

用于阿尔茨海默病诊断的权值分布特征学习*

程波¹, 丁毅², 张道强²



¹(重庆三峡学院 智能信息处理与控制重点实验室, 重庆 404020)

²(南京航空航天大学 计算机科学与技术学院, 江苏 南京 210016)

通讯作者: 张道强, E-mail: dqzhang@nuaa.edu.cn

摘要: 针对当前基于机器学习的早期阿尔茨海默病(AD)诊断中有标记训练样本不足的问题,提出一种基于多模态特征数据的权值分布稀疏特征学习方法,并将其应用于早期阿尔茨海默病的诊断.具体来说,该诊断方法主要包括两大模块:基于权值分布的 Lasso 特征选择模型(WDL)和大间隔分布分类机模型(LDM).首先,为了获取多模态特征之间的数据分布信息,对传统 Lasso 模型进行改进,引入权值分布正则化项,从而构建出基于权值分布的 Lasso 特征选择模型;然后,为了有效地利用多模态特征之间的数据分布信息,以保持多模态特征之间的互补性,直接采用大间隔分布学习算法训练分类器.选取国际阿尔茨海默症数据库(ADNI)中 202 个多模态特征的被试者样本进行实验,分类 AD 最高平均精度为 97.5%,分类轻度认知功能障碍(MCI)最高平均精度为 83.1%,分类轻度认知功能障碍转化为 AD(pMCI)最高平均精度为 84.8%.实验结果表明,所提 WDL 特征学习方法可从串联的多模态特征学到性能更优的特征子集,并能根据权值分布获取多模态特征之间的数据分布信息,从而提高早期阿尔茨海默病诊断的性能.

关键词: 权值分布;多模态;阿尔茨海默病;稀疏特征学习;大间隔分布学习

中图法分类号: TP18

中文引用格式: 程波,丁毅,张道强.用于阿尔茨海默病诊断的权值分布特征学习.软件学报,2019,30(4):1002-1014. <http://www.jos.org.cn/1000-9825/5371.htm>

英文引用格式: Cheng B, Ding Y, Zhang DQ. Feature learning of weight-distribution for diagnosis of Alzheimer's disease. Ruan Jian Xue Bao/Journal of Software, 2019, 30(4): 1002-1014 (in Chinese). <http://www.jos.org.cn/1000-9825/5371.htm>

Feature Learning of Weight-distribution for Diagnosis of Alzheimer's Disease

CHENG Bo¹, DING Yi², ZHANG Dao-Qiang²

¹(Key Laboratory of Intelligent Information Processing and Control of Chongqing Municipal Institutions of Higher Education, Chongqing Three Gorges University, Chongqing 404020, China)

²(College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract: In the field of medical imaging analysis using machine learning, the challenge is lack of training sample. In order to solve the problem, a weight-distribution based Lasso (Least absolute shrinkage and selection operator) feature learning model is proposed and applied to early diagnosis of Alzheimer's Disease (AD). Specifically, the proposed diagnosis method is consisted of two components: weight-distribution based Lasso feature selection (WDL) and large margin distribution machine (LDM) for classification. Firstly, in order to capture data

* 基金项目: 国家自然科学基金(61602072, 61422204, 61473149, 61732006, 61573023); 重庆市基础研究与前沿探索项目(cstc2016jcyjA0063, cstc2018jcyjAX0502, cstc2014jcyjA40035, cstc2014jcyjA1316, cstc2016jcyjA0521); 重庆市教委科学技术研究(KJ1501014, KJ1601003, KJ1710248, KJ1401010, KJ1601015); 重庆市高校市级重点实验室资助项目([2017]3)

Foundation item: National Natural Science Foundation of China (61602072, 61422204, 61473149, 61732006, 61573023); Chongqing Cutting-edge and Applied Foundation Research Program (cstc2016jcyjA0063, cstc2018jcyjAX0502, cstc2014jcyjA40035, cstc2014jcyjA1316, cstc2016jcyjA0521); Scientific and Technological Research Program of Chongqing Municipal Education Commission (KJ1501014, KJ1601003, KJ1710248, KJ1401010, KJ1601015); Chongqing Municipal Key Laboratory of Institutions of Higher Education ([2017]3)

收稿时间: 2016-12-05; 修改时间: 2017-01-28, 2017-05-30; 采用时间: 2017-08-24

distribution information among multimodal features, the WDL feature selection model was built, to improve on the conventional Lasso model via adding a regularization item of weight-distribution. Secondly, in order to achieve better generalization and accuracy on classification, and also to keep complementary information among multimodal features, the LDM algorithm is used for the training of the classifier. To evaluate the effectiveness of the proposed learning model, 202 subjects from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database with multimodal features were employed. Experimental results on the ADNI database show that it can recognize AD from Normal Controls (NC) with 97.5% accuracy, recognize Mild Cognitive Impairment (MCI) from NC with 83.1% accuracy, and recognize progressive MCI (pMCI) patients from stable MCI (sMCI) ones with 84.8% accuracy, which demonstrate that it can significantly improve the performance of early AD diagnosis and achieve feature ranking in terms of discrimination via optimized weight vector.

Key words: weight-distribution; multimodal; Alzheimer's disease (AD); sparse feature learning; large margin distribution learning

阿尔茨海默病(Alzheimer's disease,简称 AD),又称为老年痴呆症,是当前危害老年人健康的重要疾病之一。根据阿尔茨海默病的病理,尽早发现并治疗能有效延缓疾病发展进程。在医学上,通常将阿尔茨海默病的病程划分成3个阶段:首先是症状前AD,然后发展为轻度认知功能障碍(mild cognitive impairment,简称 MCI),最后逐步发展成AD。因此,对早期阿尔茨海默病(MCI或者症状前AD阶段)做出诊断将十分有助于治疗或延缓AD的发展进程。近年来,基于机器学习的脑图像分析方法与技术被越来越多地应用于早期阿尔茨海默病的诊断中^[1-12]。这些脑图像通常有结构磁共振成像(magnetic resonance imaging,简称 MRI)、正电子发射型计算机断层成像(positron emission tomography,简称 PET)等等。除了脑图像,生物标志物检测对早期阿尔茨海默病的诊断也具有较高的灵敏度和特异性,其中,典型的生物标志物有脑脊液(cerebrospinal fluid,简称 CSF)中的生化标记物水平检测。研究认为,综合使用脑图像与生物标志物能够有效提高早期阿尔茨海默病诊断的准确度与敏感度。所以,越来越多的多模态机器学习方法被应用于早期阿尔茨海默病的诊断中^[4,5,7,9,11-16]。

