑功能连接特征分组策略实现的具体步骤为: 首先计算每个脑功能连接特征的特征可分性度量值; 然后根据特征可分性度量值对脑功能连接特征进行降序排序; 最后设定划分步长 $\lambda (2 \leq \lambda \leq |C|)$, 依次对脑功能连接特征进行分组, 从而得到脑功能连接特征组合集 $C' = \{f_1, f_2, \dots, f_{\lfloor |C|/\lambda \rfloor}\}$, 其中 $\lfloor |C|/\lambda \rfloor$ 表示对 $|C|/\lambda$ 向下取整. 于是, 脑功能连接特征的搜索空间从 $|C|$ 维的细粒度搜索空间降为 $\lfloor |C|/\lambda \rfloor$ 维的粗粒度搜索空间. 在搜索脑功能连接判别特征时, 不再是逐个地将特征添加至当前脑功能连接特征子集中, 而是将重要度最大的特征组合中包含的特征添加至当前脑功能连接特征子集, 从而避免大量冗余脑功能连接特征重要度的计算, 并迅速增大当前脑功能连接特征子集的正域, 提高搜索脑功能连接判别特征的效率.

2.3 算法描述及时间复杂度分析

基于前面的描述, 给出基于邻域决策粗糙集的脑功能连接生物标记物识别算法的伪代码如下.

算法1 基于邻域决策粗糙集的脑功能连接生物标记物识别算法.

step 1: 输入被试的脑功能连接特征集 C 、被试的标签、邻域半径 δ 、概率阈值 α 、划分步长 λ ;

step 2: 初始化脑功能连接判别特征集 $R = \emptyset$;

step 3: 基于哈希映射和邻域对称性生成被试关于 C 的邻域, 并计算 $|\text{POS}_C(\pi_D)|$;

step 4: 基于特征可分性度量对 C 进行分组;

step 5: 结合基于哈希映射和邻域对称性的快速邻域生成策略计算每一个特征组合的重要度, 并将重要度最大的特征组合中的脑功能连接特征添加至 R , 直至 $|\text{POS}_R(\pi_D)| \geq |\text{POS}_C(\pi_D)|$;

step 6: 删除 R 中重要度为 0 的脑功能连接特征;

step 7: 输出 R .

算法1的时间复杂度取决于 step 5, 设哈希数组的数量为 n , 每一个哈希数组中包含的被试数量为 $|U|/n$, 则 step 5 的时间复杂度为 $O(\lfloor |C|/\lambda \rfloor |U|^2/n)$. 因此, 算法1的时间复杂度为 $O(\lfloor |C|/\lambda \rfloor |U|^2/n)$. 而若使用基本邻域决策粗糙集特征约简方法实现脑功能连接生物标记物的识别, 时间复杂度为 $O(|C||U|^2)$. 由于 $O(\lfloor |C|/\lambda \rfloor |U|^2/n) < O(|C||U|^2)$, 本文所提出的 BNDRS 能够更加高效地识别脑功能连接生物标记物.

3 实验结果与分析

本节通过实验来验证所提出方法的有效性, 实验的运行环境为 Windows 10 操作系统, 处理器为 Inter(R) Core(TM) i5-6500 CPU 3.20G Hz, RAM 为 12 G. 算法利用 Matlab R2018b 运行工具编写的 Matlab 代码完成.

3.1 实验数据及评价标准

3.1.1 实验数据

本文实验所使用的数据集来自: 自闭症脑成像数据交换联盟发布的 ABIDE I 数据集和阿尔茨海默症神经影像学联盟发布的 ADNI 数据集. 其中 ABIDE I 数据集包含 569 名正常被试和 525 名自闭症患者, ADNI 数据集包括 66 名正常被试和 53 名阿尔茨海默症患者.

