

轻度认知障碍磁共振信号中固有频率动态功能性连接的聚类研究

李 栋¹, 吴海锋^{1,2}, 保 涵¹, 马 佳¹, 曾 玉^{1,2}

(1. 云南民族大学电气信息工程学院, 昆明 650500; 2. 云南省高校智能传感网络及信息系统创新团队, 昆明 650500)

摘 要: 功能性连接(Functional connectivity, FC)可以表示脑区的协同工作能力,目前广泛采用动态功能性连接(Dynamic functional connectivity, DFC)和聚类分析相结合的方法研究疾病的显著性差异分析和分类。但现有研究,对于聚类个数的确定和聚类结果选用并没有明确的标准,且传统的DFC无法表示不同频率的FC信息。因此,本文对轻度认知障碍(Mild cognitive impairment, MCI)磁共振信号中固有频率DFC聚类问题进行研究。首先对被试的时间进程(Time course, TC)数据做噪音辅助的多元经验模态分解并计算DFC;然后通过评判辅助的聚类方法做聚类分析,再采用最小二乘对聚类结果做拟合;最后采用分类器做分类。实验采用阿尔茨海默病神经影像学(Alzheimer's disease neuroimaging, ADNI)数据库的数据对本文算法进行测试。实验结果表明,有监督聚类分类准确率高于无监督聚类;引入固有频率的DFC分类准确率要高于传统的DFC;最小二乘拟合能提升分类准确率。

关键词: 磁共振成像;经验模态分解;最小二乘;动态功能性连接;K-均值

中图分类号: TB183 **文献标志码:** A

Clustering Related Factors of Intrinsic Frequency Dynamic Functional Connection in MRI Signal of Mild cognitive Impairment

LI Dong¹, WU Haifeng^{1,2}, BAO Han¹, MA Jia¹, ZENG Yu^{1,2}

(1. School of Electrical and Information Technology, Yunnan Minzu University, Kunming 650500, China; 2. Intelligent Sensor Network & Information System Innovative Research Team in Science and Technology in University of Yunnan Province, Kunming 650500, China)

Abstract: Functional connectivity (FC) can represent the ability of brain regions to work together. At present, a combination of dynamic functional connectivity (DFC) and cluster analysis is widely used to study the significant difference analysis and classification of diseases. However, in the existing study, there is no clear standard for the determination of the number of clusters and the selection of clustering results, and the traditional DFC cannot represent the FC information of different frequencies. Therefore, this paper studies the clustering related factors of intrinsic frequency DFC in MRI signal of mild cognitive impairment (MCI). First, the noise-assisted multivariate empirical mode decomposition of the time course (TC) data is performed and the DFC is calculated. Then, the cluster is analyzed through the evaluation-assisted

clustering method, and the least square method is used to fit the clustering results. Finally, classifier is used for classification. The contribution of this paper is to suggest a more reasonable clustering method and a more number of clusters to obtain functional connections at different intrinsic frequencies. In the experiment, we used the Alzheimer's disease neuroimaging (ANDI) database to test the proposed method. The experimental results show that the accuracy of supervised clustering used in this paper is higher than that of unsupervised clustering; the classification accuracy of DFC with natural frequency is higher than that of traditional DFC; the least square fitting can improve classification accuracy.

Key words: magnetic resonance imaging; empirical model decomposition; least square; dynamic functional connectivity; K-means

引 言

静息态功能核磁共振成像(Rest-state functional magnetic resonance imaging, rs-fMRI)技术具有无辐射性和无创伤性,且采集时不需受试者做认知任务,因此广泛应用于脑科学研究中^[1]。通常,大脑的工作是由不同脑区协作完成,即使在静息态时也是如此。由于通过rs-fMRI所得到的功能性连接(Functional connectivity, FC)可以表示脑区间相互作用强度,因此FC可以适当地表示大脑的协同工作模式。已有研究^[2-4]表明,一些精神类疾病患者的脑区协同工作模式会发生改变,因此FC对研究这些疾病诊断具有一定的现实意义。例如,轻度认知障碍(Mild cognitive impairment, MCI)作为阿尔兹海默症(Alzheimer's disease, AD)前期的必经阶段,已被证实其患者大脑的FC和正常人具有显著性差异^[3-4]。

在rs-fMRI信号中,有两种较常见计算FC的方法。一种是通过自动解剖标签(Anatomical automatic labeling, AAL)的兴趣区(Region of interest, ROI)^[5]提取血氧水平依赖(Blood oxygenation level dependent, BOLD)信号,然后求其皮尔逊相关^[6]。另外一种常见方法是对fMRI信号进行盲分离得到时间进程(Time course, TC)^[7-8],然后由静息网络(Resting state network, RSN)模板筛选出功能性脑区,对脑区对应的TC求皮尔逊相关。由于大脑FC会展现一些非平稳性^[9-10],但基于皮尔逊相关的FC无法反映时变特性,因此可在TC信号上加上若干时间窗口,在每个窗口内求Person相关,得到动态功能性连接(Dynamic FC, DFC)^[11]。遗憾的是,人脑神经元的放电频率并不一定相同^[12],DFC方法虽能表现信号的时变性,但却无法展现不同频率上的FC信息。小波变换是在时频域上对信号进行变换,近年来也被用于rs-fMRI的FC研究,其基本做法是对TC信号做小波变换后再分析其相关特性,以此得到相应的FC^[13]。虽然小波的FC方法可以得到频率上的信息,但众所周知,其相关值往往与选取的母函数相关。经验模态分解(Empirical model decomposition, EMD)是一种无需依赖母函数非平稳信号处理方法,通过不断迭代筛选将信号分解为有限个不同固有模态频率(Intrinsic mode functions, IMF)信号和一个残差信号^[14],最近也被常用于fMRI信号计算FC^[15-16]。但是直接对信号做EMD分解会出现端点效应和模态混叠现象。为解决这个问题,Huang等提出了集成经验模态分解(Empirical EMD, EEMD)^[17],该方法是将原始信号加入幅度较小的白噪声,再进行分解,做多次试验求平均值。EEMD分解在一定程度上缓解了端点效应和模态混叠的问题,但该方法以增加算法复杂度为代价,需要做多次试验才能消除白噪声带来的影响。而且EMD和EEMD都是对受试者的每个TC信号进行独立分解,会导致各TC的IMF分量数和频率各不相同,就难以计算各个层上的相关系数。多元经验模态分解(Multivariate EMD, MEMD)^[18-19]是一种多通道的经验模态分解,可实现多个TC时间序列的同步联合分析,确保多个TC信号的IMF在时间和频率上均保持匹配,从而利于在各层上做相关分析。但MEMD仍然会存在模态混叠的问题,因此,2011年Rehman等提出了噪音辅助的多元经验模态分解(Noise assisted MEMD, NA-MEMD)算法,将额外的白噪声通道拼接接到输入信号中,通过MEMD分解后再舍弃白噪

声通道,保留原始信号的IMF分量。NA-MEMD结合了EEMD和MEMD的优点,可以多通道处理非平稳信号,并有效减少了模态混叠现象,而且相比于EEMD算法无需做多次分解求平均,具有更低的计算复杂度,因此广泛应用于非平稳信号的处理中。

目前还存在一类利用无监督聚类来研究FC动态性的方法,它们可将DFC表达为一系列时间状态序列特征^[20-22]或聚类中心的线性组合的特征^[23-25],通过分析这些特征表达来寻找疾病的显著性差异,也可以采用隐马尔科夫模型(Hidden Markov model, HMM)^[26]和支持向量机(Support vector machine, SVM)^[27]等分类器对其分类。然而,无论是显著性差异的分析还是分类器分类,其性能与聚类结果紧密相关,不同的聚类结果将得到不同性能。在无监督聚类中,设定不同的聚类数会得到不同的聚类结果,现有的方法设定聚类数大多采用经验值,并根据类内距离和类间距离评判聚类数性能^[28-30]。例如,在路易体痴呆的研究中Ma等将聚类数设为2^[30],在自闭症谱系障碍的研究中Guo等将聚类数设为5^[31],在精神分裂症的研究中Miller等将聚类数设为5^[32],在睡眠质量的研究中Zhou等将聚类数设为4^[33],认知任务下的AD研究中Lee等将聚类数设为4^[34]。合理的聚类数应设为多少,不同的聚类数会得到什么样的性能,以上文献并没有进行详细描述。另外,无监督聚类即使设置聚类数,但其初始聚类中心随机产生,聚类结果仍具有随机性,每一次聚类结果都不尽相同。从多个聚类结果中采用何种标准选用何种也会影响FC的分析,比如根据状态占有率(State occupancy rate, SOCR)^[35-36]选取结果,根据正常对照(Normal control, NC)组、疾病组占有率^[37]等参数评价结果。

