



基于广义稀疏逻辑回归的全脑分类

王敬, 张宝, 谢晓, 李健, 张莉, 郭华平

引用本文:

王敬, 张宝, 谢晓, 李健, 张莉, 郭华平. 基于广义稀疏逻辑回归的全脑分类[J]. 信阳师范学院学报自然科学版, 2022, 35(3): 488–493. doi: 10.3969/j.issn.1003–0972.2022.03.024

WANG Jing,ZHANG Bao,XIE Xiao,LI Jian,ZHANG Li,GUO Huaping. Whole-brain Classification Based on Generalized Sparse Logistic Regression[J]. Journal of Xinyang Normal University (Natural Science Edition), 2022, 35(3): 488–493. doi: 10.3969/j.issn.1003–0972.2022.03.024

在线阅读 View online: <https://doi.org/10.3969/j.issn.1003–0972.2022.03.024>

您可能感兴趣的其他文章

Articles you may be interested in

Volterra核优化的SRC人脸识别算法

Face Recognition Based on SRC with Volterra Kernels Optimization

信阳师范学院学报自然科学版, 2022, 35(1): 141–144. <https://doi.org/10.3969/j.issn.1003–0972.2022.01.024>

不平衡类分类问题的逻辑判别式算法

Logistic Discrimination Algorithms for Imbalance Classification Problems

信阳师范学院学报自然科学版, 2016, 29(2): 274–278. <https://doi.org/10.3969/j.issn.1003–0972.2016.02.029>

保泛化性能的最优上边界回归模型辨识

Optimal Upper Boundary Regression Model Identification with Generalization Performance–Guaranteed

信阳师范学院学报自然科学版, 2020, 33(3): 454–461. <https://doi.org/10.3969/j.issn.1003–0972.2020.03.020>

低维显式语义空间下的语义关联度计算方法

Semantic Relatedness Computation Method Under Low Dimensional Explicit Semantic Space

信阳师范学院学报自然科学版, 2019, 32(4): 675–682. <https://doi.org/10.3969/j.issn.1003–0972.2019.04.029>

相位一致性加权的引导图像滤波去噪算法

Weighted Guided Image Filtering Algorithm Using Phase Congruency for Image Denoising

信阳师范学院学报自然科学版, 2017, 30(3): 464–468. <https://doi.org/10.3969/j.issn.1003–0972.2017.03.025>

基于广义稀疏逻辑回归的全脑分类

王 敬^{a,b,*}, 张 宝^a, 谢 晓^{a,b}, 李 健^{a,b}, 张 莉^{a,b}, 郭华平^{a,b}

(信阳师范学院 a. 计算机与信息技术学院; b. 河南省教育大数据分析与应用重点实验室, 河南 信阳 464000)

摘要:已有的稀疏逻辑回归算法不能充分利用大脑的空间结构信息,因此不适合直接应用于全脑 fMRI 数据分类。针对此问题,将表征稀疏性的惩罚项和表征空间结构的惩罚项同时引入逻辑回归算法中,提出了一类广义稀疏逻辑回归算法。具体而言,设计了一种灵活地表征空间结构的惩罚项。通过对该惩罚项调节参数,可以更加充分地利用大脑的空间结构信息。在优化最大化框架下,设计了一个迭代流程来求解对应的优化问题,保证每步迭代都有显式解,并且最终能得到该问题的局部最优解。实验结果表明,所提出的算法相对于已有的稀疏逻辑回归算法在分类准确率上具有明显优势。

关键词:全脑分类;逻辑回归;稀疏;优化最大化;空间结构

中图分类号:TP391 **文献标识码:**A **开放科学(资源服务)标识码(OSID):** 

Whole-brain Classification Based on Generalized Sparse Logistic Regression

WANG Jing^{1,2}, ZHANG Bao¹, XIE Xiao^{1,2}, LI Jian^{1,2}, ZHANG Li^{1,2}, GUO Huaping^{1,2}

(a. College of Computer and Information Technology; b. Henan Key Laboratory of Analysis and Applications of Education Big Data, Xinyang Normal University, Xinyang 464000, China)

Abstract:Previously proposed sparse logistic regression algorithms cannot make the most of the spatial structure information of brain, so it is not appropriate to be applied directly in whole-brain fMRI data classification. To address the problem, two penalty terms of characterizing sparsity and spatial structure are simultaneously introduced into traditional logistic regression algorithm, which produces a generalized sparse logistic regression algorithm. Specifically, a flexible penalty term that characterizes the spatial structure is designed. By tuning parameters of this penalty term, the spatial structure information of brain can be fully utilized. Under the framework of minorization-maximization, an iterative process is designed to solve the corresponding optimization problem, which guarantees that an explicit solution can be obtained in each iteration and a locally optimal solution for the problem can be found. Experimental results show that the proposed algorithm outperforms existing sparse logistic regression algorithms in classification accuracy.