在医学实践中,收集大量脑图像来训练机器学习算法是一件费时且代价昂贵的任务。所以,通常情况下只有少量训练样本可以使用,但是,为了获得较好的诊断性能,很多研究采用了多模态脑图像与生物标志物来提高学习算法的性能^[4,5,7,9,11,12,14,16],然而这样又增加了特征维数。面对高维小样本的多模态学习问题,大量研究设计了特征选择来避免过学习^[4,5,7,9,13]。其中,有一些研究采用稀疏学习来设计多模态特征选择模型。例如,Zhang等人^[14]根据多模态脑图像与生物标志物特征之间对AD诊断具有互补性原理,提出一种多模态多任务学习方法,能从分类与回归学习任务中学习出共同的最优特征子集;Jie等人^[4]提出一种半监督多模态稀疏学习方法,可从标记和未标记的MRI与PET图像上学习出共同的最优特征子集;Zhu等人^[9]提出一种基于矩阵相似性度量的多任务学习方法,可同时从多模态多任务训练集中学习出最优特征子集。此外,Ye等人^[13]提出一种基于Lasso(least absolute shrinkage and selection operator)模型的稳定特征选择算法,可从串联的多模态脑图像与生物标志物上学习出稳定的最优特征子集。虽然以上研究都能获得判别性较好的特征子集,并能完成较好的分类性能,但是这些研究并未深入挖掘多模态特征之间的数据分布信息。

在当前基于神经图像的早期AD诊断中,现存很多研究通过设计稀疏特征学习模型以获取最优特征子集,其中有不少研究利用Lasso模型进行特征选择^[4,5,10,13,14,16],其原理是根据Lasso模型优化学到的权值进行特征筛选,大多数研究都是选择非零权值对应的特征,而无法获取多模态特征之间的数据分布信息。由于传统Lasso模型只包含权值的 L_1 范数正则化项,而它通常用于选择那些能降低分类器泛化误差的特征,无法获取多模态特征的数据分布信息。实际上,针对分类问题,Lasso模型中的权值可以理解为调整特征贡献度的权重值,若对传统Lasso模型进行改进,便可获取特征的数据分布信息。此外,一些研究认为,多模态脑图像与生物标志物特征之间对AD的诊断具有互补性^[4,5,7,9,11,12,14,16]。获取多模态特征的数据分布信息,就是为了保持从多模态特征中选出的特征子集仍具有良好的互补性。

传统支持向量机(support vector machine,简称 SVM)以其优良的分类性能,被广泛应用于基于神经图像的早期AD诊断中^[12,14,16,17],其核心思想是最大化“最小间隔”,以解决机器学习中“过学习”问题。然而现有一些研究表明,相比于最小间隔,“间隔分布”能够更好地降低分类器的泛化误差^[18-20]。因此,大间隔分布学习机(large margin distribution machine,简称 LDM)受此启发,提出通过同时最大化间隔均值并最小化间隔方差来优化间隔分布。另一方面,Lasso模型中的“权值”与SVM中的“间隔”对学习器的泛化性能起着相似的作用。此外,充分利用数据分

布或变量分布先验信息能够有效提高学习算法的性能,比如 Geng 提出的标记分布学习^[21]能够解决除了单标记与多标记学习之外更多标记学习问题.因此,为了保持多模态特征之间的互补性,应充分利用多模态特征之间的分布信息.为了从多模态特征获得泛化性能更好的特征子集,受大间隔分布学习^[18-20]与标记分布学习^[21]的共同启发,对传统 Lasso 模型进行改进,通过增加同时最小化权值均值并最大化权值方差来优化权值分布,以获取多模态特征之间的分布信息,提出了基于权值分布的 Lasso 特征选择模型(weight-distribution based Lasso,简称 WDL).

此外,由于多模态脑图像与生物标志物特征之间对 AD 诊断具有互补性,设计权值分布 Lasso 特征选择方法就是为了使选出的最优特征子集不但具有良好的泛化性与可分性,还要尽量保持多模态特征之间的互补性.相比于支持向量机 SVM,大间隔分布学习机 LDM 能够利用样本数据分布的先验信息,所以能够更有效地利用多模态特征之间的互补性.因此,LDM 更适合基于多模态特征的早期 AD 诊断.综合以上研究动机,本文提出了基于权值分布的稀疏特征学习框架,由于该框架结合了 WDL 特征选择模型和 LDM 分类器,因此将该分类学习模型表示为 WDL-LDM.为了验证本文提出的学习框架用于早期阿尔茨海默病诊断的有效性,采用国际老年痴呆症数据库(Alzheimer's disease neuroimaging initiative,简称 ADNI)上的 MRI、PET 和 CSF 多模态特征数据,并与一些相关方法进行比较.实验结果表明,WDL-LDM 方法能够有效地提高分类性能.

1 方法概述

为了获得泛化性较好且分类精度较高的分类模型,本文根据 Lasso 特征选择模型以及大间隔分布学习模型,提出了基于权值分布稀疏特征学习的早期阿尔茨海默病诊断框架.图 1 给出了该学习框架用于早期阿尔茨海默病诊断流程图.具体来说,使用该学习框架进行早期阿尔茨海默病诊断包括以下 3 个步骤.

(1) 首先,分别对采集的 MRI 和 PET 脑图像进行预处理,其中,预处理方法采用文献[12]中的方法.通过预处理,分别从 MRI 和 PET 图像上提取了 93 维感兴趣区域体积(volume)特征;

(2) 将提取的 MRI、PET 特征以及 CSF 检测值串联成 183 维的多模态特征矩阵 X ,再结合类别标记向量 y ,设计出基于权值分布的 Lasso 特征选择模型,用于从多模态特征中寻找最优特征子集以解决高维小样本学习问题;

(3) 将第(2)步求解的最优特征子集训练数据以及类别标记向量 y 输入到大间隔分布分类器.

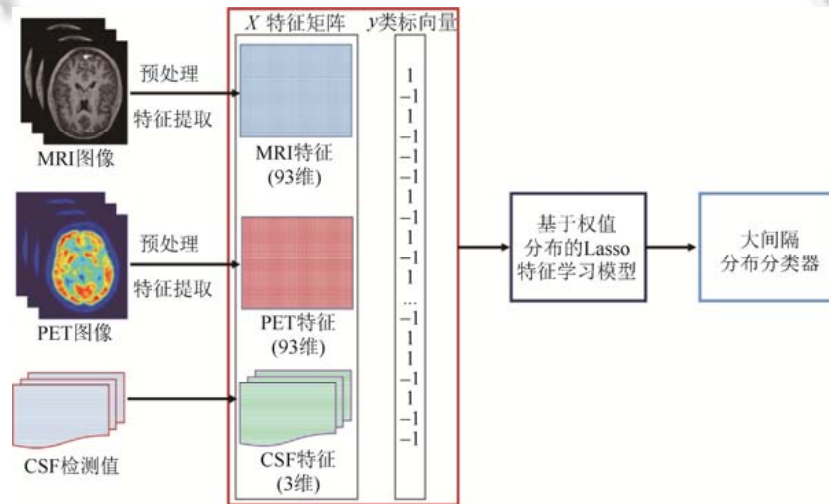


Fig.1 The system diagram of weight-distribution based sparse feature learning framework for early diagnosis of AD