从以上数据集中获取的原始 fMRI 数据首先需要经过层间时间校正、头动校正等一系列图像预处理操作, 然后采用 AAL 模板进行脑区的划分, 并使用皮尔森相关系数计算不同脑区之间相互关联的程度, 得到脑功能连接矩阵, 最后将脑功能连接矩阵的上三角或下三角展开, 即可获得脑功能连接数据. 由于神经精神类疾病主要由大脑皮层脑区之间功能连接的障碍引起, 本文只选取位于大脑皮层的 90 个脑区进行脑区功能关联的计算, 最终获取的脑功能连接特征的维数为 4 005.

3.1.2 评价标准

本文主要从以下 3 个方面评价脑功能连接生物标记物识别方法的性能: 获得的脑功能连接判别特征的数量 Nu ; 时间消耗 t , 单位 s ; 获得的脑功能连接判别特征的分类能力. 其中脑功能连接判别特征的分类能力以 libsvm 工具箱中 SVM 分类器的结果为依据进行评价, 并采用 4 种常见的衡量分类能力的评价指标: 准确率 (Accuracy, Acc)、精确率 (Precision, Pr)、召回率 (Recall, Re) 和 F 度量 (F -measure, Fm). 它们的计算公式分别如下:

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}, \quad (21)$$

$$\text{Pr} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (22)$$

$$\text{Re} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (23)$$

$$\text{Fm} = \frac{2 \times \text{Pr} \times \text{Re}}{\text{Pr} + \text{re}}. \quad (24)$$

其中: TP 是将患者被试预测为患者的数目, TN 是将正常被试预测为正常人的数目, FP 是将正常被试预测为患者的数目, FN 是将患者被试预测为正常人的数目.

3.2 参数设置

本文测试了BNDRS在不同参数取值下所得脑功能连接判别特征的分类能力,在ABIDE I和ADNI数据集上,测试的概率阈值的范围为:0.5、0.6、0.7、0.8、0.9,测试的邻域半径的范围为:0.4、0.5、0.6、0.7、0.8.在ABIDE I数据集上测试的划分步长的范围为:11、15、19、23、27,在ADNI数据集上测试的划分步长的范围为:5、10、15、20、25.

测试结果显示在ABIDE I数据集和ADNI数据集上,对于不同的参数取值,BNDRS所得脑功能连接判别特征的分类能力整体上较平稳,并且,当邻域半径取0.4、概率阈值取0.6、划分步长取19时,BNDRS在ABIDE I数据集上获得了分类能力最强的脑功能连接判别特征.当邻域半径取0.8、概率阈值取0.9、划分步长取5时,BNDRS在ADNI数据集上获得了分类能力最强的脑功能连接判别特征.因此,为了获得分类能力最强的脑功能连接判别特征,本文在ABIDE I数据集上将邻域半径设置为0.4、概率阈值设置为0.6、划分步长设置为19.在ADNI数据集上将邻域半径设置为0.8、概率阈值设置为0.9、划分步长设置为5.

3.3 策略验证

本节实验验证BNDRS中基于哈希映射和邻域对称性的快速邻域生成策略以及基于特征可分性度量的脑功能连接特征分组策略的效果.实验设计了2个BNDRS的变异算法:BNDRS-F和BNDRS-H.其中,BNDRS-F在BNDRS基础上去除了基于哈希映射和邻域对称性的快速邻域生成策略,BNDRS-H在BNDRS基础上去掉了基于特征可分性度量的脑功能连接特征分组策略.实验使用的数据集是ABIDE I数据集,实验结果如表1所示.

表1 BNDRS与两个变异算法的比较

算法	Acc	Pr	Re	Fm	Nu	t
BNDRS	0.667	0.659	0.643	0.646	25	33.13
BNDRS-F	0.667	0.659	0.643	0.646	25	124.93
BNDRS-H	0.639	0.630	0.601	0.614	17	20 902.77

对比BNDRS和BNDRS-F可以看出,两者在Acc、Pr、Re、Fm、Nu五个指标上均相同,说明基于哈希映射和邻域对称性的快速邻域生成策略可以正确获取被试的邻域,不会影响算法的准确性.同时还可从表1看出,BNDRS的耗时小于BNDRS-F.实验结果表明,基于哈希映射和邻域对称性的快速邻域生成策略可以在不影响算法所得脑功能连接判别特征分类能力的前提下,提高算法的时间性能.