Key words: whole-brain classification; logistic regression; sparsity; minorization-maximization; spatial structure

0 引言

大脑解码,即从测量得到的生理数据推断大脑的认知状态,是揭示人脑奥秘的一个重要途径^[1-2]。大量神经影像学研究已经证实了基于功能磁共振成像(functional magnetic resonance imaging, fMRI)数据预测被试大脑认知状态的可行性^[3]。

在相关研究中,传统的机器学习算法如支持向量机、线性判别分析、逻辑回归等得到了广泛应用^[4-5]。然而这些算法不能充分利用 fMRI 数据的特点,因此将它们直接应用于大脑解码时存在局限性。

fMRI 数据的主要特点包括维度高、样本数目少、蕴含空间结构等。传统机器学习算法应用到高维数据上时容易碰到维数灾难的问题^[6],通常解决

收稿日期:2021-07-21;修订日期:2021-11-19; * 通信联系人, E-mail: wangjing@xynu.edu.cn

基金项目:国家自然科学基金项目(31900710.62006205);河南省自然科学基金项目(222300420275);信阳师范学院青年科研基金项目(2021-QN-036);信阳师范学院“南湖学者奖励计划”青年项目

作者简介:王敬(1989—),男,湖北麻城人,讲师,博士,硕士生导师,主要从事神经影像学和机器学习研究。

这个问题的方法是通过主成分分析来降维,或者通过特征选择方法来提取与类别相关的特征。当数据维度高并且样本数目少时,传统的机器学习算法又容易产生过拟合的问题^[6]。解决这个问题的常用方法是在分类模型中加入惩罚项。fMRI 数据的前两个特点,即维度高和样本数目少,并非 fMRI 数据独有。比如在人脸识别中,人脸数据也常具备这两种特点。因此,维数灾难和过拟合的问题在机器学习和模式识别领域已经有了广泛和成熟的研究^[6]。针对这两个问题,一种行之有效的解决方法是引入稀疏惩罚项^[7],稀疏惩罚项能够使得与分类无关的特征得到抑制,并且能够有效地避免算法的过拟合。已经有大量的稀疏算法^[8-10]被应用到 fMRI 数据分析上。

对传统机器学习算法进行稀疏化改造,虽然可以避免维数灾难和过拟合问题,但仍无法利用到大脑的空间结构信息,因此在应用于全脑 fMRI 数据分类时存在不足。传统的机器学习算法通常要求将每个样本,比如二维人脸图像、三维全脑 fMRI 数据等,排列成一个一维向量,然后才能进行分析。将多维样本排列成一维向量会使得样本的维度信息丢失。人脸识别中的这个问题可以通过构造二维算法得到有效解决^[11]。通过利用人脸图像的二维空间结构,该类算法能够在使用较少特征的前提下有效地提高分类准确率,然而这种思路无法直接推广到三维全脑 fMRI 数据上。要利用 fMRI 数据的空间结构,一种可行的方法是将表征空间结构的惩罚项引入到机器学习算法中^[9,12-13]。

基于以上思路,本文在经典逻辑回归(Logistic Regression, LR)^[8]算法的基础上,同时引入表征稀疏的惩罚项和表征空间结构的惩罚项,构建一种新的全脑 fMRI 数据分类算法,即广义稀疏逻辑回归(Generalized Sparse Logistic Regression, GSLR)算法,然后在优化最大化框架下,设计了一个迭代流程来求解该算法对应的优化问题,最后通过实验来证明该算法的有效性。

1 广义稀疏逻辑回归算法

1.1 稀疏逻辑回归算法

逻辑回归算法^[8]常用于二分类。假设有 n 个样本类别对 $(\mathbf{x}_i, y_i)$ ($i=1, 2, \dots, n$), 其中 $\mathbf{x}_i \in \mathbb{R}^d$ 为样本, $y_i \in \{-1, 1\}$ 为类别, 样本之间互相独立且满足同一分布。定义 $\mathbf{X}=[\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$, $\mathbf{y}=[y_1, y_2, \dots, y_n]^T \in \mathbb{R}^n$ 。定义 Sigmoid 函数

$$\sigma(x) = \frac{1}{1 + \exp(-x)} \quad (1)$$

根据逻辑回归模型,在已知权重向量 $\mathbf{w} \in \mathbb{R}^d$ 和截距 $w_0 \in \mathbb{R}$ 时,样本 $\mathbf{x}_i$ 属于类别 y_i 的概率可以表示为

$$P(y_i | w_0, \mathbf{w}, \mathbf{x}_i) = \frac{1}{1 + \exp(-y_i(w_0 + \mathbf{w}^T \mathbf{x}_i))} = \sigma(y_i(w_0 + \mathbf{w}^T \mathbf{x}_i)) \quad (2)$$