图 1 基于权值分布稀疏特征学习的早期阿尔茨海默病诊断框架示意图

2 基于权值分布稀疏特征学习的早期 AD 诊断方法

2.1 脑图像预处理

对来自 ADNI 数据库上的 MRI 和 PET 脑图像采用文献[12]中的预处理方法进行预处理.具体过程为:首先采用 N3 算法^[22]对所有图像进行前连合与后连合校正,再使用 BSE(brain extraction tool)工具包^[23]将颅骨剥离,去除小脑.然后使用 FSL 工具^[24]把 MRI 图像分割成 3 个不同的组织:灰质、白质、脑脊髓液.然后使用 HAMMER 算法^[25]对所有 MRI 和 PET 图像进行配准,并标注出所有图像的 93 个感兴趣区域.接着,再计算每个被试者 MRI 图像的 93 个感兴趣区域的灰质体积(volume)值.而对 PET 图像计算各个感兴趣区域的图像平均强度值.经过预处理后得到的特征是:来自 MRI 图像的 93 个特征、来自 PET 图像的 93 个特征和来自非图像数据 CSF 的 3 个特征(分别是 CSF A β 42、CSF t -tau 和 CSF p -tau 水平检测值).

2.2 基于权值分布的 Lasso 特征学习模型

在基于机器学习的多模态阿尔茨海默病诊断领域里,通常面临着高维小样本学习问题.为了克服高维小样本容易导致的“过学习”问题,本文采用权值 L_1 范数来限制模型复杂性以获得稀疏特征子集.另外,为了充分利用多模态脑图像与生物标志物特征之间对 AD 诊断的互补性,本文通过获取多模态特征的数据分布信息来加以实现.大间隔分布学习通过增加“间隔分布”正则化项,深入挖掘出支持向量机中间隔分布方面先验信息,可有效降低支持向量机的泛化误差.而传统 Lasso 模型中的权值与特征存在一一对应关系,获取多模态特征的数据分布信息,可以通过获取权值分布信息来实现.综合以上动机,本文提出基于权值分布的 Lasso 特征学习模型 WDL.实际上,WDL 就是在传统 Lasso 模型中同时增加最小化权值均值与最大化权值方差正则化项.由于在本文中大间隔分布学习一方面作为引出基于权值分布的 Lasso 特征学习模型的铺垫,另一方面也直接作为分类的分类器,所以在第 2.2.1 节中将加以介绍.

2.2.1 大间隔分布学习机

在基于神经图像的阿尔茨海默病辅助诊断领域里,大量研究采用支持向量机(SVM)作为分类器,且能获得较好的分类性能,其核心思想是通过最大化“最小间隔”来优化模型复杂度,以获得泛化误差较小的分类模型.实际上,间隔理论不仅在支持向量机中起着重要作用,还可以用于解释其他优秀学习算法的推广性能,例如 boosting 集成学习^[18].此外,根据间隔学习原理,近年来发展出代价敏感大间隔分布学习^[20]与大间隔分布学习^[19].其中的大间隔分布学习机(LDM)^[19]是近年来发展起来的学习模型,该模型认为优化“间隔分布”比优化“最小间隔”更加重要,提出同时增加最大化间隔均值并最小化间隔方差正则化项,以降低分类器的泛化误差.通过引入间隔分布正则化项,研究结果显示,LDM 能够获得优于 SVM 的分类性能与推广性能^[19].

给定一组训练样本集 $S=\{(x_1,y_1),(x_2,y_2),\dots,(x_n,y_n)\}$,其中, n 为训练样本的数量.为了解决线性不可分的问题,支持向量机 SVM 需要将原始特征空间的任意一个训练样本 x_i 通过 ϕ 映射到高维空间.此外,支持向量机是传统的大间隔学习方法,核心思想是最大化“最小间隔”以降低分类器的泛化误差.其中,支持向量机 $y=\omega^T\phi(x)$ 中样本 (x_i,y_i) 的间隔定义为^[26]

$$\gamma_i = y_i \omega^T \phi(x_i), \forall i = 1, 2, \dots, n \quad (1)$$

其中,最小间隔是所有训练样本中样本间隔的最小值.

大间隔分布学习机 LDM 通过同时最大化间隔均值并最小化间隔方差来优化间隔分布,以进一步降低分类器的泛化误差.那么最能表示间隔分布的统计量就是间隔均值和间隔方差了.为了方便公式表达和推导,设 $\mathbf{X}=[\phi(x_1),\phi(x_2),\dots,\phi(x_n)]$, $\mathbf{y}=[y_1,y_2,\dots,y_n]^T$.这样,根据式(1)则可以计算得到间隔均值 $\bar{\gamma}$ 与方差 $\hat{\gamma}$ 分别为

$$\left. \begin{aligned} \bar{\gamma} &= \frac{1}{n} \sum_{i=1}^n y_i \omega^T \phi(x_i) = \frac{1}{n} (\mathbf{X}\mathbf{y})^T \omega, \\ \hat{\gamma} &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (y_i \omega^T \phi(x_i) - y_j \omega^T \phi(x_j))^2 = \frac{2}{n^2} (n\omega^T \mathbf{X}\mathbf{X}^T \omega - \omega^T \mathbf{X}\mathbf{y}\mathbf{y}^T \mathbf{X}^T \omega) \end{aligned} \right\} \quad (2)$$

实际上,LDM就是在原SVM优化问题基础之上增加了间隔均值 $\bar{\gamma}$ 和间隔方差 $\hat{\gamma}$ 正则化项,其形式化目标函数^[19]为

$$\left. \begin{aligned} \min_{\omega, \xi} & \frac{1}{2} \omega^T \omega + \beta_1 \hat{\gamma} - \beta_2 \bar{\gamma} + C \sum_{i=1}^N \xi_i, \\ \text{s.t.} & y_i \omega^T \phi(x_i) \geq 1 - \xi_i, \\ & \xi_i \geq 0, i = 1, \dots, n \end{aligned} \right\} \quad (3)$$

其中, $\beta_1 \geq 0$ 和 $\beta_2 \geq 0$ 分别是用于调节模型复杂度与间隔方差和间隔均值的正则化参数,惩罚参数 $C > 0$ 用于调节分类模型的误差.式(3)的优化问题通过文献[19]提出的平均随机梯度下降(average stochastic gradient descent,简称ASGD)算法进行求解.

2.2.2 基于权值分布的Lasso特征选择模型

为了从串联的183维多模态特征向量中寻找泛化性能更好的特征子集,并获取特征分布信息以保持多模态特征之间的互补性,受大间隔分布学习启发,本文提出了基于权值分布的Lasso特征选择模型.与传统Lasso特征选择模型不同的是,基于权值分布的Lasso模型增加了权值分布正则化项,亦即权值均值和权值方差.与大间隔分布学习不同的是,一方面通过最小化权值均值使优化的权值向量更稀疏,另一方面最大化权值方差可以使判别性好的特征具有较大权重,从而更有利于从高维小样本上学习最优特征子集.