对比BNDRS和BNDRS-H可以看出,与BNDRS-

H相比,BNDRS在Acc、Pr、Re、Fm四个指标上均大于BNDRS-H.并且BNDRS的耗时远小于BNDRS-H的耗时.实验结果表明,基于特征可分性度量的脑功能连接特征分组策略不仅能大幅度提高邻域决策粗糙集的时间性能,还可以提高邻域决策粗糙集所获得的脑功能连接判别特征的分类能力.

3.4 特征约简能力验证

本节对BNDRS的特征约简能力进行验证,实验设计了1个变异算法:BNRS,该算法将BNDRS中的概率阈值设为1,不具备噪声处理能力.实验使用ABIDE I数据集,实验结果如表2所示.表2中C代表未约简的原脑功能连接特征,BNDRS_C代表使用BNDRS得到的脑功能连接判别特征,BNRS_C代表使用BNRS得到的脑功能连接判别特征.

表2 不同脑功能连接特征分类能力和数量的比较

特征	Acc	Pr	Re	Fm	Nu
C	0.639	0.630	0.606	0.614	4005
BNDRS_C	0.667	0.659	0.643	0.646	25
BNRS_C	0.639	0.625	0.616	0.618	17

由表2可以看出,与第1行未约简的原脑功能连接特征对比,BNDRS所得脑功能连接判别特征的数量大幅减少,并且BNDRS所得脑功能连接判别特征在指标Acc、Pr、Re、Fm上都较高.该实验结果说明BNDRS有能力删除大量的冗余脑功能连接特征,获得分类能力更强的脑功能连接判别特征.与第3行不具备噪声处理能力的BNRS相比,BNDRS所得脑功能连接判别特征在指标Acc、Pr、Re、Fm上也较高.该实验结果说明BNDRS在约简高维脑功能连接数据的过程中,可对噪声数据进行较好地处理.以上实验结果验证了BNDRS具有较强的特征约简能力.

3.5 鲁棒性验证

本节在ABIDE I数据集上验证BNDRS的鲁棒性.首先随机挑选ABIDE I数据集上5种不同比例(2%、4%、6%、8%、10%)的被试,然后为这些选出的被试添加高斯噪声,最后在5个带有不同比例噪声的ABIDE I数据集上测试BNDRS所得到的脑功能连接判别特征的分类能力.实验结果如图2所示.

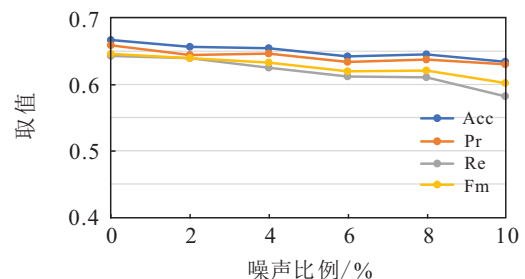


图2 不同噪声比例下的实验结果

由图2可以看出,当噪声比例为2%、4%、6%、8%时,与没有添加噪声的数据(噪声比例为0%)相比,指标Acc、Pr、Re、Fm的值波动较小,表明BNDRS在上述不同比例的噪声下仍能获得分类能力比较强的脑功能连接判别特征.当噪声比例达到10%时,与没有添加噪声的数据(噪声比例取0%)对比,指标Pr、Re、Fm的值下降相对较多,但指标Acc依然保持在一个不错的水平.上述实验结果表明,BNDRS在10%以内的噪声比例下有不错的抗噪声干扰能力,具有较强的鲁棒性.

3.6 脑功能连接生物标记物识别方法对比

本节将所提出的BNDRS与PCA^[6]、LLE^[7]、*F-score*^[8]、mRMR^[9]、NDRS^[12]进行比较来验证BNDRS的性能,实验所使用的数据集包括ABIDE I和ADNI两个数据集.