已知训练样本及其对应的类别,通过最大化后验概率计算出 $\mathbf{w}$,然后将测试样本代入式(2),计算出该样本属于某一类别的概率,即可达到对测试样本进行分类的目的。通常构造增广矩阵来避免考虑 w_0 , 即 $\mathbf{x}_i \leftarrow [1; \mathbf{x}_i]$, $\mathbf{w} \leftarrow [w_0; \mathbf{w}]$, 从而得到联合概率密度分布函数:

$$P(\mathbf{y} | \mathbf{w}, \mathbf{X}) = \prod_{i=1}^n \frac{1}{1 + \exp(-y_i \mathbf{w}^T \mathbf{x}_i)} = \prod_{i=1}^n \sigma(y_i \mathbf{w}^T \mathbf{x}_i) \quad (3)$$

根据贝叶斯理论,在已知 $\mathbf{X}$ 和 $\mathbf{y}$ 的前提下, $\mathbf{w}$ 的似然函数可以表示为

$$P(\mathbf{w} | \mathbf{y}, \mathbf{X}) \propto P(\mathbf{y} | \mathbf{w}, \mathbf{X}) P(\mathbf{w}), \quad (4)$$

最大化该似然函数即可求出 $\mathbf{w}$ 。

最大化式(4)中的似然函数,首先需要确定先验因子 $P(\mathbf{w})$ 。假设权重之间不相关时,由广义正态分布理论,权重 w_j ($j=1, 2, \dots, d$) 的概率密度函数为

$$P(w_j) = \frac{\beta}{2\alpha \Gamma(1/\beta)} \exp\left(-\left(\frac{|w_j - \mu|}{\alpha}\right)^\beta\right), \quad (5)$$

其中 α, β, μ 为调节参数。取均值 $\mu=0$, 调节 β 可以得到两种常见的先验因子。 $\beta=1$ 对应拉普拉斯先验因子:

$$P_{\text{Laplacian}}(\mathbf{w}) \propto \exp(-\lambda_1 \|\mathbf{w}\|_1) \quad (6)$$

$\beta=2$ 对应高斯先验因子:

$$P_{\text{Gaussian}}(\mathbf{w}) \propto \exp(-\lambda_2 \|\mathbf{w}\|_2^2) \quad (7)$$

将两种先验因子同时应用于式(4)中,然后对似然函数取对数可得式(8)的优化问题

$$\mathbf{w} = \arg \max_{\mathbf{w}} \ln P(\mathbf{y} | \mathbf{w}, \mathbf{X}) - \lambda_1 \|\mathbf{w}\|_1 - \lambda_2 \|\mathbf{w}\|_2^2, \quad (8)$$

即同时带有 l_1 范数和 l_2 范数的稀疏逻辑回归算法,记作 LR12。当 $\lambda_1=0$ 时,LR12 退化为普通的逻辑回归算法,记作 LR2。当 $\lambda_2=0$ 时,LR12 退化为仅带 l_1 范数惩罚项的逻辑回归算法,记作 LR1。

拉普拉斯先验因子和高斯先验因子分别对应 LR12 中的 l_1 范数和 l_2 范数惩罚项。 l_2 范数惩罚项也称岭(Ridge)惩罚项,能抑制无关特征对应的权重,避免过拟合。 l_1 范数惩罚项也称拉索(LASSO)惩罚项,能使无关特征对应的权重缩减至零,从而得到稀疏的结果。岭惩罚项和拉索惩罚项结合在一起则构成弹性网(Elastic-net)惩罚项。该惩罚项同时具有岭惩罚项和拉索惩罚项的优点,因而得到了广泛应用。

1.2 空间结构惩罚项

为了利用大脑的空间结构信息,在高斯先验因子中对权重之间的相关性进行如下建模。假设两个特征 w_i 和 w_j 的坐标分别为 (a_i, b_i, c_i) 和 (a_j, b_j, c_j) , 则 w_i 和 w_j 在三维空间中的距离为

$$d_{ij} = \|(a_i, b_i, c_i) - (a_j, b_j, c_j)\|_2. \quad (9)$$

定义邻接矩阵 $N = (N_{ij})$,

$$N_{ij} = \begin{cases} \exp(-\frac{d_{ij}^2}{2\delta^2}), d_{ij} \leq \epsilon, \\ 0, \text{其他}, \end{cases} \quad (10)$$