本文将MCI与NC的rs-fMRI作为处理对象,关注动态FC中有关聚类的相关问题,不仅研究加窗的DFC的聚类性能,而且研究了固有频率DFC(DFC of intrinsic frequencies, IF-DFC)的聚类性能,该IF-DFC通过噪音辅助的多元经验模态分解和加窗皮尔逊算法得到。此外,利用提出的评判辅助聚类方法(Evaluation assisted clustering, EA Clustering)对DFC和IF-DFC进行聚类,以评价各种聚类方式的性能。进一步,本文还对比了由各种聚类所得特征来进行分类的性能,包括时间状态序列以及所提出的最小二乘(Least square, LS)拟合特征。另外本文引入K-means++算法^[38],相比于传统K-means算法,K-means++通过优化初始聚类中心的选取,具有收敛更快、聚类效果更好的优点。

1 材料和方法

1.1 数据采集和预处理

本实验采用(Alzheimer's disease neuroimaging initiative, ADNI)数据库提供的rs-fMRI公开数据,其下载地址为<http://adni.loni.usc.edu/>,数据参数见表1。首先,利用数据处理和脑成像分析(Data processing & analysis of brain imaging, DPABI)工具箱^[5]对所下载的数据进行预处理,其下载地址为:<http://rfmri.org/dpabi>,具体过程如下:

(1) 对原始数据去除前10帧图像,保留130帧图像;

(2) 以第48个切片为基准对所有的切片进行校准,使每个切片上的数据具有相同时间点;

(3) 对所有被试进行头动校正,将头动校正到同一位置,并为接下来的图像质量控制提供数据;

(4) 将与rs-fMRI数据匹配的MRI数据进行颅骨剥离,然后同功能图像进行配准,将MRI数

表1 fMRI数据相关参数

Table 1 Parameters related to fMRI data

参数	描述
数据库	ADNI-2
磁场强度	3.0Tesla
采集设备	Philip
翻转角/(°)	80
TR/ms	3 000
TE/ms	30
像素大小/(mm×mm)	3.3×3.3
切片数量	48
时间点	140

据变换到 rs-fMRI 数据的空间,使得 rs-fMRI 数据可以被分割成脑脊髓液、灰质和白质信号;

(5) 去除一些混淆因素,将 6 个头动参数(包含 3 个转动参数和 3 个平移参数)、脑脊髓液和白质信号被作为无关变量去除;

(6) 评估被试者头部的旋转和平移,发现所有参与者都没有表现出过度的头部运动;

(7) 进行归一化和平滑,并将所有数据都过滤到一个频率范围(0.01~0.08 Hz)。

再次,采用 GIFT(Group ICA of fMRI Toolbox)^[7]软件从预处理数据中提取功能区,其下载地址为 <http://trendscenter.org/trends/software/gift/index.html>。该软件将 fMRI 信号视为空间信息经过时间进程的混合,采用最小描述长度(Minimum description length, MDL)准则估计出独立脑区成分个数,通过独立成分分析(Independent component analysis, ICA)算法分离出若干个独立脑区。实验中,Gift 软件分离算法选择 ICASSO 方法,重复做 10 次实验选取每个脑区的最佳估计,并与每个静息态网络(Resting state network, RSN)模板做空间多元回归分析,筛选出回归系数最大的 53 个脑区,其中 RSN 模板下载地址为 https://trendscenter.org/trends/data/neuromark/Neuromark_good.nii。该模板将 RSN 分为 7 大类,分别为下皮层(Sub-cortical, SC)、听觉(Auditory, AUD)、感觉运动(Sensorimotor, SM)、视觉(Visual, VIS)、认知控制(Cognitive-control, CC),默认模式网络(Default-mode network, DM)和小脑(Cerebellar, CB)。

1.2 IF-DFC 相关矩阵

本文采用 NA-MEMD 来计算 IF-DFC,其可表征不同频率和时间的 FC 强度,计算如图 1 所示。若每一个被试的 TC 数为 N ,添加的噪音序列数为 M ,执行 $N + M$ 通道的 MEMD 后,每个通道将得到 Q 个 IMF,那么前 N 个通道的 IMF 即为分解结果。对分解得到的 IMF 划分为若干个时间窗口,在每个窗口内计算皮尔逊相关,即可得到第 i 个序列和第 j 个序列在 q 层窗口 w 的 IF-DFC 相关矩阵,有

$$R^{(q,w)} = [r_{ij}^{(q,w)}] \in \mathbf{R}^{N \times N} \quad (1)$$

式中: N 取决于 RSN 模板,本文采用的模板 N 为 53; M 为噪音序列数,本文分别尝试使用不同的噪音序列数做 NA-MEMD 分解,得到 $M = 5$ 时能取得较好的效果。

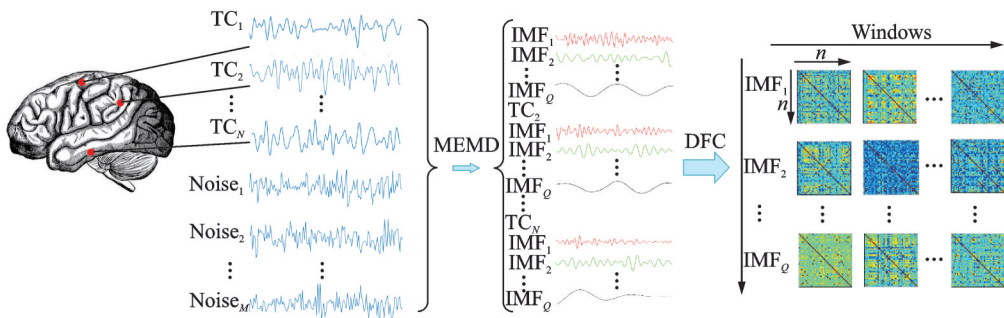


图 1 IF-DFC 算法示意图

Fig.1 Schematic diagram of IF-DFC algorithm

1.3 EA Clustering 方法

由于 IF-DFC 相关矩阵为对称阵,因此取其上三角部分来聚类,得到 IF-DFC 矢量,表示为

$$d^{(q,w)} = \text{tru}(R^{(q,w)}) \quad (2)$$

式中: $d^{(q,w)} \in \mathbf{R}^{V \times 1}$; $V = N \times (N - 1) / 2$ 。

采用 EA Clustering 对该相关矢量进行聚类,分别采用有监督聚类和无监督聚类两种方式。对于有

监督聚类方法,若聚2类,第1类的聚类中心为训练集所有MCI受试者在某一层IF-DFC矢量的组平均,第2类聚类中心为训练集所有NC受试者在某一层IF-DFC矢量的组平均;若聚3类,则第3类聚类中心为训练集所有受试者在某一层IF-DFC矢量的组平均。再将训练集和测试集所有被试IF-DFC矢量分别与聚类中心求欧氏距离,比较距离大小得到每个受试者在某一层的转态矢量。无监督方式对某一层所有被试的IF-DFC矢量采用K均值聚类,分别将聚类数设为2、3、4或5。为避免大部分矢量只聚到少量几类上,无监督方式在运行时还需设置最小SOCR,定义为

$$o = \frac{n^*}{N_s} \times 100\% \quad (3)$$

式中: $n^* = \min\{n_1, n_2, \dots, n_K\}$, $k=1, 2, \dots, +K$ 为第 k 个类上所聚的矢量数, $N_s = n_1 + n_2 + \dots + n_K$ 。无监督聚类的最小SOCR的设置可参见表2。另外,通过聚类结果还可得到一被试者IF-DFC在第 q 层的状态序列矢量,有

$$\mathbf{S}^{(q)} = [s_1^{(q)}, s_2^{(q)}, \dots, s_W^{(q)}] \quad (4)$$

式中 $s_w^{(q)}$, $w=1, 2, \dots, W$ 表示为IF-DFC在第 w 个窗口上所聚类的对应状态,用一系列数字表示,即 $s_w^{(q)} \in \{1, 2, \dots, K\}$ 。