给定一组训练样本集 $\{(x_i, y_i)\}_{i=1}^n$,设 $X = \{x_i\}_{i=1}^n$, $y = \{y_i\}_{i=1}^n$,向量 $x_i \in R^f$ 为训练集 X 中任意一个样本, $y_i \in \{+1, -1\}$ 为训练样本对应的类别标记,其中 f 表示训练样本特征维数, n 表示训练集样本的数量.另外, $w \in R^f$ 表示要优化的权值向量,其任意一个分量 w^j 表示第 j 特征对应的权值系数.那么,基于权值分布的Lasso(weight-distribution based Lasso,简称WDL)模型的目标函数 $J(w)$ 为

$$\min_w \frac{1}{2} \|Xw - y\|_2^2 + \lambda_1 \|w\|_1 + \lambda_2 \bar{w} - \lambda_3 \hat{w} \quad (4)$$

式(4)中, $\|Xw - y\|_2^2$ 为最小平方损失函数,其作用是调整权值向量 w ,使训练集上学到的类别标记向量 $y'(y'=Xw)$ 更逼近真实类别标记向量 y ,亦即最小平方误差最小化; $\|w\|_1$ 为权值向量 w 的 L_1 范数,其作用是使权值向量 w 中部分分量 w^j 的值为0,从而用于筛选最优特征子集; \bar{w} 为权值向量 w 的均值, \hat{w} 为权值向量 w 的方差,权值向量 w 的均值和方差合称为权值分布正则化项,其作用是保持最优特征具有较大权重,排除无关的特征以及判别性较差的特征具有较小的权重,也能实现最优特征的排序; $\lambda_1 \lambda_2 \lambda_3 > 0$ 为正则化参数,用以调节式(4)中各项的贡献度.

对式(4)目标函数的优化问题,可以采用近端梯度下降(PGD)法^[27,28]进行求解.具体来说,式(4)的目标函数 $J(w)$ 可以分成两个部分:平滑可求导部分 $G(w)$ 和不可求导部分 $H(w)$,它们分别为

$$G(w) = \frac{1}{2} \|Xw - y\|_2^2 + \lambda_2 \bar{w} - \lambda_3 \hat{w} \quad (5)$$

$$H(w) = \lambda_1 \|w\|_1 \quad (6)$$

则式(4)的目标函数 $J(w)$ 又可表示为

$$J(w) = \min_w G(w) + \lambda_1 \|w\|_1 \quad (7)$$

由于 $H(w)$ 不可导,因此不能采用传统梯度下降法进行求解,但是根据文献[27,28],可以采用式(4)目标函数的近邻梯度来求解.由于 $G(w)$ 可导,若 $G(w)$ 梯度函数 $\nabla G(w)$ 满足Lipschitz条件,即存在常数 $L > 0$,又根据泰勒公式对 $G(w)$ 展开,使得式(4)目标函数 $J(w)$ 在 w_k 上基于近邻梯度迭代的求解等式为

$$w_{k+1} = \arg \min_w \frac{1}{L} \left\| w - \left(w_k - \frac{1}{L} \nabla G(w_k) \right) \right\|_2^2 + \lambda_1 \|w\|_1 \quad (8)$$

对式(8),可先计算 $t = w_k - \frac{1}{L} \nabla G(w_k)$,然后求解:

$$w_{k+1} = \arg \min_w \frac{1}{L} \|w - t\|_2^2 + \lambda_1 \|w\|_1 \quad (9)$$

令 w^j 表示 w 的第 j 个分量,其中, w 的各分量互不影响,于是式(9)有闭式解:

$$w_{k+1}^j = \begin{cases} t^j - \lambda_1 / L, & \lambda_1 / L < t^j \\ 0, & |t^j| \leq \lambda_1 / L \\ t^j + \lambda_1 / L, & t^j < -\lambda_1 / L \end{cases} \quad (10)$$

其中, w_{k+1}^j 与 t^j 分别表示 w_{k+1} 与 t 的第 j 个分量.因此,通过近端梯度下降法可以快速求解基于权值分布的 Lasso 特征选择模型.

3 实验分析

3.1 ADNI数据库

本文实验数据均来自于国际阿尔茨海默病(ADNI)数据库(<http://adni.loni.usc.edu/>).ADNI 由国际老年研究所(National Institute on Aging,简称 NIA)、生物医学成像和生物工程研究所(National Institute of Biomedical Imaging and Bioengineering,简称 NIBIB)、美国食品和药物管理局(Food and Drug Administration,简称 FDA)、民营医药企业和非营利组织在 2003 年启动建立,其主要目标是测试能否通过组合 MRI、PET、其他生物标志物以及临床神经心理学评估来测定 MCI 和早期 AD 的进展(progression).该数据库包含的数据模态有:基于时间序列的 MRI 图像数据、PET 图像数据以及其他类型生物标记值,例如:CSF,还包括一些临床的神经心理学评估评分值,例如:简易精神状态检查量表(mini-mental state examination,简称 MMSE)和老年痴呆症评定量表(Alzheimer's disease assessment scale-cognitive,简称 ADAS-Cog).这些数据类别主要为:早期阿尔茨海默病患者(AD)、轻度认知障碍患者(MCI)和认知正常对照组(NC).ADNI 数据库的被试者从横跨美国与加拿大 50 个网站招募而来.他们的最初目标是招募 800 个成年志愿者,年龄段是 55 岁~90 岁;其中有 200 人是连续 3 年跟踪测试认知正常的老人,有 400 人是连续 3 年跟踪测试的轻度认知障碍患者,还有 200 人是连续 2 年跟踪测试的阿尔茨海默病患者.这些被试者的个人基本信息可从 ADNI 官网上获得.

本文同时选择了具有 MRI、PET 和 CSF 模态的被试者样本数据进行实验,并只选择这些被试者的基准时间点采集的数据.在国际阿尔茨海默病数据库中,同时具有以上 3 个模态数据的被试者为 202 个,更多详情请参考文献[12].表 1 列出这些被试者的人口统计资料信息.

Table 1 Subject information (mean±std)

表 1 被试者样本信息(均值±标准差)

类别	被试者个数	年龄	受教育年限	MMSE	ADAS-Cog
AD	51	75.2±7.4	14.7±3.6	23.8±2.0	18.3±6.0
NC	52	75.3±5.2	15.8±3.2	29.0±1.2	7.4±3.2
MCI-C	43	75.8±6.8	16.1±2.6	26.6±1.7	12.9±3.9
MCI-NC	56	74.7±7.7	16.1±3.0	27.5±1.5	10.2±4.3

3.2 实验设置

为了测试分布学习模型性能,本文采用 10 折交叉验证策略来评价算法的分类性能.具体来说,将样本集平均划分成 10 份,逐一地选择其中 1 份作为测试集,剩余 9 份作为训练集.计算这 10 次实验的平均精度(ACCuracy)、敏感度(SENsitivity)、特异度(SPEcificity)、受试者工作特征(receiver operating characteristic,简称 ROC)曲线下面积(area under curve,简称 AUC)值,以作为 1 次划分的实验结果.然后随机地交换样本排列顺序,再进行 1 次 10 折交叉验证的划分,并计算平均精度、敏感度、特异度、AUC 值.重复 10 次划分,并计算这 10 次划分的平均精度 ACC、敏感度 SEN、特异度 SPE 和 AUC 值.