其中 ϵ 用来调节矩阵 N 的稀疏程度, ϵ 越小, N 越稀疏; δ 用来调节矩阵 N 中非零元素的大小, δ 越小, N 中非零元素的值越小。在矩阵 N 的基础上定义精度矩阵 Q^{-1} 如下:

$$Q^{-1} = D - N, \quad (11)$$

其中 $D = \text{diag}(\mathbf{1}_d^T N)$, $\mathbf{1}_d$ 表示一个 d 行的全 1 向量, 则有

$$\mathbf{w}^T Q^{-1} \mathbf{w} = \frac{1}{2} \sum_{i,j=1}^d (w_i - w_j)^2 N_{ij}. \quad (12)$$

以式(12)作为惩罚项代替式(8)中的 l_2 范数惩罚项可得

$$\mathbf{w} = \arg \max_{\mathbf{w}} \ln P(\mathbf{y} | \mathbf{w}, \mathbf{X}) - \lambda_1 \|\mathbf{w}\|_1 - \lambda_2 \mathbf{w}^T Q^{-1} \mathbf{w}, \quad (13)$$

即广义稀疏逻辑回归(Generalized Sparse Logistic Regression, GSLR)算法。

GSLR 算法基于两个特征的空间距离来调节两者权重的残差。直观地讲,两个特征 w_i 和 w_j 在空间上距离越近,由式(10)定义的邻接矩阵中对应的元素 N_{ij} 越大,最大化式(13)时得到 w_i 和 w_j 的残差就越小。因此, GSLR 可以保证特征 $\mathbf{w}$ 在空间上的连续性。

GSLR 模型是已有的稀疏逻辑回归模型的泛化形式。当 Q 为单位阵时, GSLR 退化为 LR12。自然地, GSLR 也是 LR2 和 LR1 的泛化形式。当

$$N_{ij} = \begin{cases} 1, d_{ij} = 1, \\ 0, \text{其他} \end{cases} \quad (14)$$

时 GSLR 算法退化为带 GraphNet 惩罚项的逻辑回归算法^[12]。在不至于混淆的前提下,本文中不带 GraphNet 惩罚项的逻辑回归算法记作 GraphNet。相对于其他几种稀疏逻辑回归算法, GSLR 算法能更充分地利用特征的空间信息,预期可以在全脑分类上取得更好的效果。

GSLR 与平滑稀疏逻辑回归(Smooth Sparse Logistic Regression, SSLR)^[9]的主要区别在于构造了不同的空间结构惩罚项。GSLR 中先用空间距离的高斯函数定义邻接矩阵 N , 然后通过式(11)计算出精度矩阵 Q^{-1} , 而 SSLR 中直接使用空间距离的高斯函数来计算协方差矩阵 Q 。因此, GSLR 可以退化为 GraphNet, 而 SSLR 则不行。另外, GSLR 中可以通过调节参数 ϵ 改变邻接矩阵 N 的稀疏度, 从而得到稀疏的精度矩阵 Q^{-1} , 而 SSLR 中没有设置对应参数, 计算得到的协方差矩阵 Q 是稠密的。考虑到全脑 fMRI 数据维度很高, 稠密的协方差矩阵在存储和运算时都需要消耗更多的硬件资源。基于以上原因, GSLR 选择与 SSLR 不同的方式来构造空间结构惩罚项。

1.3 迭代求解方法

在优化最大化(minimization-maximization, MM)^[14]框架下构造 GLSR 算法对应优化问题的迭代求解方法, 令

$$l(\mathbf{w}) = \ln P(\mathbf{y} | \mathbf{w}, \mathbf{X}), \quad (15)$$

则 GSLR 的目标函数可以写为

$$f(\mathbf{w}) = l(\mathbf{w}) - \lambda_1 \|\mathbf{w}\|_1 - \lambda_2 \mathbf{w}^T Q^{-1} \mathbf{w}. \quad (16)$$

为方便调节 $\|\mathbf{w}\|_1$ 和 $\mathbf{w}^T Q^{-1} \mathbf{w}$ 两者间的相对权重, 引入参数 λ 和 ρ , 将式(16)改写为

$$f(\mathbf{w}) = l(\mathbf{w}) - \lambda(1 - \rho) \|\mathbf{w}\|_1 - \lambda \rho \mathbf{w}^T Q^{-1} \mathbf{w}. \quad (17)$$