EA Clustering将尝试不同的监督方式、聚类数和SOCR来得到不同的聚类结果,通过引入评判聚类性能指标,聚类时间 t 和MCI与NC占比差 Δ 等推断聚类的合理性,其中 Δ 定义为

$$\Delta = \frac{1}{K} \sum_{k=1}^K \left| \frac{\delta_k}{n_k} \right| \times 100\% \quad (5)$$

式中 δ_k 为聚在第 k 个类上MCI组和NC组矢量之差,越少的聚类时间和越高的占比差视为较合理的聚类,该算法步骤如表3所示。

表3 EA Clustering算法步骤

Table 3 Steps of EA Clustering algorithm

输入: IF-DFC 相关矩阵 $R^{(q,w)}$
输出: 第 q 层的聚类中心, 占比差 Δ 及状态序列矢量 $\mathbf{S}^{(q)}$
步骤:
(1) 由式(4)得到 IF-DFC 矢量 $\mathbf{d}^{(q,w)}$ 。
(2) 设定聚类数。
(3) 若聚类方式设为有监督则转到(a);若设为无监督则转到(d)。
(a) 若聚类数设为2转到(b);若聚类数设为3转到(c);
(b) 分别计算训练组中MCI和NC的组平均作为聚类中心;
(c) 分别计算训练组中MCI, NC和所有被试的组平均作为聚类中心;
(d) 根据设定的聚类数由表2设定最小SOCR;
(e) K 均值聚类;
(f) 若聚类结果满足条件(d)则结束,否则重复(e);
(g) 得到聚类中心,并由式(6)和(7)分别得到状态序列矢量 $\mathbf{S}^{(q)}$ 和占比差 Δ 。

1.4 LS 特征提取

为进一步评判 EA Clustering 的聚类性能并进一步提取特征矢量,将每一被试在第 q 层上的 IF-DFC 矢量表达为聚类中心的线性组合,通过 LS 拟合求解该线性组合,所得拟合系数作为进一步分类的特征,较高分类准确率对应的聚类方式被推断为有较好合理性,特征提取具体如下。

令 $\mathbf{a}_k^{(q)} \in \mathbf{R}^{V \times 1}$ 为 EA Clustering 在第 q 层上第 k 个聚类中心矢量,则由 K 个聚类中心矢量组成的矩阵表示为

$$\mathbf{A}^{(q)} = [\mathbf{a}_1^{(q)} \mathbf{a}_2^{(q)}, \dots, \mathbf{a}_K^{(q)}] \quad (6)$$

若一被试在第 q 层上第 w 个窗口的 IF-DFC 矢量表示为聚类中心矢量的线性组合,即

$$\mathbf{d}^{(q,w)} = \mathbf{A}^{(q)} \boldsymbol{\theta}^{(q,w)} + \boldsymbol{\xi}^{(q,w)} \quad (7)$$

式中 $\boldsymbol{\xi}^{(q,w)}$ 为误差矢量,那么系数 $\boldsymbol{\theta}^{(q,w)}$ 可通过 LS 拟合为

$$\hat{\boldsymbol{\theta}}^{(q,w)} = \mathbf{A}^{(q)\dagger} \mathbf{d}^{(q,w)} \quad (8)$$

式中 $(\cdot)^\dagger$ 表示伪逆。最后将该被试 W 个窗口的拟合系数矢量合并,形成最终的特征矢量

$$\boldsymbol{\Theta}^{(q)} = [\boldsymbol{\theta}^{(q,1)\text{H}}, \boldsymbol{\theta}^{(q,2)\text{H}}, \dots, \boldsymbol{\theta}^{(q,W)\text{H}}]^\text{H} \quad (9)$$

上述计算步骤如表 4 所示。

表 4 LS 特征提取步骤

Table 4 LS feature extraction steps

输入:第 q 层 IF-DFC 矢量 $\mathbf{d}^{(q,w)}$, $w=1, 2, \dots, W$

输出:第 q 层特征矢量 $\boldsymbol{\Theta}^{(q)}$

步骤:

- (1) 在第 q 层上,由表 3 的聚类得到式(5)中聚类中心矩阵 $\mathbf{A}^{(q)}$;
- (2) 由式(7)拟合得到第 q 层第 w 个窗口的系数矢量 $\hat{\boldsymbol{\theta}}^{(q,w)}$;
- (3) 将 W 个窗口的拟合系数合并得到式(8)的特征矢量 $\boldsymbol{\Theta}^{(q)}$ 。

1.5 分 类

HMM 是常用于时间序列的分类模型,由表 3 可以得到一系列的时间状态序列,因此本文采用 HMM 对该状态序列做分类,通过期望最大化算法分别对 MCI 和 NC 训练集受试者的时间状态序列训练得到 2 个 HMM 模型,再分别求每个模型对测试集时间状态序列的对数似然,比较其大小做分类。从另外一个角度看,MCI 与 NC 分类实际上是一种二分类过程,而 SVM 通过寻找特征空间最优决策超平面进行分类,是目前应用最为广泛的二分类模型,因此使用 SVM 分类器对表 3 得到的转态矢量和表 4 得到的 LS 特征做分类。本文所采用的 HMM 软件下载地址为 <https://www.cs.ubc.ca/~murphyk/Software/HMM/hmm.html>,SVM 采用 MATLAB 自带程序,分类器的输入选用了 DFC 和 IF-DFC 聚类结果,即由式(6)得到的状态序列矢量和由式(9)得到的 LS 特征,以下是各分类方法的具体说明:

(1) UnHMM-DFC:由表 3 对 DFC 作无监督聚类得到转态矢量,其作为 HMM 的输入进行分类;其中,DFC 转态矢量特征提取来自文献[39],HMM 分类方法来自文献[40-41],注意到表 3 算法的处理对象是 IF-DFC,需将其替换为 DFC,下面的分类算法中若出现 DFC,均是如此处理;

(2) SupHMM-DFC:由表 3 对 DFC 作有监督聚类得到转态矢量,使用 HMM 分类。

(3) UnHMM-IF-DFC:由表 3 对 IF-DFC 作无监督聚类得到转态矢量,使用 HMM 分类。

(4) SupHMM-IF-DFC:由表3对IF-DFC作有监督聚类,使用HMM分类。

(5) UnSVM-DFC:由表3对DFC作无监督聚类得到转态矢量,使用SVM分类;其中,SVM分类模型与文献[42]相同。

(6) SupSVM-DFC:由表3对DFC作有监督聚类,使用SVM分类。

(7) UnSVM-IF-DFC:由表3对IF-DFC作无监督聚类,使用SVM分类。

(8) SupSVM-IF-DFC:由表3对IF-DFC作有监督聚类,使用SVM分类。

(9) SupLS-SVM-DFC:由表3对DFC作有监督聚类,由表4得到LS特征,采用SVM分类;同表3类似,表4中的IF-DFC需替换为DFC,后面的算法也是如此。

(10) SupLS-SVM-IF-DFC:由表3对IF-DFC作有监督聚类,由表4得到LS特征,采用SVM分类。

(11) UnLS-SVM-DFC:由表3对DFC作无监督聚类,由表4得到LS特征,采用SVM分类,其中,采用无监督聚类,再表达为LS拟合特征方法来自于文献[23],后面的算法也是如此。

(12) UnLS-SVM-IF-DFC:由表3对IF-DFC作无监督聚类,由表4得到LS特征,采用SVM分类。

(13) UnLS-SVM-IF-DFC++:将表3的K均值方法替换为K-means++算法^[38],再由表3对IF-DFC作无监督聚类,由表4得到LS特征,采用SVM分类。

测试以上算法时,均采用5折交叉验证,把总数据样本随机分为5组,1组作为测试集,其他4组作为训练集,5组中每1组都会作为1次测试集。为了避免偶然性,实验中共做了10次5折交叉验证,因此最终分类准确率为50次实验结果的平均值。另外,在上述的分类算法中,无监督聚类方法对测试集和训练集的所有数据进行聚类,有监督聚类算法仅对训练集数据进行聚类。

2 实验结果

2.1 EA Clustering 聚类结果

本节分别给出了EA Clustering对DFC和IF-DFC的聚类结果,主要是占比差 Δ 和聚类时间,共独立做了50次实验,实验结果为50次实验的均值。另外,本节还给出了聚类中心所对应的FC矩阵,由于每一次聚类结果,聚类中心变化较大,因此该FC矩阵为50次实验中随机选取的一次结果,仅为了显示较直观的结果。