另外,传统支持向量机 SVM 以及其他需要采用支持向量机分类的比较方法,均采用 LIBSVM(a Library for Support Vector Machine,简称 LIBSVM)工具箱^[29]仿真实现.其中,所有使用 SVM 分类的方法都采用线性核,其他

参数均采用系统缺省值.在基于权值分布的 Lasso 特征选择步骤中,正则化参数 $\lambda_1, \lambda_2, \lambda_3$ 通过网格搜索方法进行优化,其搜索范围为 0~1,步长为 0.01.在采用大间隔分布学习机进行分类的步骤中,采用文献[19]中提出的大间隔分布学习机工具包(http://lamda.nju.edu.cn/code_LDM.ashx),参数 β_1, β_2, C 也通过网格搜索方法进行优化,其搜索范围为 0~100,步长为 1.Lasso 特征选择模型采用 Zhou 等人提出的 MALSAR 工具包(<http://www.yelab.net/software/MALSAR/>),正则化参数同样也通过网格搜索方法进行优化,其搜索范围为 0~1,步长为 0.01.值得注意的是,本文所有方法的参数优化都是在训练集上通过内嵌的 10 折交叉验证优化得到的.此外,采用文献[12]中的归一化方法进行数据归一化.

3.3 实验结果

3.3.1 不同方法分类性能比较

为了验证本文提出的分布学习模型(WDL-LDM)用于早期阿尔茨海默病诊断的性能,分别和与之相关性较强的 4 种分类方法进行比较.这 4 种分类方法分别是:传统支持向量机(SVM)、大间隔分布分类机(LDM)、基于 t-检验(t-test)的特征选择方法结合支持向量机分类器以及基于 Lasso 特征选择方法结合支持向量机分类器.表 2 给出了 WDL-LDM 方法与其他 4 种方法在串联的多模态数据上分别分类 AD/NC、MCI/NC、pMCI/sMCI 的性能比较实验结果.在表 2 中,SVM 表示只采用 SVM 分类模型进行分类实验;LDM 表示只采用 LDM 分类模型进行分类实验;t-test 表示先采用 t-检验进行特征选择,再使用 SVM 分类模型进行分类实验;Lasso 表示先采用 Lasso 模型进行特征选择,再使用 SVM 分类模型进行分类实验;SLasso-LDM 表示先采用文献[30,31]提出的 Lasso-path 方法进行特征选择,再使用 LDM 分类模型进行分类实验.其中,表 2 中列出的所有实验结果都是通过 10 次 10 折交叉验证划分的平均值.此外,为了进一步比较各种分类方法的性能,图 2 绘制了表 2 中所有方法对应的 ROC 曲线.

Table 2 For three binary classification problems (AD/NC, MCI/NC, and pMCI/sMCI), our proposed method (WDL-LDM) compare with other state-of-the-art methods (SVM, Lasso, LDM, SLasso-LDM and t-test) using multimodal data

表 2 不同方法在多模态数据上分别分类 AD/NC、MCI/NC、pMCI/sMCI 的性能比较

AD vs. NC				
方法	精度(%)	敏感度(%)	特异度(%)	AUC
SVM	90.3	90.1	90.5	0.963
LDM	90.9	91.4	90.3	0.959
t-test	90.7	90.5	90.9	0.963
Lasso	90.5	90.3	90.7	0.963
SLasso-LDM	94.6	92.7	96.3	0.989
WDL-LDM	97.5	99.1	95.9	0.996
MCI vs. NC				
方法	精度(%)	敏感度(%)	特异度(%)	AUC
SVM	76.9	82.3	66.8	0.814
LDM	72.0	84.2	65.3	0.799
t-test	77.4	82.6	67.5	0.817
Lasso	80.5	85.0	71.9	0.881
SLasso-LDM	81.9	94.6	74.5	0.919
WDL-LDM	83.1	95.3	76.4	0.924
pMCI vs. sMCI				
方法	精度(%)	敏感度(%)	特异度(%)	AUC
SVM	62.3	57.0	66.4	0.659
LDM	68.1	72.8	64.5	0.661
t-test	63.7	58.5	67.6	0.671
Lasso	74.1	70.5	76.9	0.809
SLasso-LDM	79.9	82.1	78.1	0.884
WDL-LDM	84.8	87.2	82.9	0.925

表 2 和图 2 的实验结果表明,采用特征选择步骤能够改进分类模型诊断早期阿尔茨海默病的性能.其中,在 AD/NC 分类实验中,采用本文提出的基于权值分布的 Lasso(WDL)特征选择方法能够明显提高仅采用 LDM 分

类器的分类性能.然而,相比于仅采用 SVM 分类器,采用 t-test 和 Lasso 特征选择方法只是轻微提高了分类精度、敏感度和特异度,而 AUC 值并未改变,这表明,采用以上两种特征选择方法不能有效改进分类模型的性能.其次,在 MCI/NC 分类实验中,采用特征选择步骤能够较明显地改进分类模型的性能.其中,采用本文提出的 WDL 特征选择方法能够显著提高 LDM 分类器的分类性能;采用 t-test 和 Lasso 特征选择方法也能改进一些 SVM 分类器的分类性能;而仅采用 LDM 分类器的性能不如此采用 SVM 分类器的性能.最后,在 pMCI/sMCI 分类实验中,采用特征选择步骤能够显著改进分类模型的性能.其中,采用 t-test 特征选择方法只能轻微提高分类性能,然而分别采用 WDL 和 Lasso 特征选择方法能够显著改进 LDM 与 SVM 分类器的分类性能;尤其是 WDL 特征选择方法能够十分显著地改进 LDM 分类器性能;此外,仅采用 LDM 分类器的性能也明显优于仅采用 SVM 分类器的性能.同时,为了进一步验证 WDL-LDM 的性能,实验中也采用文献[30,31]提出的 Lasso-path 方法来替换 WDL-LDM 中的 WDL 特征选择方法,亦即表 2 中的 SLasso-LDM 方法;实验结果显示,在 3 组分类实验中,WDL 仍优于 Lasso-path 方法.综合以上表 2 和图 2 的实验分析结果可以得出以下结论:本文提出的 WDL 特征选择方法能够有效地改进分类器性能,其中,对 pMCI/sMCI 分类性能的改进尤其显著,并结合 LDM 分类器能够完成较高的分类性能,因此本文提出的 WDL-LDM 方法能够很好地应用于早期阿尔茨海默病的诊断.