对 $l(\mathbf{w})$ 在 $\mathbf{w}^{(k)}$ 处进行二阶泰勒展开, 其中 $\mathbf{w}^{(k)}$ 是 $\mathbf{w}$ 在第 k 次迭代后得到的结果, 由中值定理可知, 存在 $\theta \in [0, 1]$, 使得

$$l(\mathbf{w}) = l(\mathbf{w}^{(k)}) + (\mathbf{w} - \mathbf{w}^{(k)})^T \frac{\partial l(\mathbf{w}^{(k)})}{\partial \mathbf{w}} + \frac{1}{2} (\mathbf{w} - \mathbf{w}^{(k)})^T \frac{\partial^2 l(\theta \mathbf{w} + (1 - \theta) \mathbf{w}^{(k)})}{\partial \mathbf{w} \partial \mathbf{w}^T} (\mathbf{w} - \mathbf{w}^{(k)}). \quad (18)$$

定义

$$\mathbf{s} = [\sigma(y_1 \mathbf{w}^T \mathbf{x}_1), \dots, \sigma(y_n \mathbf{w}^T \mathbf{x}_n)]^T = \sigma(\mathbf{y} \circ (\mathbf{X}^T \mathbf{w}^{(k)})), \quad (19)$$

其中 $\mathbf{a} \circ \mathbf{b}$ 表示 Hadamard 积, 即向量 $\mathbf{a}$ 和向量 $\mathbf{b}$

对应元素相乘得到的向量; Sigmoid 函数 $\sigma(\cdot)$ 作用在向量上, 表示对向量中的每个元素分别求 Sigmoid 函数, 则 $l(\mathbf{w})$ 的梯度和 Hessian 矩阵分别为

$$\mathbf{g}(\mathbf{w}) = \frac{\partial l(\mathbf{w})}{\partial \mathbf{w}} = \mathbf{X} [(\mathbf{1}_n - \mathbf{s}) \circ \mathbf{y}], \quad (20)$$

$$\mathbf{H}(\mathbf{w}) = \frac{\partial^2 l(\mathbf{w})}{\partial \mathbf{w} \partial \mathbf{w}^T} = \mathbf{X} \text{diag} [-(\mathbf{1}_n - \mathbf{s}) \circ \mathbf{s}] \mathbf{X}^T, \quad (21)$$

其中 $\mathbf{1}_n$ 表示一个 n 行的全 1 向量。由式(21)可得

$$\mathbf{H}(\mathbf{w}) \geq \mathbf{B} = -\frac{1}{4} \mathbf{X} \mathbf{X}^T, \quad \forall \mathbf{w}. \quad (22)$$

对于两个同样规模的矩阵 $\mathbf{A}$ 和 $\mathbf{B}$, $\mathbf{A} \geq \mathbf{B}$ 表示 $\mathbf{A} - \mathbf{B}$ 为半正定矩阵。将式(20)和式(21)代入式(18), 再结合式(22)可得

$$l(\mathbf{w}) \geq l(\mathbf{w}^{(k)}) + (\mathbf{w} - \mathbf{w}^{(k)})^T \mathbf{g}(\mathbf{w}^{(k)}) + \frac{1}{2} (\mathbf{w} - \mathbf{w}^{(k)})^T \mathbf{B} (\mathbf{w} - \mathbf{w}^{(k)}). \quad (23)$$

另外, 由文献[11]知,

$$\|\mathbf{w}\|_1 \leq \frac{1}{2} (\mathbf{w}^T \mathbf{U} \mathbf{w} + \|\mathbf{w}^{(k)}\|_1), \quad (24)$$

其中 $\mathbf{U} = \text{diag}(\text{abs}(\mathbf{w}^{(k)}))^{-1}$ 。将式(23)和式(24)应用到式(17)中等式的右侧构造如下函数:

$$\begin{aligned} g(\mathbf{w} | \mathbf{w}^{(k)}) &= l(\mathbf{w}^{(k)}) + (\mathbf{w} - \mathbf{w}^{(k)})^T \mathbf{g}(\mathbf{w}^{(k)}) + \\ &\frac{1}{2} (\mathbf{w} - \mathbf{w}^{(k)})^T \mathbf{B} (\mathbf{w} - \mathbf{w}^{(k)}) - \\ &\frac{1}{2} \lambda (1 - \rho) (\mathbf{w}^T \mathbf{U} \mathbf{w} + \|\mathbf{w}^{(k)}\|_1) - \\ &\lambda \rho \mathbf{w}^T \mathbf{Q}^{-1} \mathbf{w}. \end{aligned} \quad (25)$$

该函数满足 MM 框架的两个条件, 因此是一个合理的中间函数。根据 MM 框架, 可以通过式(26)迭代地最大化 $g(\mathbf{w} | \mathbf{w}^{(k)})$ 来达到最大化 $f(\mathbf{w})$ 的目的,

$$\mathbf{w}^{(k+1)} = \arg \max_{\mathbf{w}} g(\mathbf{w} | \mathbf{w}^{(k)}). \quad (26)$$