表5给出了DFC的聚类时间和占比差。从表5可以看出,有监督EA Clustering聚类时间要比无监督聚类时间低得多,且占比差也比无监督大。对于无监督聚类,其占比差从高到低排列为聚类数5、4、3和2,聚类时间从高到低排列为聚类数5、4、3和2。该结果表明,无监督设置的聚类数多,聚类中心间的差异越大。

表5 DFC EA Clustering 聚类结果

Table 5 DFC EA Clustering results

聚类方式	有监督聚类	有监督聚类	无监督聚类	无监督聚类	无监督聚类	无监督聚类
聚类个数	2	3	2	3	4	5
占有率差/%	100	66.67	13.26	18.45	23.15	30.79
聚类时间/s	0.02	0.04	1.74	18.38	27.63	31.17

表6给出了IF-DFC的聚类结果。其中,由于该聚类需要在某一层的IMF上进行,因此表6给出了IMF₁、IMF₃和IMF₄层上的聚类结果,选取原则是本文算法在这几层IMF分类效果较好。从表5可以看到有监督聚类时间仍然远少于无监督,且占比差也高于无监督。对于无监督聚类,无论在何层IMF,其

表6 IF-DFC EA Clustering 聚类结果
Table 6 IF-DFC EA Clustering results

聚类方式	有监督聚类	有监督聚类	无监督聚类	无监督聚类	无监督聚类	无监督聚类
聚类个数	2	3	2	3	4	5
占有率差/%	100	66.67	11.31/1 6.78/3 14.58/4	12.68/1 17.08/3 17.63/4	27.28/1 21.72/3 22.40/4	27.46/1 31.63/3 28.88/4
聚类时间/s	0.02	0.04	4.09/1 3.38/3 2.70/4	46.87/1 16.30/3 19.27/4	125.48/1 27.50/3 30.51/4	185.65/1 19.75/3 23.42/4

注:/1、/3和/4分别表示 IMF₁、IMF₃、IMF₄。

占比差从高到低大约为5类、4类、3类和2类,聚类时间4和5类最长,3类次之,2类最少。该排列顺序和表5的结果类似,但表6中某些IMF层上的占比差却比表5略有提升,例如表6中IMF₁层上4类的占比差高于表5。对于聚类时间,表6的某些IMF层上的聚类时间也略少于表5,例如IMF₄的5类。不过,总体而言,表5,6的占比差和聚类时间还是比较接近。

图2和图3分别给出了DFC和IF-DFC在IMF₁层上各聚类中心的FC矩阵图。与表5,6一致,有监督聚类中,MCI占比高和NC占比高的聚类中心间显示的差异性更大。另外,还可以看到,NC占比高的聚类

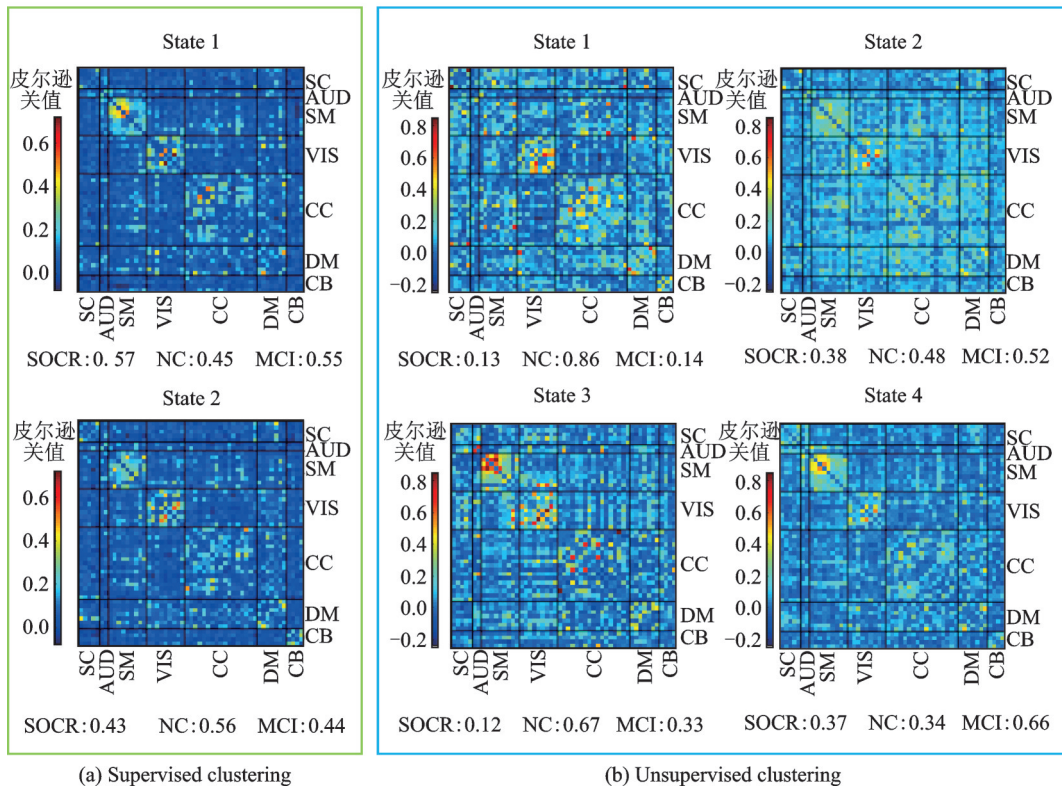
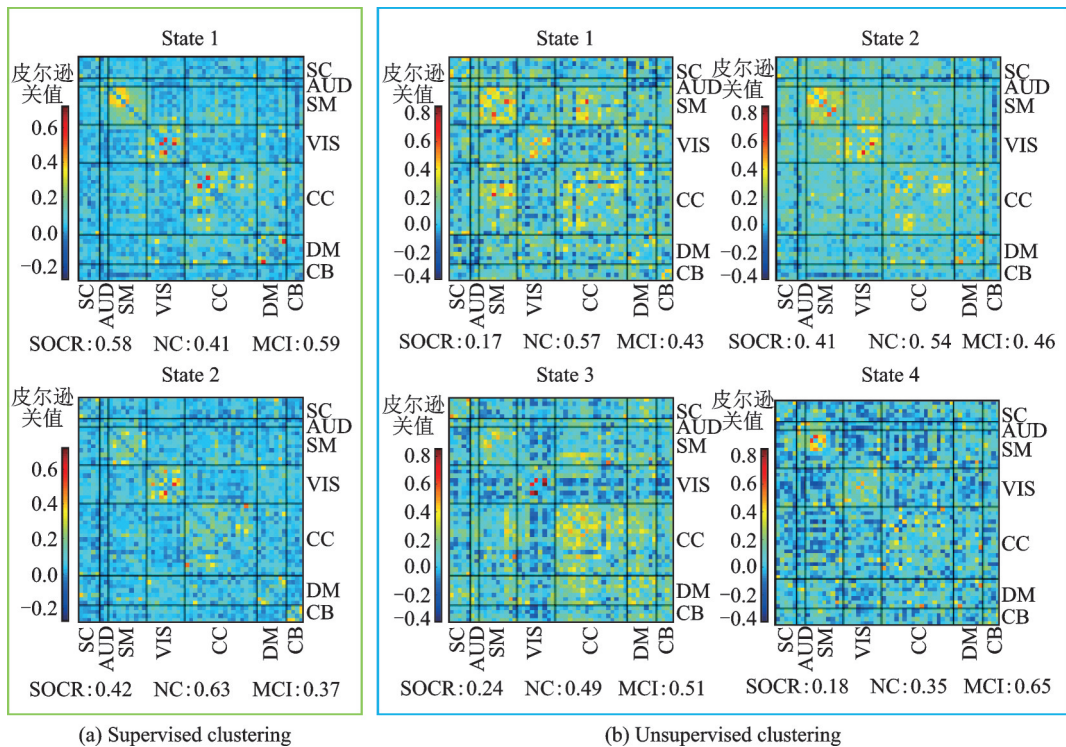


图2 DFC在IMF₁层上EA Clustering 聚类结果图

Fig.2 Clustering result of DFC EA Clustering on IMF₁ layer

图3 IF-DFC在 IMF_1 层上EA Clustering聚类结果图Fig.3 Clustering result of IF-DFC EA Clustering on IMF_1 layer