3.6.1 不同方法的分类能力对比

图3展示了不同方法所得脑功能连接判别特征的分类指标值.由图3可以看出,对于实验使用的两个数据集,BNDRS在Acc、Pr、Re、Fm四个指标上的值均大于对比方法,且在ADNI数据集上,BNDRS所得脑功能连接判别特征分类能力的优势尤为显著.该实验结果表明,与对比方法相比,BNDRS所得脑功能连接判别特征具有更强的分类能力.

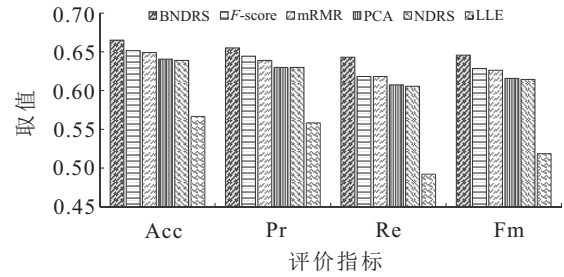
表3 不同方法时间性能及所得脑功能连接判别特征数量的比较

数据集	指标	BNDRS	PCA	LLE	mRMR	<i>F-score</i>	NDRS
ABIDE I	<i>t</i>	33.13	0.92	14.4	74.19	0.16	26 106.66
	Nu	25	152	44	25	25	17
ADNI	<i>t</i>	2.67	0.05	0.19	1.90	0.07	550.78
	Nu	20	58	26	20	20	17

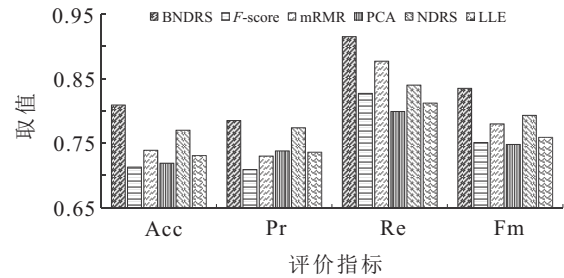
由表3还可以看出,BNDRS获得的脑功能连接判别特征的数量在ABIDE I和ADNI数据集上均仅多于NDRS,这是因为BNDRS具有强大的特征约简能力,可去除大量的冗余脑功能连接特征.

3.6.3 生物标记物对比

本文对比的所有方法中,PCA和LLE对脑功能连接数据的特征形式进行了转换,无法根据约简获得的特征定位到相关联的脑区,所以本节只对BNDRS、mRMR、*F-score*以及NDRS四种方法获得的脑功能连接判别特征进行关联脑区的定位.脑区定位的方法为:将4种方法得到的脑功能连接判别特征放回到原始的脑功能连接矩阵中,再对照AAL模板定位与自闭症和阿尔茨海默症相关的脑区.现有研究已证实与自闭症相关的脑区包括:丘脑(THA.L, THA.R)、



(a) ABIDE I数据集上不同方法评价指标的对比



(b) ADNI数据集上不同方法评价指标的对比

图3 各数据集上不同方法评价指标的对比

3.6.2 时间性能及脑功能连接判别特征数量对比

不同方法的约简耗时及所得脑功能连接判别特征数量的对比如表3所示.由表3可以看出,在实验使用的所有数据集上,BNDRS获得脑功能连接判别特征的时间消耗均远小于NDRS.此外,在ABIDE I数据集上,BNDRS获得脑功能连接判别特征的时间消耗小于mRMR.

楔前叶(PCUN.L, PCUN.R)、梭状回(FFG.L, FFG.R)、前扣带与旁扣带脑回(ACG.L, ACG.R)^[20-23]等脑区.与阿尔茨海默症相关的脑区包括:后扣带回(PCG.L, PCG.R)、楔前叶(PCUN.L, PCUN.R)、颞中回(MTG.L, MTG.R)、颞下回(ITG.L, ITG.R)、中央前回(PreCG.L, PreCG.R)^[24-26]等脑区.