去掉 $g(\mathbf{w} | \mathbf{w}^{(k)})$ 中与 $\mathbf{w}$ 的无关项得

$$\begin{aligned} \hat{g}(\mathbf{w} | \mathbf{w}^{(k)}) &= \\ &\frac{1}{2} \mathbf{w}^T (\mathbf{B} - \lambda (1 - \rho) \mathbf{U} - 2\lambda \rho \mathbf{Q}^{-1}) \mathbf{w} + \\ &\mathbf{w}^T (\mathbf{g}(\mathbf{w}^{(k)}) - \mathbf{B} \mathbf{w}^{(k)}). \end{aligned} \quad (27)$$

最大化 $\hat{g}(\mathbf{w} | \mathbf{w}^{(k)})$ 的问题是一个关于 $\mathbf{w}$ 的二次优化问题, 存在显式解, 即

$$\begin{aligned} \mathbf{w}^{(k+1)} &= (\mathbf{B} - \lambda (1 - \rho) \mathbf{U} - 2\lambda \rho \mathbf{Q}^{-1})^{-1} \cdot \\ &(\mathbf{B} \mathbf{w}^{(k)} - \mathbf{g}(\mathbf{w}^{(k)})). \end{aligned} \quad (28)$$

GSLR 的优化过程如算法 1 所示。

算法 1 GSLR 算法

输入: 训练样本 $\mathbf{X}$, 类别 $\mathbf{y}$

初始化: $\mathbf{w}^{(0)} = \mathbf{1}$, $k = 0$; $\lambda, \rho, \epsilon, \delta, \eta = 1$,

$$N_{ij} = \begin{cases} \exp(-\frac{d_{ij}^2}{2\delta^2}), & d_{ij} \leq \epsilon, \\ 0, & \text{其他}, \end{cases}$$

$$\mathbf{D} = \text{diag}(\mathbf{1}_d^T \mathbf{N})$$

$$\mathbf{Q}^{-1} = \mathbf{D} - \mathbf{N}$$

$$\mathbf{B} = -\frac{1}{4} \mathbf{X} \mathbf{X}^T$$

while $\eta > 10^{-4}$

$$\mathbf{s} = \sigma(\mathbf{y} \circ (\mathbf{X}^T \mathbf{w}^{(k)}))$$

$$\mathbf{g}(\mathbf{w}^{(k)}) = \mathbf{X} [(\mathbf{1}_n - \mathbf{s}) \circ \mathbf{y}]$$

$$\mathbf{w}^{(k+1)} = (\mathbf{B} - \lambda (1 - \rho) \mathbf{U} - 2\lambda \rho \mathbf{Q}^{-1})^{-1} \cdot (\mathbf{B} \mathbf{w}^{(k)} - \mathbf{g}(\mathbf{w}^{(k)}))$$

$$\eta = \|\mathbf{w}^{(k+1)} - \mathbf{w}^{(k)}\|_2 / \|\mathbf{w}^{(k)}\|_2$$

$$k \leftarrow k + 1$$

end while

输出: $\mathbf{w}$

使用训练样本 $\mathbf{X}$ 及其类别 $\mathbf{y}$ 训练出 $\mathbf{w}$ 后, 将测试样本代入式(2)中计算出该样本属于某一类别的概率, 以此预测该样本所属类别。最后, 结合预测类别和实际类别统计出分类准确率。

2 实验

2.1 fMRI 数据集

实验中使用了两个公开数据集, 即 StarPlus 数据集^[15]和视觉物体识别(Visual Object Recognition, VOR)数据集^[16]。关于两个数据集的详细描述如下。

StarPlus 数据集采集的是被试观看图片和句子时的大脑 fMRI 数据。实验中选取了 6 个被试的数据, 每个被试完成 40 个任务。每个任务要求被试首先看 4 s 句子/图片, 接着看 4 s 空白屏幕, 然后看 4 s 图片/句子, 最后休息 15 s。前 20 个任务先看句子后看图片, 后 20 个任务先看图片后看句子。实验过程中, 每秒采集两次大脑 fMRI 数据, 仅扫描部分脑区, 这些脑区对应 20~30 个解剖学上的区域。从 StarPlus 数据集中随机选取一个被试的某个时刻的大脑图像, 是一个三维的灰度图像, 其典型的切片图如图 1 所示。