中心的连接性要强于MCI占比高的聚类中心,这个结果与MCI被认为是功能性丧失的理论是一致的。

2.2 LS特征分类结果

本节将给出引入了LS特征后的分类结果,该LS特征分别由DFC和IF-DFC利用EA Clustering聚类所得,并尝试用SVM分类器做分类。同时,还将以上结果与仅用EA Clustering聚类后的状态矢量的分类结果进行了对比。图4先给出了IF-DFC用EA Clustering聚类的结果,该IF-DFC选用NA-MEMD IMF_1 的分类结果,原因该层的分类准确率较高。从平均准确率看,大部分引入LS特征的分类算法的准确率要高于仅用转态矢量的分类准确率。特别地,对于LS特征,采用有监督聚类的分类准确率要高于无监督的,无论是将聚类数设为2或3均是如此,其准确率已到达63%,而不再是接近50%。而从标准差看,大部分引入LS特征的分类标准差要小于仅用转态矢量的标准差,这表明引入LS特征再分类要更具稳定性。此外,对于无监督的LS算法,还将K-means++算法与传统的K-means聚类算法作比较,可以看出将聚类数设为3,4和5时,无论是分类准确率还是标准差,K-means++算法都要优于K-means算法,而将聚类数设为2时,K-means++算法分类效果略比K-means算法差。

图5将分类器的输入由IF-DFC换成了DFC。可以看出,有监督聚类引入LS特征的分类算法的准确率高其他算法,而分类标准差低于其他算法。除此之外,采用IF-DFC的有监督的LS分类准确率已在63%左右,已超过采用DFC分类算法3个百分点。

图4给出的IF-DFC的分类结果是选取了具有较高准确率的 IMF_1 层,而图6给出了IF-DFC在多个 IMF 层的分类结果。同样地,将这些 IMF 层上的IF-DFC通过聚类后计算LS特征,再用SVM进行分类。从图中可以看到, IMF_1 、 IMF_2 和 IMF_4 层准确率较高, IMF_1 、 IMF_2 和 IMF_3 层分类标准差较低,因此

这也是图4中有监督LS-SVM选取IMF₁的原因。

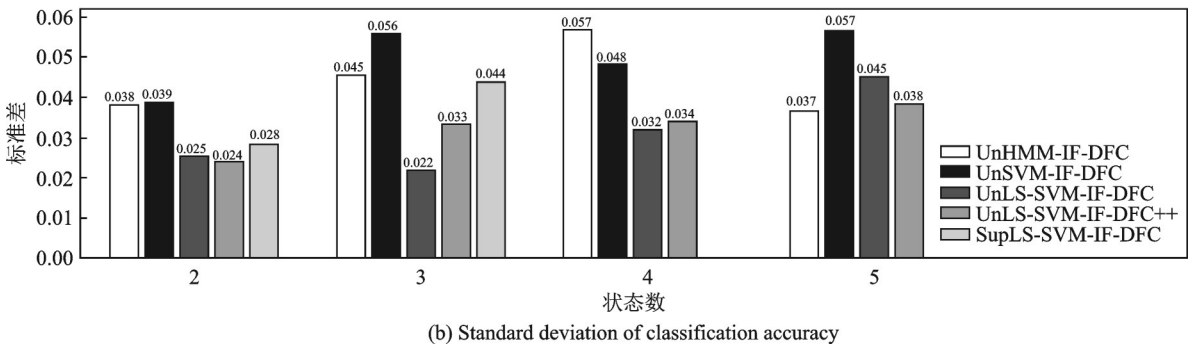
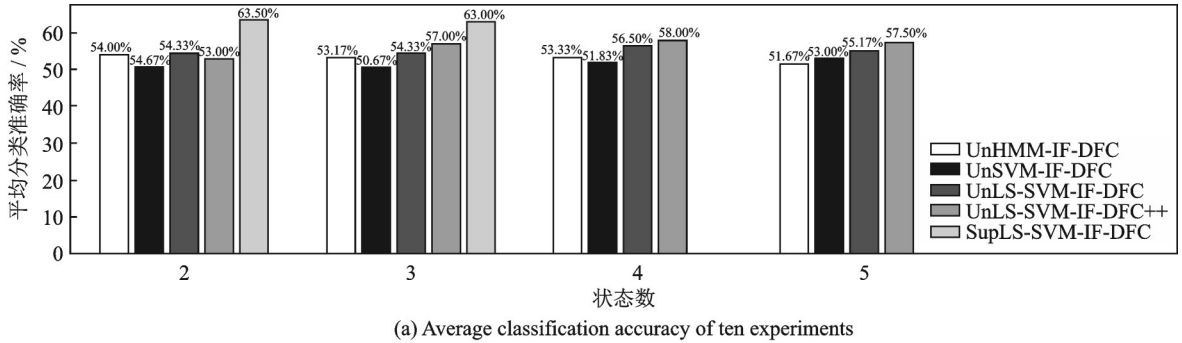


图4 IF-DFC各算法分类结果图

Fig.4 IF-DFC classification results of each algorithm

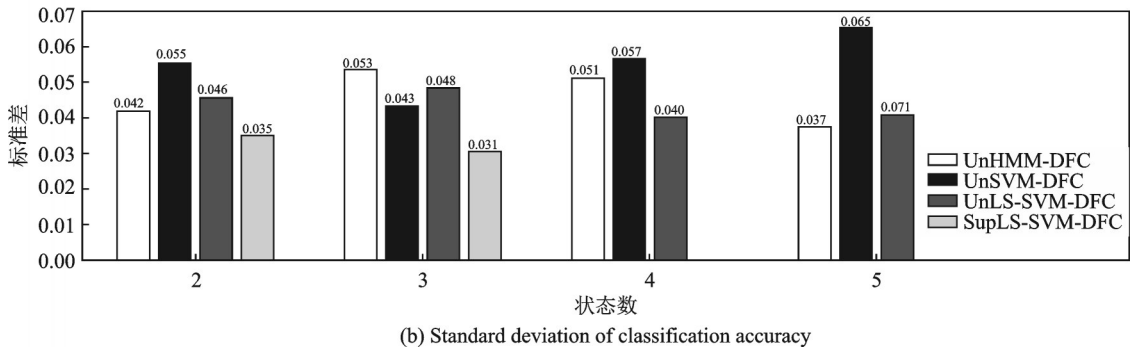
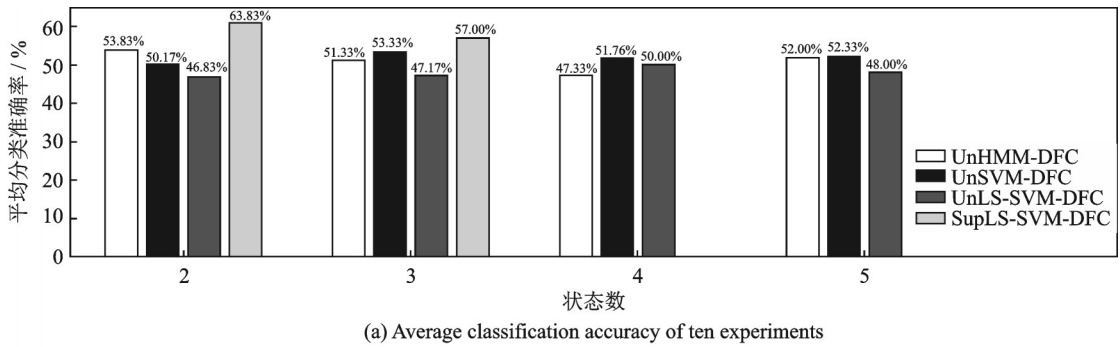
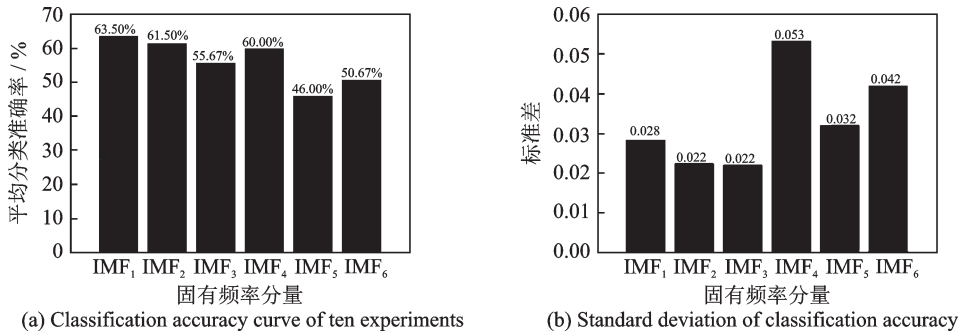