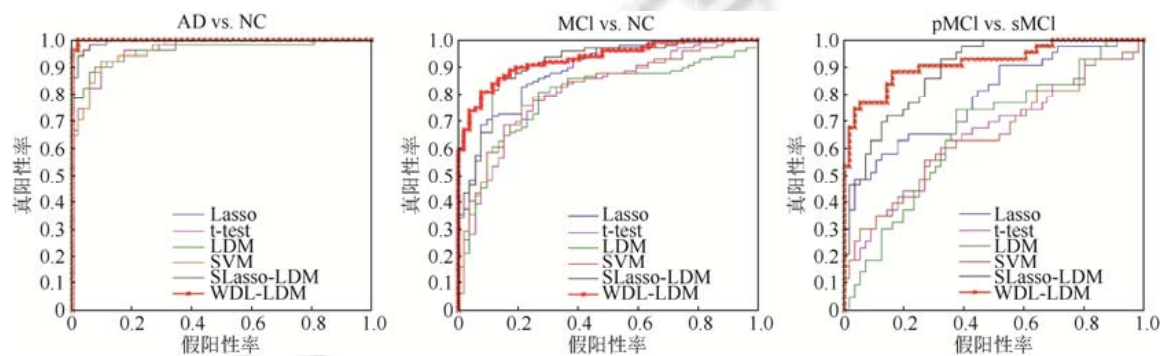


Fig.2 ROC curves of different methods for three binary classification problems (AD/NC, MCI/NC, and pMCI/sMCI)

图 2 不同方法在多模态数据上分别分类 AD/NC、MCI/NC、pMCI/sMCI 的 ROC 曲线

3.3.2 多模态与单模态数据上的分类性能比较

为了验证本文提出的分布学习模型(WDL-LDM)在多模态与单模态数据上的分类性能,分别将多模态特征数据(MRI+PET+CSF,简称 MPC)和单模态特征数据(MRI、PET、CSF)输入到 WDL-LDM.表 3 就是 WDL-LDM 方法分别运行在多模态特征和单模态特征数据上分类 AD/NC、MCI/NC、pMCI/sMCI 的实验结果.为了更直观、更全面地比较这些方法的差异性,图 3 绘制了表 3 中所有方法对应的 ROC 曲线.表 3 和图 3 的实验结果表明,采用多模态特征数据能够获得更好的分类性能,这是因为,这些多模态特征数据之间存在互补性的诊断信息.这也说明,WDL-LDM 方法能够有效融合多模态特征数据.在 3 组都采用单模态特征数据的分类实验中,采用 MRI 模态特征能够获得最好的分类性能,这表明,早期阿尔茨海默病患者在脑组织的结构形态上逐渐发生了萎缩.此外,除了使用 MRI,采用 PET 或 CSF 模态特征也能获得不错的诊断性能,这说明,早期阿尔茨海默病患者不仅在脑组织结构上出现了变化,也在脑组织功能上出现了变化.然而,单独使用某一种模态特征都无法获得同时使用这些多模态特征的性能,这也再次验证了 WDL-LDM 方法能从多模态特征中找到那些具有互补性的特征子集,从而避免多模态特征易引起的过学习问题,达到有效提高分类模型性能的目的.

Table 3 Comparison of performances that the WDL-LDM method performs on multimodal and single modality data

表 3 分布学习模型分别在多模态数据与单模态上分类性能的比较

AD vs. NC				
模态	精度(%)	敏感度(%)	特异度(%)	AUC
MRI	93.5	95.9	90.9	0.969
PET	91.4	95.9	86.8	0.968
CSF	83.7	84.4	83.2	0.866
MPC	97.5	99.1	95.9	0.996
MCI vs. NC				
模态	精度(%)	敏感度(%)	特异度(%)	AUC
MRI	77.4	91.3	69.9	0.891
PET	69.8	88.1	59.9	0.869
CSF	70.9	80.8	65.4	0.753
MPC	83.1	95.3	76.4	0.924
pMCI vs. sMCI				
模态	精度(%)	敏感度(%)	特异度(%)	AUC
MRI	75.4	79.9	71.9	0.856
PET	72.0	80.5	65.4	0.826
CSF	63.1	69.3	58.4	0.646
MPC	84.8	87.2	82.9	0.925

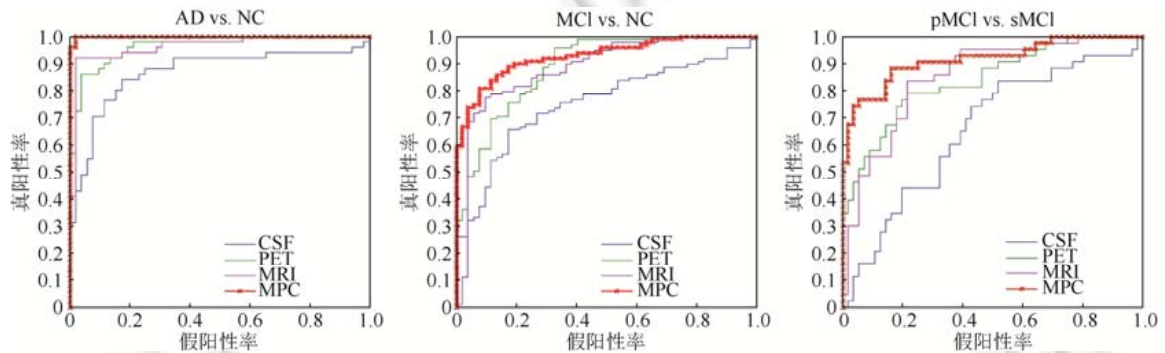


Fig.3 ROC curves of the WDL-LDM method with multi-modality and single-modality data, respectively

图 3 WDL-LDM 分别在多模态数据与单模态上分类 AD/NC、MCI/NC、pMCI/sMCI 的 ROC 曲线

3.3.3 判别性脑区检测

为了验证本文提出的分布学习模型(WDL-LDM)用于特征选择与排序的有效性,表 4~表 6 列出了在 3 组分类实验中 WDL 方法选出的稳定特征子集,然后根据选出特征对应的平均权值进行了排序,最后为了验证这个排序的准确性,表中列出了这些特征在不同分类实验中的 P 值.表 4~表 6 中稳定特征就是在 10 次 10 折交叉验证实验中都曾被 WDL 方法选中的特征,这些特征具有很好的泛化性和判别性.类似地,平均权值也就是 10 次 10 折交叉验证实验权值的均值.本文提出的 WDL 特征选择方法与传统 Lasso 最大的区别就是可利用权值分布信息进行多模态特征选择,这归因于最小化权值向量均值与 L_1 范数保证特征的稀疏性,同时最大化权值向量方差保持最优特征子集的有序性.表 4~表 6 中根据权值进行的稳定特征排序与 P 值排序保持一致,这证实了 WDL 方法进行最优特征子集排序的有效性.此外,在表 4~表 6 中,从 MRI 模态上选出的稳定脑区特征均多于 PET 和 CSF 模态,这也表明,单独 MRI 模态在早期阿尔茨海默病的诊断性能优于单独采用 PET 或 CSF 模态.在表 4~表 6 的 3 组分类实验中,选出的稳定特征子集均包含了 3 个模态的特征,这验证了多模态脑图像与生物标志物特征对 AD 诊断具有互补性.表 4~表 6 中还有一个值得注意的现象,就是除了那些 P 值小于 0.05 的特征被选出以外,还有一些 P 值大于 0.05 的特征也被选中,这表明,t-test 方法采用 P 值进行特征选择的局限性,也解释了表 2 中 WDL-LDM 方法分类性能明显优于 t-test 方法的原因.WDL 方法选出的稳定脑区在其他相关文献中都有报道^[4,12-14,32,33],例如:hippocampal、amygdala、temporal lobe、precuneus、insula 等等,这些脑区在医学上被证实

与早期阿尔茨海默病的诊断密切相关.