图 1 StarPlus 数据示例图

Fig. 1 An illustration of the StarPlus data

VOR 数据集采集的是被试观看 8 种类别的图片时的全脑 fMRI 数据。8 种类别的图片包括脸、房子、猫、瓶子、剪刀、鞋、椅子和混乱的图片。共 6 个被试参与实验,每个被试完成 12 个任务,每个任务选取 8 类图片各 1 张,随机打乱顺序,然后依次给被试观看。每张图片展示 0.5 s,间隔 1.5 s。从 VOR 数据集中随机选取一个被试,其典型的大脑图像如图 2 所示。

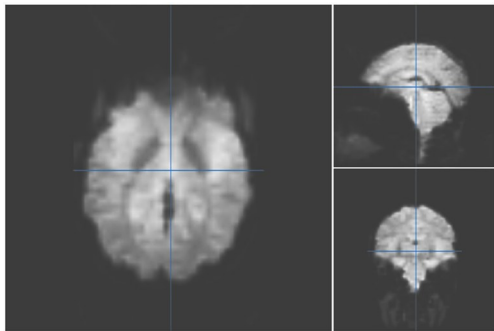


图 2 VOR 数据示例图

Fig. 2 An illustration of the VOR data

2.2 fMRI 数据预处理

对于 StarPlus 数据集,网上公开的是已经预处理好的数据,无须额外的预处理步骤。对于 VOR 数据集,使用集成 SPM12 的静息态 fMRI 数据预处理工具 DPABI^[17],对原始数据进行预处理,包括以下步骤:首先校正不同切片的扫描时间差异,接着对齐图像以校正扫描过程中的头部运动,然后使用蒙特利尔神经学研究所模板大脑对图像进行标准化,使用三线性插值将图像重采样使得新图像分辨率为 4 mm 的立方,最后去除线性漂移。

2.3 实验设置

对 StarPlus 数据集和 VOR 数据集中每个被试的数据进行十重交叉验证,并计算平均分类准确

率。具体地讲,将每个被试的所有样本随机等分为 10 份,每次取不同的 1 份用于测试,其他用于训练,计算分类准确率,然后将得到的 10 次分类准确率取平均,从而得到对于单个被试的分类准确率。将一个数据集中的所有被试对应的分类准确率取平均,得到对于该数据集的平均分类准确率,用来衡量算法的分类准确性。

实验中选取 LR1^[8]、LR12^[8] 和 GraphNet^[12] 作为 GSLR 的对比方法。GSLR 包括 4 个调节参数 λ 、 ρ 、 ϵ 和 δ 。参数 λ 用于调节惩罚项的总权重,从 $\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3\}$ 中选取。参数 ρ 用于调节表征稀疏的惩罚项 $\|w\|_1$ 和表征空间结构的惩罚项 $w^T Q^{-1} w$ 两者的相对权重,从 $0 \sim 1$ 中选取,步长为 0.1。参数 ϵ 用于调节矩阵 N 的稀疏度,从 $\{1, \sqrt{2}, \sqrt{3}\}$ 中选取。参数 δ 用于调节矩阵 N 中非零元素的大小,从 $0.5 \sim 1.5$ 中选取,步长为 0.1。LR1 仅有参数 λ 。LR12 和 GraphNet 都包含两个参数,即 λ 和 ρ 。三种对比方法的参数选取范围跟 GSLR 一致。

2.4 实验分析

对于 StarPlus 数据集,当 $\lambda = 10$ 时,LR1 的最高分类准确率为 61.25%;当 $\lambda = 10$ 、 $\rho = 0.4$ 时,LR12 的最高分类准确率为 68.75%;当 $\lambda = 10$ 、 $\rho = 0.3$ 时,GraphNet 的最高分类准确率为 75.00%;当 $\epsilon = \sqrt{3}$ 、 $\delta = 1.2$ 、 $\lambda = 10$ 、 $\rho = 0.75$ 时,GSLR 的最高分类准确率为 75.00%。4 种对比算法在 StarPlus 数据集上的最高分类准确率如表 1 所示。

表 1 4 种算法在 StarPlus 数据集上的分类准确率

Tab. 1 The classification accuracies of four algorithms on the StartPlus dataset

算法	准确率/%
LR1	61.25
LR12	68.75
GraphNet	75.00
GSLR	75.00

对于 VOR 数据集,当 $\lambda = 10$ 时,LR1 的最高分类准确率为 62.50%;当 $\lambda = 10$ 、 $\rho = 0.5$ 时,LR12 的最高分类准确率为 70.83%;当 $\lambda = 10$ 、 $\rho = 0.4$ 时,GraphNet 的最高分类准确率为 72.22%;当 $\epsilon = \sqrt{3}$ 、 $\delta = 1.1$ 、 $\lambda = 10$ 、 $\rho = 0.5$ 时,GSLR 的最高分类准确率为 72.22%。4 种对比算法在 VOR 数据集上的最高分类准确率如表 2 所示。