图5 DFC各算法分类结果图

Fig.5 Classification results of each algorithm of DFC



(a) Classification accuracy curve of ten experiments

(b) Standard deviation of classification accuracy

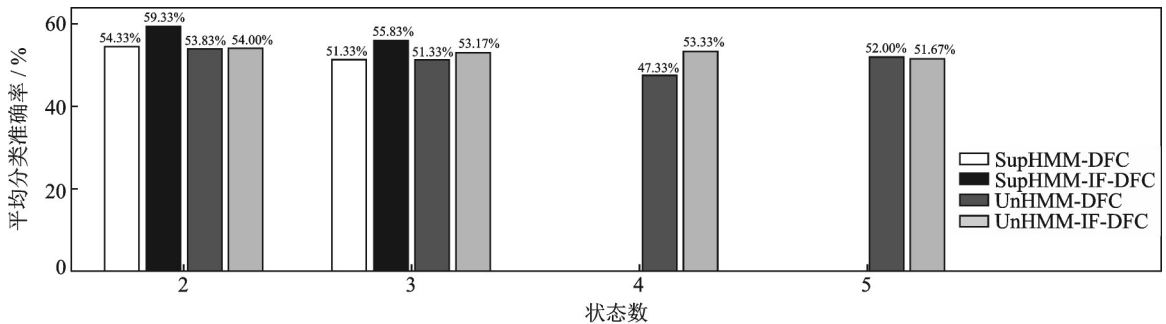
图6 LS-SVM有监督聚类二类IMF₁~IMF₆的分类结果

Fig.6 LS-SVM supervised clustering classification results of two classes IMF₁—IMF₆.

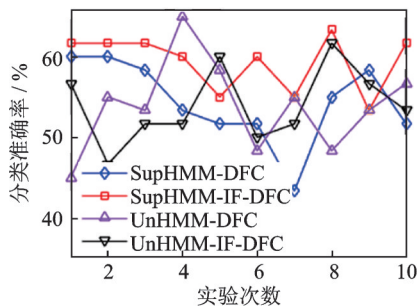
2.3 HMM和SVM分类器结果

本节将分别给出HMM和SVM采用式(9)中聚类状态序列矢量的分类准确率,该序列矢量分别由DFC和IF-DFC采用EA Clustering获得。图7给出DFC和IF-DFC在设置不同聚类数时,采用EA Clustering聚类后获得的分类准确率,其中IF-DFC选用的是IMF层中较高准确率的分类结果。其中SupHMM-IF-DFC是在IMF₄的计算结果,UnHMM-IF-DFC是在IMF₁的计算结果,从图7(a)给出的平均分类准确率看,平均分类准确率在47%~59%,难以将MCI组和NC组区分,但也应看到IF-DFC的分类准确率略高于传统DFC,除聚类数为5的情况除外,而有监督聚类分类准确率都略高于无监督聚类。图7(b~e)给出了10次实验的分类准确率曲线。可以看出:将聚类数设为2和3时,分别有1和2次实验有监督聚类IF-DFC分类准确率低于DFC;将聚类数设为3和4时,也仅有3和2次实验无监督聚类IF-DFC分类准确率低于DFC。

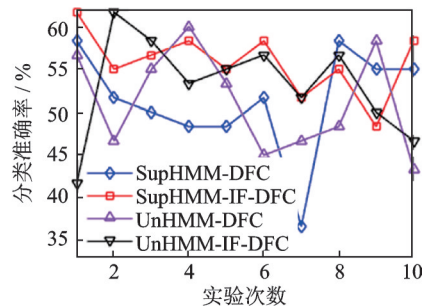
图8给出了SVM对DFC和IF-DFC EA Clustering无监督聚类后的分类结果。与图7的结果类似,



(a) Average classification accuracy of ten experiments



(b) Number of clusters being 2



(c) Number of clusters being 3

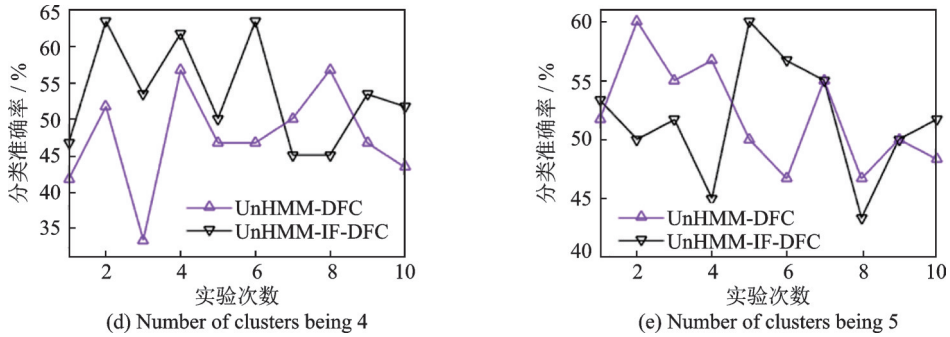
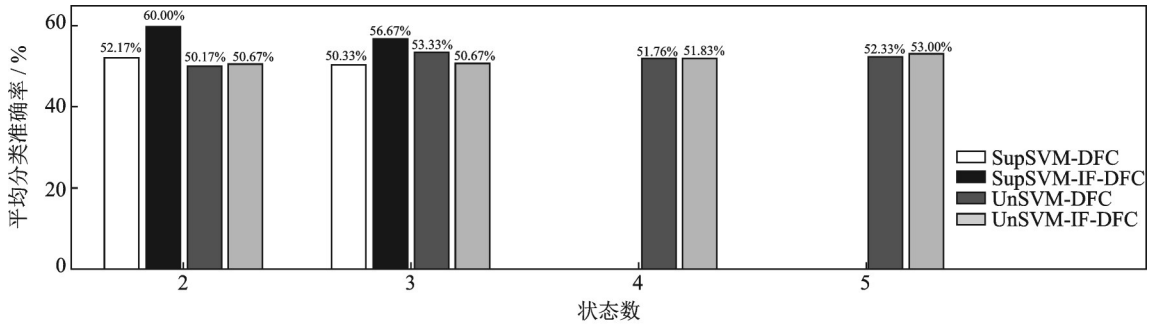


图 7 HMM算法分类结果图

Fig.7 Classification results of the HMM algorithm



(a) Average classification accuracy of ten experiments

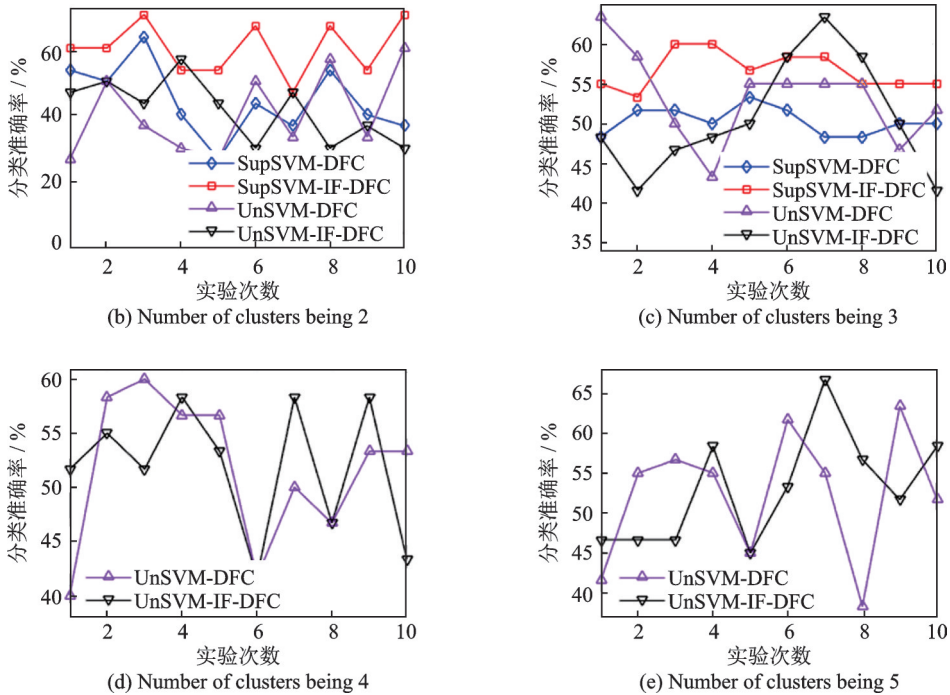


图 8 SVM算法分类结果图

Fig.8 Classification results of the SVM algorithm

IF-DFC的分类准确率略高于DFC,除无监督聚类数为3的情况除外,有监督聚类分类准确率略高于无监督,除DFC聚类数为3的情况除外。

3 讨 论

通过FC研究MCI与NC差异是较常见技术,因为MCI作为一种功能性连接丧失症,其在FC的表现与NC组将有所区别。由于FC的非平稳性,对其分段加窗后聚类也并不是种新技术,但现有文献似乎并未过多关心该技术中的聚类方式,比如无监督方式还是有监督组平均,甚至聚类数也各不相同,从2类到5类都有。本文试图从聚类本身所展示特性来评估聚类的合理性,同时引入机器学习的分类方法,从分类准确率进一步评估聚类方法的合理性,主要关心的仍是聚类方式及聚类数。