图4展示了4种方法在ABIDE I数据集上得到的脑功能连接判别特征所对应的脑区连接情况.由图4可以看出,在ABIDE I数据集上,BNDRS得到的脑功能连接判别特征涉及上述与自闭症相关的所有脑区,而NDRS和mRMR得到的脑功能连接判别特征不涉及与自闭症相关的梭状回(FFG.L, FFG.R),*F-score*得到的脑功能连接判别特征不涉及与自闭症相关的丘脑(THA.L, THA.R).

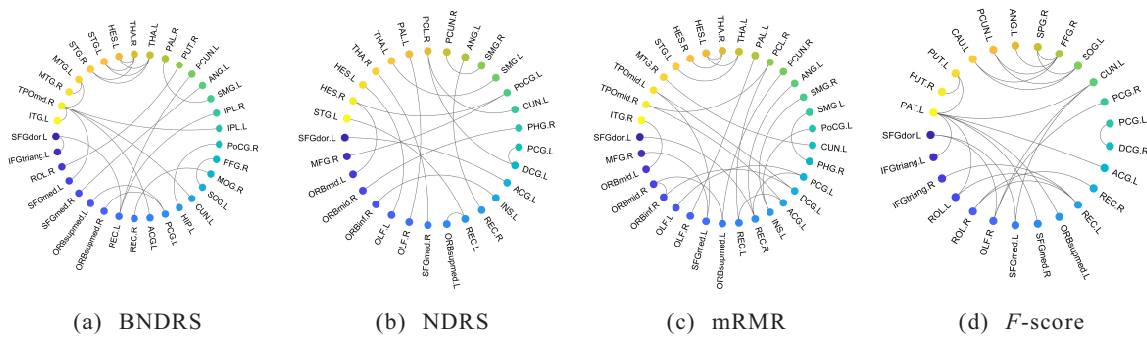


图4 ABIDE I数据集上不同方法对应生物标记物的对比

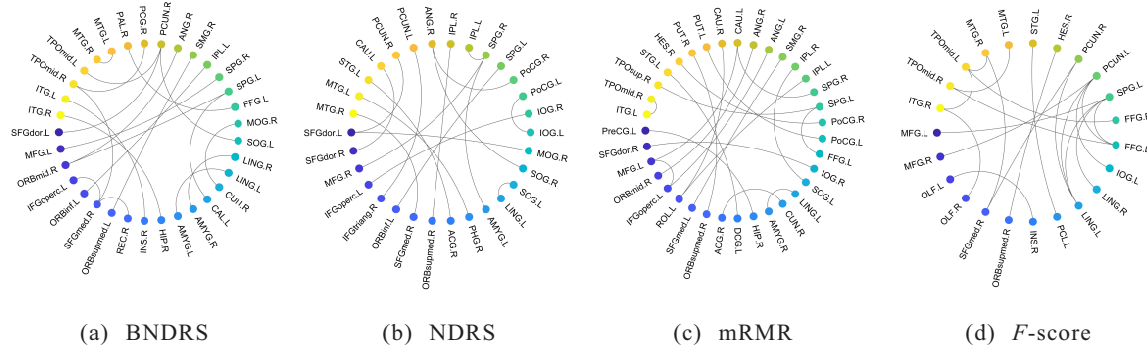


图5 ADNI数据集上不同方法对应生物标记物的对比

图5展示了4种方法在ADNI数据集上得到的脑功能连接判别特征所对应的脑区连接情况. 可以看出, 在ADNI数据集上, BNDRS涉及与阿尔茨海默症相关的后扣带回(PCG.R)、楔前叶(PCUN.R)、颞中回(MTG.L, MTG.R)、颞下回(ITG.L, ITG.R); NDRS涉及与阿尔茨海默症相关的颞中回(MTG.L, MTG.R)、楔前叶(PCUN.L, PCUN.R); mRMR涉及与阿尔茨海默症相关的颞下回(ITG.L)、中央前回(PreCG.L); F-score涉及与阿尔茨海默症相关的楔前叶(PCUN.L, PCUN.R)、颞中回(MTG.L)、颞下回(ITG.L).