表 2 4 种算法在 VOR 数据集上的分类准确率
Tab. 2 The classification accuracies of four algorithms on the VOR dataset

算法	准确率/%
LR1	62.50
LR12	70.83
GraphNet	72.22
GSLR	72.22

GSLR 的最高分类准确率超过 LR1 和 LR12 的最高分类准确率,与 GraphNet 的最高分类准确率相等。与 GraphNet 相比,GSLR 在构造邻接矩阵 $\mathbf{N}$ 时更加灵活:可以通过调节参数 ϵ 来调节改变邻接矩阵的稀疏度;可以通过调节参数 δ 可以调节邻接矩阵中非零元素的大小。因此,GSLR 能够更加充分地利用到大脑的空间结构信息,在特征选择方面相对于 GraphNet 具有优势。

参考文献:

[1] XU L C, XU M P, JUNG T P, et al. Review of brain encoding and decoding mechanisms for EEG-based brain-computer interface[J]. Cognitive Neurodynamics, 2021, 15(4): 569-584.

[2] CAO L, HUANG D D, ZHANG Y, et al. Brain decoding using fNIRS[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(14): 12602-12611.

[3] HUANG S, SHAO W, WANG M L, et al. fMRI-based decoding of visual information from human brain activity: A brief review[J]. International Journal of Automation and Computing, 2021, 18(2): 170-184.

[4] KHOSLA M, JAMISON K, NGO G H, et al. Machine learning in resting-state fMRI analysis[J]. Magnetic Resonance Imaging, 2019, 64: 101-121.

[5] COHEN A D, CHEN Z Y, PARKER JONES O, et al. Regression-based machine-learning approaches to predict task activation using resting-state fMRI[J]. Human Brain Mapping, 2020, 41(3): 815-826.

[6] DUDA R O, HART P E, STORK D G. Pattern classification[M]. 2nd Ed. Hoboken, New Jersey: Wiley-Interscience, 2000: 27-32.

[7] ZHANG Z, XU Y, YANG J, et al. A survey of sparse representation: Algorithms and applications[J]. IEEE Access, 2015, 3: 490-530.

[8] RYALI S, SUPEKAR K, ABRAMS D A, et al. Sparse logistic regression for whole-brain classification of fMRI data[J]. NeuroImage, 2010, 51(2): 752-764.

[9] DE BRECHT M, YAMAGISHI N. Combining sparseness and smoothness improves classification accuracy and interpretability[J]. NeuroImage, 2012, 60(2): 1550-1561.

[10] WEN Z F, YU T Y, YU Z L, et al. Grouped sparse Bayesian learning for voxel selection in multivoxel pattern analysis of fMRI data[J]. NeuroImage, 2019, 184: 417-430.

[11] WANG J. Generalized 2-D principal component analysis by Lp-Norm for image analysis[J]. IEEE Transactions on Cybernetics, 2016, 46(3): 792-803.

[12] GROSENICK L, KLINGENBERG B, KATOVICH K, et al. Interpretable whole-brain prediction analysis with GraphNet[J]. NeuroImage, 2013, 72: 304-321.

[13] WATANABE T, KESSLER D, SCOTT C, et al. Disease prediction based on functional connectomes using a scalable and spatially-informed support vector machine[J]. NeuroImage, 2014, 96: 183-202.

[14] HUNTER D R, LANGE K. A tutorial on MM algorithms[J]. The American Statistician, 2004, 58(1): 30-37.

[15] MITCHELL T M, HUTCHINSON R, NICULESCU R S, et al. Learning to decode cognitive states from brain images[J]. Machine Learning, 2004, 57(1): 145-175.

[16] HAXBY J V, GOBBINI M I, FUREY M L, et al. Distributed and overlapping representations of faces and objects in ventral temporal cortex[J]. Science, 2001, 293(5539): 2425-2430.

[17] 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.

责任编辑:郭红建

3 总结和展望

将表征稀疏的惩罚项和表征空间结构的惩罚项同时引入逻辑回归中,特别是设计了一种具有一般性的表征空间结构的惩罚项,从而提出了一种新的广义稀疏逻辑回归算法。该算法是多种已有的稀疏逻辑回归算法的泛化形式。实验结果表明,本文提出的算法相对于已有算法在全脑 fMRI 分类上具有明显优势。另外,由于本文提出的算法能够更加充分有效地利用大脑的空间结构信息,因此相对于其他方法更适合用于特征选择。

本文存在一些不足之处需要在今后的研究加以改进。一是使用到的数据种类和数量都有限,没有充分体现不同算法之间的差距。二是有必要以更精细的粒度来调节邻接矩阵中的参数,从而找到最有利于分类和特征选择的参数组合。