EA Clustering引入聚类参数来体现聚类的合理性,例如SOCR,聚类数,聚类时间,聚类方式以及占比差等指标。大部分无监督聚类中心指向并不明确,虽占比差超过0,但由于所聚的类代表的簇既有MCI被试也有NC被试,因此其表现的特征归为何组就不易断定。同时,无监督聚类还需设定最小SOCR条件,否则所聚的矢量只会集中到少数几类中,极端时只集中到一类,这将给分析带来极大困难。但是,设置SOCR条件又会使聚类时间增加,因为聚类结果不满足SOCR条件,算法只能重新开始。这说明若无监督聚类不受干预,聚类其实更倾向于聚较少类,而不是更多的类,由此可以猜测无监督的自组织性存在一定欠缺。而通过组平均的方式得到的状态指向更加明确,且其运算只需要加法和除法而已,而无需像无监督聚类求欧氏距离并反复迭代,因此计算组平均的时间要远小于无监督聚类。

从另外一方面,利用组平均或无监督聚类结果,可以将动态连接性看成一系列有限状态序列,该有限状态序列其实就是所聚的类。因此,可以将该有限状态序列作为特征输入到分类器中,但由于该方法的分类准确率均接近50%,即接近二分类中1/2的概率,因此考虑了另外一种特征提取的方法,把动态功能性连接拟合为各有限状态的线性组合,将该拟合系数作为分类特征。采用这种方式的分类,同样是采用组平均方式所得到的特征的分类精度要高于无监督聚类,且分类准确率可以达到63%。另外,本文在计算动态FC时,考虑另外一种方式,即将TC先经EMD分解后再计算动态FC,这样做的目的在于可以计算每一层固有频率上的动态FC。从实验结果看,在IMF层上聚类后,总有一些层上的分类准确率要高于未分层的动态FC。一个可能的解释是,分层以后会把一些MCI组和NC组的共同噪声进行分离,如原始fMRI信号中来自于呼吸和心跳的共同噪声^[43-45],从而提高分类准确率。另外一个可能的解释是,神经元放电频率并不一定相同,在不同的固有频率上进行聚类,以此寻找MCI组与NC组间的差异性应该具有更好的合理性。当然,本文在固有频率上的研究并不充分。首先,IMF信号聚类在何层具有更好的性能,或者具有更好的分类准确率,需要进一步分析。从实验结果看,利用LS特征进行分类时,最好的分类准确率在第1层,第2层和第4层,均具有超过60%的分类准确率。然而,IMF信号仅能体现信号在一固有频率的特征,该固有频率仅仅是一个频带,因此只能得到在中低频率上具有较好性能的结论,而无法具体到频率点。下一步工作可采用希尔伯特变换计算IMF信号的瞬时频率,从而确定聚类具体在何频率带进行,这对完善IMF聚类具有积极意义。比如,本文现在采用的多通道EMD仅能保证一个被试的各功能区在相同IMF层上的固有频率匹配,但被试与被试间在相同的层上匹配并不严格,因为该分解是独立对每一个被试分别进行。然而,计算出具体频带后,可根据该结果选择被试与被试在何层具有最佳的频率匹配。

最后需要强调的是,本文并未将重心放在如何提高MCI与NC的分类准确率上。从以往的文献看,大多对DFC进行聚类的方法^[30-34]主要也是分析疾病组与控制组间的显著性差异,而较少地关心用其来进行分类,有可能的一个原因在于用这种方法进行分类的准确率并不高。从本文的实验结果也可以看到,其部分结果仅略超60%。因此,本文也仅仅将分类准确率作为一种评价聚类性能的指标,而不是将

聚类看成是一种分类方法。从该角度看,经IMF的有监督聚类后,分析MCI组和NC组间的显著性差异的工作也是必要的,特别与加窗DFC的方法相比,这种差异性有何不同,这部分研究也将是未来工作的重点之一。特别地,如果以提高分类准确率为目的来研究聚类方法,在未来也可以考虑利用提取特征的方法,因为目前本文的分类准确率未得到大幅提升的原因之一是FC的维度太高,至少是一个相关阵的上三角部分,一些研究已经表明,特征数多将产生较多的冗余信息。

4 结束语

MCI与NC在FC上表现了较大的差异性,且FC还具有一定的动态特性。对FC进行聚类后再研究其差异性是一种常用方法,遗憾的是现存方法并未过多讨论聚类方法本身。针对以上问题,本文提出利用EA Clustering、IF-DFC和LS拟合特征等方法研究和评价聚类的合理性,得出如下结论:(1)从聚类时间和占比差等聚类指标看,有监督聚类比无监督聚类具有更好的合理性,因为其聚类时间更少,占比差更高,且有监督聚类数可设为2或3;(2)从分类准确率的实验结果看,实验结果仍支持有监督聚类所展示的差异性比无监督聚类更显著,因为无论采用HMM和SVM的分类器,还是采用聚类所得到的状态序列和LS拟合特征,有监督聚类的分类准确率均高于无监督聚类。(3)仍从分类准确率看,将动态FC在不同固有频率层上聚类,其显著性差异要好于不划分固有频率的聚类,因为无论采用HMM和SVM的分类器,还是采用聚类所得到的状态序列和LS拟合特征,引入固有频率的聚类分类准确率要高于无固有频率层上的聚类。

参考文献:

- [1] LEE M H, SMYSER C D, SHIMONY J S. Resting-state fMRI: A review of methods and clinical applications[J]. *American Journal of Neuroradiology*, 2013, 34(10): 1866-1872.
- [2] LI C, FRONCZEK-PONCELET J, LANGE D, et al. Impact of acute sleep deprivation on dynamic functional connectivity states[J]. *Human Brain Mapping*, 2020, 41(4): 994-1005.
- [3] CÓRDOVA-PALOMERA A, KAUFMANN T, PERSSON K, et al. Disrupted global metastability and static and dynamic brain connectivity across individuals in the Alzheimer's disease continuum[J]. *Scientific Reports*, 2017, 7(1): 1-14.
- [4] SCHUMACHER J, PERAZA L R, FIRBANK M, et al. Dynamic functional connectivity changes in dementia with Lewy bodies and Alzheimer's disease[J]. *NeuroImage: Clinical*, 2019, 22: 101812.
- [5] YAN C G, WANG X D, ZUO X N, et al. DPABI: data processing & analysis for (resting-state) brain imaging[J]. *Neuroinformatics*, 2016, 14(3): 339-351.
- [6] HONEY C J, SPORNS O, CAMMOUN L, et al. Predicting human resting-state functional connectivity from structural connectivity[J]. *Proceedings of the National Academy of Sciences*, 2009, 106(6): 2035-2040.
- [7] ERHARDT E B, RACHAKONDA S, BEDRICK E J, et al. Comparison of multi-subject ICA methods for analysis of fMRI data[J]. *Human Brain Mapping*, 2011, 32(12): 2075-2095.
- [8] WU H, LI D, LU M, et al. fMRI activations via low-complexity second-order inverse-sparse-transform blind separation[J]. *Digital Signal Processing*, 2021, 117: 103137.
- [9] ALLEN J J B, COHEN M X. Deconstructing the "resting" state: Exploring the temporal dynamics of frontal alpha asymmetry as an endophenotype for depression[J]. *Frontiers in Human Neuroscience*, 2010, 4: 232.
- [10] HANDWERKER D A, ROOPCHANSINGH V, GONZALEZ-CASTILLO J, et al. Periodic changes in fMRI connectivity [J]. *Neuroimage*, 2012, 63(3): 1712-1719.
- [11] HUTCHISON R M, WOMELSDORF T, ALLEN E A, et al. Dynamic functional connectivity: Promise, issues, and interpretations[J]. *Neuroimage*, 2013, 80: 360-378.
- [12] CALVIN W H, SYPERT G W. Fast and slow pyramidal tract neurons: An intracellular analysis of their contrasting repetitive firing properties in the cat[J]. *Journal of Neurophysiology*, 1976, 39(2): 420-434.