Table 4 Selected stable features ranking using the WDL-LDM method on AD/NC classification

表 4 在 AD/NC 分类实验中 WDL-LDM 选出的稳定特征排序

模态	特征/脑区名称	P 值
MRI	Hippocampal formation right	0.124×10^{-10}
MRI	Hippocampal formation left	0.208×10^{-10}
PET	Precuneus left	0.985×10^{-10}
CSF	t-tau	0.108×10^{-9}
MRI	Inferior temporal gyrus right	0.277×10^{-9}
MRI	Uncus right	0.378×10^{-9}
CSF	A β 42	0.179×10^{-7}
MRI	Lateral occipitotemporal gyrus left	0.141×10^{-6}
PET	Anterior limb of internal capsule right	0.106×10^{-5}
MRI	Angular gyrus right	0.491×10^{-4}
PET	Entorhinal cortex right	0.000 19
PET	Amygdala left	0.000 90
PET	Postcentral gyrus left	0.000 94
MRI	Inferior occipital gyrus left	0.001 19
PET	Putamen right	0.001 31
MRI	Parahippocampal gyrus left	0.005 48
PET	Parahippocampal gyrus right	0.020 13
PET	Cingulate region right	0.022 64
PET	Superior temporal gyrus left	0.059 42
MRI	Supramarginal gyrus right	0.161 12
MRI	Insula right	0.208 78
PET	Medial frontal gyrus left	0.322 78
MRI	Anterior limb of internal capsule right	0.349 32
MRI	Medial front-orbital gyrus right	0.600 74
MRI	Fornix right	0.683 65
MRI	Superior frontal gyrus right	0.869 45
MRI	Lingual gyrus left	0.874 95
MRI	Fornix left	0.920 06

Table 5 Selected stable features ranking using the WDL-LDM method on MCI/NC classification

表 5 在 MCI/NC 分类实验中 WDL-LDM 选出的稳定特征排序

模态	特征/脑区名称	P 值
MRI	Hippocampal formation left	0.000 000 5
CSF	t-tau	0.000 001 9
CSF	A β 42	0.000 014 9
MRI	Parahippocampal gyrus left	0.000 2
PET	Precuneus right	0.000 4
MRI	Cuneus left	0.000 6
MRI	Temporal pole right	0.001 9
PET	Middle frontal gyrus left	0.007 2
PET	Entorhinal cortex right	0.015 0
MRI	Anterior limb of internal capsule left	0.018 8
MRI	Thalamus right	0.024 9
MRI	Posterior limb of internal capsule inc. cerebral peduncle left	0.040 6
PET	Superior parietal lobule right	0.082 5
PET	Caudate nucleus right	0.204 1
MRI	Middle occipital gyrus right	0.233 5
MRI	Occipital pole right	0.286 1
PET	Inferior occipital gyrus left	0.452 3
PET	Temporal pole left	0.632 7

3.3.4 正则化参数对分类性能的影响

表 2 的实验结果表明,增加 WDL 特征选择步骤能够明显改进 LDM 的分类性能,这也验证了本文提出 WDL 特征选择方法的有效性.在 WDL 特征选择模型中,通过添加权重分布正则化项到传统 Lasso 模型中以提高模型泛化性能,这是由于权重分布对应于特征分布,然而传统 Lasso 模型没有考虑特征分布信息,所以本文提出在传统 Lasso 模型引入权重分布先验信息,亦即最小化权值均值并最大化权值方差.为了验证这个假设的有效性,在 WDL-LDM 模型中通过对 WDL 进行不同正则化参数设置来探讨分类性能的变化,以此来验证引入权重分布先验信息的有效性,表 7 给出了 WDL 特征选择方法采用不同正则化参数设置时的 WDL-LDM 分类性能.

Table 6 Selected stable features ranking using the WDL-LDM method on pMCI/sMCI classification**表 6** 在 pMCI/sMCI 分类实验中 WDL-LDM 选出的稳定特征排序

模态	特征/脑区名称	P 值
MRI	Perirhinal cortex left	0.004 7
PET	Superior frontal gyrus right	0.006 5
PET	Medial frontal gyrus right	0.009 8
CSF	t-tau	0.014 4
CSF	p-tau	0.014 8
PET	Lingual gyrus left	0.026 5
PET	Hippocampal formation right	0.028 1
MRI	Anterior limb of internal capsule left	0.040 6
MRI	Inferior frontal gyrus right	0.046 9
PET	Entorhinal cortex right	0.084 6
MRI	Temporal pole left	0.086 8
MRI	Inferior frontal gyrus left	0.091 1
MRI	Angular gyrus left	0.135 2
MRI	Fornix left	0.146 0
MRI	Superior occipital gyrus left	0.208 9
PET	Cuneus left	0.286 6
MRI	Precentral gyrus left	0.413 1
MRI	Postcentral gyrus left	0.552 9
PET	Temporal lobe WM left	0.625 0
MRI	Occipital pole left	0.687 9
PET	Amygdala left	0.761 0
MRI	Inferior occipital gyrus left	0.761 7
PET	Caudate nucleus left	0.781 8
MRI	Uncus right	0.785 2
MRI	Supramarginal gyrus right	0.990 8

Table 7 Classification performance comparison of our proposed method (WDL-LDM) using different setting of regularization parameters on the step of WDL feature selection**表 7** WDL 特征选择方法采用不同正则化参数设置的 WDL-LDM 分类性能比较

正则化参数	AD vs. NC		MCI vs. NC		pMCI vs. sMCI	
	精度(%)	AUC	精度(%)	AUC	精度(%)	AUC
$\lambda_2, \lambda_3=0$	93.3	0.991	80.7	0.913	78.3	0.870
$\lambda_2=0$	94.5	0.992	81.6	0.918	81.8	0.912
$\lambda_3=0$	92.8	0.986	80.8	0.916	79.1	0.887
$\lambda_1, \lambda_2, \lambda_3>0$	97.5	0.996	83.1	0.924	84.8	0.925

在表 7 中,正则化参数设置为“ $\lambda_2, \lambda_3=0$ ”,表示没有权重分布正则化项,此时,WDL 特征选择模型就是传统的 Lasso;若设置为“ $\lambda_2=0$ ”,则表示没有最小化权值均值正则化项;若设置为“ $\lambda_3=0$ ”,则表示没有最大化权值方差正则化项;另外,若设置为“ $\lambda_1, \lambda_2, \lambda_3>0$ ”,则表示同时存在权重 L_1 范数和权重分布正则化项.表 7 的实验结果表明,在 3 组分类实验中同时采用权重 L_1 范数和权重分布正则化项的 WDL-LDM 分类性能都优于其他 3 种正则化参数设置方案.其中,不采用权重分布正则化项,WDL-LDM 分类性能明显不如“ $\lambda_1, \lambda_2, \lambda_3>0$ ”参数设置方案;单独采用最小化权值均值或者最大化权值方差正则化项,WDL-LDM 分类性能也明显差于“ $\lambda_1, \lambda_2, \lambda_3>0$ ”参数设置方案.综合以上实验结论表明,在传统 Lasso 模型中同时加入最小化权值均值和最大化权值方差正则化项能够有效改进分类模型的性能.