从以上实验结果可看出, 在ABIDE I和ADNI数据集上, BNDRS分别定位到了最多数量的与自闭症和阿尔茨海默症相关的脑区. 因此, BNDRS有望为神经精神类疾病的诊断提供更准确的脑功能连接生物标记物.

4 结论

本文提出了一种新的面向神经精神类疾病的脑功能连接生物标记物识别方法. 该方法针对脑功能连接数据的连续性和高噪声的特点, 利用能有效处理连续和高噪声数据的邻域决策粗糙集来识别与神经精神类疾病高度相关的脑功能连接判别特征; 针对脑功能连接数据的高维特点, 该方法通过快速生成邻域和缩小特征搜索空间来提高邻域决策粗糙集识别脑功能连接判别特征的效率. 实验结果表明, 该方法

可以高效地获得分类能力更强的脑功能连接判别特征, 更加准确定位与脑疾病相关的脑区, 进而为神经精神类疾病的诊断提供更准确的生物标记物.

下一步的工作将研究融合不同邻域半径下的粒度信息, 探索面向脑功能连接数据的邻域决策粗糙集多粒度特征约简方法, 进一步提高获得的脑功能连接生物标记物的准确性.

参考文献(References)

- [1] Lynall M E, Bassett D S, Kerwin R, et al. Functional connectivity and brain networks in schizophrenia[J]. The Journal of Neuroscience, 2010, 30(28): 9477-9487.
- [2] Zhang J R, Wang J H, Wu Q Z, et al. Disrupted brain connectivity networks in drug-naive, first-episode major depressive disorder[J]. Biological Psychiatry, 2011, 70(4): 334-342.
- [3] Rosa M J, Portugal L, Hahn T, et al. Sparse network-based models for patient classification using fMRI[J]. NeuroImage, 2015, 105: 493-506.
- [4] Challis E, Hurley P, Serra L, et al. Gaussian process classification of Alzheimer's disease and mild cognitive impairment from resting-state fMRI[J]. NeuroImage, 2015, 112: 232-243.
- [5] Rashid B, Arbabshirani M R, Damaraju E, et al. Classification of schizophrenia and bipolar patients using static and dynamic resting-state fMRI brain connectivity[J]. NeuroImage, 2016, 134: 645-657.
- [6] Mohanty R, Sinha A M, Remsik A B, et al. Machine learning classification to identify the