- [13] SAVVA A D, MATSOPOULOS G K, MITSIS G D. A Wavelet-based approach for estimating time-varying connectivity in resting-state fMRI[J]. *Brain Connectivity*, 2022, 12(3): 285-298.
- [14] HUANG N E, SHEN Z, LONG S R, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis[J]. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 1998, 454(1971): 903-995.
- [15] WU H, KONG L, ZENG Y, et al. Resting-state brain connectivity via multivariate EMD in Mild cognitive impairment[J]. *IEEE Transactions on Cognitive and Developmental Systems*, 2021. DOI: 10.1109/TCDS.2021.3054504.
- [16] DAS S, SAO A K, BISWAL B. Precise estimation of resting state functional connectivity using empirical mode decomposition [C]//*Proceedings of International Conference on Brain Informatics*. Switzerland: Springer, Cham, 2020: 75-84.
- [17] ZHANG Z, LIAO M, YAO Z, et al. Frequency-specific functional connectivity density as an effective biomarker for adolescent generalized anxiety disorder[J]. *Frontiers in Human Neuroscience*, 2017, 11: 549.
- [18] REHMAN N, MANDIC D P. Multivariate empirical mode decomposition[J]. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2010, 466(2117): 1291-1302.
- [19] UR REHMAN N, MANDIC D P. Filter bank property of multivariate empirical mode decomposition[J]. *IEEE Transactions on Signal Processing*, 2011, 59(5): 2421-2426.
- [20] YUAN Y, ZHANG L, LI L, et al. Distinct dynamic functional connectivity patterns of pain and touch thresholds: A resting-state fMRI study[J]. *Behavioural Brain Research*, 2019, 375: 112142.
- [21] DU M, ZHANG L, LI L, et al. Abnormal transitions of dynamic functional connectivity states in bipolar disorder: A whole-brain resting-state fMRI study[J]. *Journal of Affective Disorders*, 2021, 289: 7-15.
- [22] FU Z, TU Y, DI X, et al. Characterizing dynamic amplitude of low-frequency fluctuation and its relationship with dynamic functional connectivity: An application to schizophrenia[J]. *Neuroimage*, 2018, 180: 619-631.
- [23] MILLER R L, YAESOUBI M, TURNER J A, et al. Higher dimensional meta-state analysis reveals reduced resting fMRI connectivity dynamism in schizophrenia patients[J]. *PloS One*, 2016, 11(3): e0149849.
- [24] MILLER R L, YAESOUBI M, CALHOUN V D, et al. Higher dimensional fMRI connectivity dynamics show reduced dynamism in schizophrenia patients[C]//*Proceedings of 2014 International Workshop on Pattern Recognition in Neuroimaging*. Tuebingen, Germany: IEEE, 2014: 1-4.
- [25] ZHANG Y, XIAO L, ZHANG G, et al. Multi-paradigm fMRI fusion via sparse tensor decomposition in brain functional connectivity study[J]. *IEEE Journal of Biomedical and Health Informatics*, 2020, 25(5): 1712-1723.
- [26] GRAVES A, FERNÁNDEZ S, GOMEZ F, et al. Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks[C]//*Proceedings of the 23rd International Conference on Machine Learning*. New York, United States: ACM, 2006: 369-376.
- [27] YAN B, XU X, LIU M, et al. Quantitative identification of major depression based on resting-state dynamic functional connectivity: A machine learning approach[J]. *Frontiers in Neuroscience*, 2020, 14: 191.
- [28] YU L, ZHOU C. Determining the best clustering number of k-means based on bootstrap sampling[C]//*Proceedings of 2018 2nd International Conference on Data Science and Business Analytics (ICDSBA)*. Changsha, China: IEEE, 2018: 78-83.
- [29] RAY S, TURI R H. Determination of number of clusters in k-means clustering and application in colour image segmentation [C]//*Proceedings of the 4th International Conference on Advances in Pattern Recognition and Digital Techniques*. Guangzhou, China: Springer-Verlag, 1999: 137-143.
- [30] MA W Y, YAO Q, HU G, et al. Dysfunctional dynamics of intra-and inter-network connectivity in dementia with Lewy bodies [J]. *Frontiers in Neurology*, 2019, 10: 1265.
- [31] GUO X, DUAN X, SUCKLING J, et al. Partially impaired functional connectivity states between right anterior insula and default mode network in autism spectrum disorder[J]. *Human Brain Mapping*, 2019, 40(4): 1264-1275.
- [32] MILLER R L, YAESOUBI M, CALHOUN V D. Cross-frequency rs-fMRI network connectivity patterns manifest differently for schizophrenia patients and healthy controls[J]. *IEEE Signal Processing Letters*, 2016, 23(8): 1076-1080.
- [33] ZHOU Z, CAI B, ZHANG G, et al. Prediction and classification of sleep quality based on phase synchronization related whole-brain dynamic connectivity using resting state fMRI[J]. *NeuroImage*, 2020, 221: 117190.

- [34] LEE Y B, YOO K, ROH J H, et al. Brain-state extraction algorithm based on the state transition (BEST): A dynamic functional brain network analysis in fMRI study[J]. Brain Topography, 2019, 32(5): 897-913.
- [35] SENDI M S E, ZENDEHROUH E, FU Z, et al. Alzheimer's disease projection from normal to mild dementia reflected in functional network connectivity: A longitudinal study[J]. Frontiers in Neural Circuits, 2020, 14: 87.
- [36] SENDI M S E, PEARLSON G D, MATHALON D H, et al. Multiple overlapping dynamic patterns of the visual sensory network in schizophrenia[J]. Schizophrenia Research, 2021, 228: 103-111.
- [37] YAESOUBI M, MILLER R L, CALHOUN V D. Mutually temporally independent connectivity patterns: A new framework to study resting state brain dynamics with application to explain group difference based on gender[J]. Neuroimage, 2015, 107: 85.
- [38] ARTHUR D, VASSILVITSKII S. K-means++: The advantages of careful seeding[R].[S.l.]: Stanford, 2006.
- [39] LIAO W, WU G R, XU Q, et al. DynamicBC: A Matlab toolbox for dynamic brain connectome analysis[J]. Brain Connectivity, 2014, 4(10): 780-790.
- [40] KHOSLA M, JAMISON K, NGO G H, et al. Machine learning in resting-state fMRI analysis[J]. Magnetic Resonance Imaging, 2019, 64: 101-121.
- [41] SUK H I, WEE C Y, LEE S W, et al. State-space model with deep learning for functional dynamics estimation in resting-state fMRI[J]. NeuroImage, 2016, 129: 292-307.
- [42] RÄTSCHE G, SONNENBURG S, SCHÄFER C. Learning interpretable SVMs for biological sequence classification[J]. BMC Bioinformatics, 2006, 7(1): 1-14.
- [43] BIRN R M, DIAMOND J B, SMITH M A, et al. Separating respiratory-variation-related fluctuations from neuronal-activity-related fluctuations in fMRI[J]. Neuroimage, 2006, 31(4): 1536-1548.
- [44] CHANG C, CUNNINGHAM J P, GLOVER G H. Influence of heart rate on the BOLD signal: The cardiac response function [J]. Neuroimage, 2009, 44(3): 857-869.
- [45] SHMUELI K, VAN GELDEREN P, DE ZWART J A, et al. Low-frequency fluctuations in the cardiac rate as a source of variance in the resting-state fMRI BOLD signal[J]. Neuroimage, 2007, 38(2): 306-320.

作者简介:



李栋(1996-),男,硕士研究生,研究方向:生物医学信号处理、机器学习, E-mail: lidongim@126.com。



吴海锋(1977-),通信作者,男,博士,教授,研究方向:深度学习、生物信号处理, E-mail:whf5469@gmail.com。



保涵(1998-),女,硕士研究生,研究方向:生物医学信号处理、机器学习, E-mail: baohanhan@163.com。



马佳(1999-),女,硕士研究生,研究方向:生物医学信号处理、机器学习, E-mail: 1609186984@qq.com。



曾玉(1981-),女,讲师,博士研究生,研究方向:生物医学信号处理, E-mail:yv.zeng@gmail.com。

(编辑:刘彦东)