- stage of brain-computer interface therapy for stroke rehabilitation using functional connectivity[J]. *Frontiers in Neuroscience*, 2018, 12: 353-366.
- [7] Shen H, Wang L B, Liu Y D, et al. Discriminative analysis of resting-state functional connectivity patterns of schizophrenia using low dimensional embedding of fMRI[J]. *NeuroImage*, 2010, 49(4): 3110-3121.
- [8] Kong Y Z, Gao J L, Xu Y P, et al. Classification of autism spectrum disorder by combining brain connectivity and deep neural network classifier[J]. *Neurocomputing*, 2019, 324: 63-68.
- [9] Hamdi S M, Aydin B, Filali Boubrahimi S, et al. Biomarker detection from fMRI-based complete functional connectivity networks[C]. *The 1st International Conference on Artificial Intelligence and Knowledge Engineering*. Laguna Hills, 2018: 17-24.
- [10] Hu Q H, Yu D R, Liu J F, et al. Neighborhood rough set based heterogeneous feature subset selection[J]. *Information Sciences*, 2008, 178(18): 3577-3594.
- [11] Yao Y Y, Zhao Y. Attribute reduction in decision-theoretic rough set models[J]. *Information Sciences*, 2008, 178(17): 3356-3373.
- [12] Li W W, Huang Z Q, Jia X Y, et al. Neighborhood based decision-theoretic rough set models[J]. *International Journal of Approximate Reasoning*, 2016, 69: 1-17.
- [13] Pan R L, Wang X M, Yi C S, et al. Multi-objective optimization method for thresholds learning and neighborhood computing in a neighborhood based decision-theoretic rough set model[J]. *Neurocomputing*, 2017, 266: 619-630.
- [14] 苑红星, 卓雪雪, 竺德, 等. 基于矩阵的混合型邻域决策粗糙集增量式更新算法[J]. *控制与决策*, 2022, 37(6): 1621-1631.
(Yuan H X, Zhuo X X, Zhu D, et al. Incremental updating algorithms of neighborhood decision-theoretic rough set model for hybrid data based on matrix[J]. *Control and Decision*, 2022, 37(6): 1621-1631.)
- [15] 孙颖, 蔡天使, 张毅, 等. 基于合理粒度的局部邻域决策粗糙计算方法[J]. *南京大学学报: 自然科学*, 2021, 57(2): 262-271.
(Sun Y, Cai T S, Zhang Y, et al. Justifiable granularity based local neighborhood decision-theoretic rough set approach[J]. *Journal of Nanjing University: Natural Science*, 2021, 57(2): 262-271.)
- [16] Yao Y Y, Zhao Y, Wang J. On reduct construction algorithms[C]. *Proceedings of the 1st International Conference on Rough Sets and Knowledge Technology*. Chong qing, 2006: 239-304.
- [17] Liu Y, Huang W L, Jiang Y L, et al. Quick attribute reduct algorithm for neighborhood rough set model[J]. *Information Sciences*, 2014, 271: 65-81.
- [18] 胡清华, 于达仁, 谢宗霞. 基于邻域粒化和粗糙逼近的数值属性约简[J]. *软件学报*, 2008, 19(3): 640-649.
(Hu Q H, Yu D R, Xie Z X. Numerical attribute reduction based on neighborhood granulation and rough approximation[J]. *Journal of Software*, 2008, 19(3): 640-649.)
- [19] Ma Y Y, Luo X Y, Li X L, et al. Selection of rich model steganalysis features based on decision rough set -positive region reduction[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2019, 29(2): 336-350.
- [20] Li H L, Parikh N A, He L L. A novel transfer learning approach to enhance deep neural network classification of brain functional connectomes[J]. *Frontiers in Neuroscience*, 2018, 12: 491.
- [21] Heinsfeld A S, Franco A R, Craddock R C, et al. Identification of autism spectrum disorder using deep learning and the ABIDE dataset[J]. *NeuroImage: Clinical*, 2018, 17: 16-23.
- [22] Lynch C J, Uddin L Q, Supekar K, et al. Default mode network in childhood autism: Posteromedial cortex heterogeneity and relationship with social deficits[J]. *Biological Psychiatry*, 2013, 74(3): 212-219.
- [23] Uddin L Q, Supekar K, Lynch C J, et al. Saliency network-based classification and prediction of symptom severity in children with autism[J]. *JAMA Psychiatry*, 2013, 70(8): 869-879.
- [24] He Y, Wang L, Zang Y F, et al. Regional coherence changes in the early stages of Alzheimer's disease: A combined structural and resting-state functional MRI study[J]. *NeuroImage*, 2007, 35(2): 488-500.
- [25] Buckner R L, Sepulcre J, Talukdar T, et al. Cortical hubs revealed by intrinsic functional connectivity: Mapping, assessment of stability, and relation to Alzheimer's disease[J]. *The Journal of Neuroscience*, 2009, 29(6): 1860-1873.
- [26] Daianu M, Jahanshad N, Nir T M, et al. Rich club analysis in the Alzheimer's disease connectome reveals a relatively undisturbed structural core network[J]. *Human Brain Mapping*, 2015, 36(8): 3087-3103.

作者简介

冀俊忠(1969—),男,教授,博士生导师,从事机器学习、计算智能等研究, E-mail: jjz01@bjut.edu.cn;

龙腾(1997—),男,硕士生,从事粗糙集理论、计算智能的研究, E-mail: longteng_bjut@163.com;

杨翠翠(1985—),女,副教授,博士,从事机器学习、计算智能等研究, E-mail: yangcc@bjut.edu.cn.

(责任编辑: 孙艺红)