4 结束语

为了获得泛化性较好且分类精度较高的分类模型,本文根据 Lasso 特征选择模型以及大间隔分布学习模型,提出了基于权值分布稀疏特征学习的早期阿尔茨海默病诊断框架.具体来说,该诊断框架包含 3 个部分:首先对采集的 MRI 和 PET 脑图像进行一系列预处理并提取特征;然后根据传统 Lasso 模型提出基于权值分布的 Lasso 特征选择模型(WDL),可用于从多模态特征中寻找性能更好的特征子集;最后采用大间隔分布分类机(LDM)训练分类模型.在国际老年痴呆症数据库(ADNI)上,采用 MRI、PET 和 CSF 多模态数据进行实验验证,并与一些相关的方法进行比较.实验结果表明,WDL-LDM 方法能够有效提高分类性能,并能有效利用多模态特

征分布信息。

尽管本文取得了较好的结果,但仍有改进空间。一方面,本文只针对数量有限的标记训练数据设计了监督学习算法,而临床上存在许多未标记的多模态数据还可利用,考虑将当前的模型扩展为半监督学习方式是未来的发展方向之一。另一方面,临床上也存在大量不完全的多模态数据,充分利用这些不完全的多模态标记数据,不但可以增加训练样本数量,针对不完全的多模态数据设计学习算法,还可以提高模型推广性能。

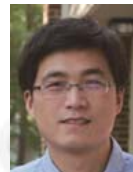
References:

- [1] Hu K, Wang Y, Chen K, Hou L, Zhang X. Multi-scale features extraction from baseline structure MRI for MCI patient classification and AD early diagnosis. *Neurocomputing*, 2016,175(1):132–145. [doi: 10.1016/j.neucom.2015.10.043]
- [2] Wachinger C, Reuter M. Domain adaptation for Alzheimer’s disease diagnostics. *NeuroImage*, 2016,139(1):470–479. [doi: 10.1016/j.neuroimage.2016.05.053]
- [3] Liu M, Zhang D, Shen D. Relationship induced multi-template learning for diagnosis of Alzheimer’s disease and mild cognitive impairment. *IEEE Trans. on Medical Imaging*, 2016,35(6):1463–1474. [doi: 10.1109/TMI.2016.2515021]
- [4] Jie B, Zhang D, Cheng B, Shen D. Manifold regularized multitask feature learning for multimodality disease classification. *Human Brain Mapping*, 2015,36(2):489–507. [doi: 10.1002/hbm.22642]
- [5] Cheng B, Liu M, Zhang D, Munsell BC, Shen D. Domain transfer learning for MCI conversion prediction. *IEEE Trans. on Biomedical Engineering*, 2015,62(7):1805–1817. [doi: 10.1109/TBME.2015.2404809]
- [6] Moradi E, Pepe A, Gaser C, Huttunen H, Tohka J. Machine learning framework for early MRI-based Alzheimer’s conversion prediction in MCI subjects. *NeuroImage*, 2015,104(1):398–412. [doi: 10.1016/j.neuroimage.2014.10.002]
- [7] Liu F, Wee CY, Chen HF, Shen D. Inter-modality relationship constrained multi-modality multi-task feature selection for Alzheimer’s disease and mild cognitive impairment identification. *NeuroImage*, 2014,84(1):466–475. [doi: 10.1016/j.neuroimage.2013.09.015]
- [8] Liu M, Zhang D, Shen D. Inherent structure based multi-view learning with multi-atlas feature representation for Alzheimer’s disease diagnosis. *IEEE Trans. on Biomedical Engineering*, 2016,63(7):1473–1482. [doi: 10.1109/TBME.2015.2496233]
- [9] Zhu X, Suk H, Shen D. A novel matrix-similarity based loss function for joint regression and classification in AD diagnosis. *NeuroImage*, 2014,100(1):91–105. [doi: 10.1016/j.neuroimage.2014.05.078]
- [10] Zhou J, Liu J, Narayan VA, Ye J. Modeling disease progression via multi-task learning. *NeuroImage*, 2013,78(1):233–248. [doi: 10.1016/j.neuroimage.2013.03.073]
- [11] Cheng B, Liu M, Suk H, Shen D, Zhang D. Multimodal manifold-regularized transfer learning for MCI conversion prediction. *Brain Imaging and Behavior*, 2015,9(4):913–926. [doi: 10.1007/s11682-015-9356-x]
- [12] Zhang D, Wang Y, Zhou L, Yuan H, Shen D. Multimodal classification of Alzheimer’s disease and mild cognitive impairment. *NeuroImage*, 2011,55(3):856–867. [doi: 10.1016/j.neuroimage.2011.01.008]
- [13] Ye J, Farnum M, Yang E, Verbeeck R, Lobanov V, Raghavan N. Sparse learning and stability selection for predicting MCI to AD conversion using baseline ADNI data. *BMC Neurology*, 2012,12(1):1–12. [doi: 10.1186/1471-2377-12-46]
- [14] Zhang D, Shen D. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer’s disease. *NeuroImage*, 2012,59(2):895–907. [doi: 10.1016/j.neuroimage.2011.09.069]
- [15] Dukart J, Sambataro F, Bertolino A. Accurate prediction of conversion to Alzheimer’s disease using imaging, genetic, and neuropsychological biomarkers. *Journal of Alzheimer’s Disease*, 2016,49(4):1143–1159. [doi: 10.3233/JAD-150570]
- [16] Zhang D, Shen D. Predicting future clinical changes of MCI patients using longitudinal and multimodal biomarkers. *PLoS One*, 2012,7(3):e33182. [doi: 10.1371/journal.pone.0033182]
- [17] Cheng B, Zhang D, Chen S, Kaufer DI, Shen D. Semi-supervised multimodal relevance vector regression improves cognitive performance estimation from imaging and biological biomarkers. *Neuroinformatics*, 2013,11(3):339–353. [doi: 10.1007/s12021-013-9180-7]
- [18] Gao W, Zhou Z. On the doubt about margin explanation of boosting. *Artificial Intelligence*, 2013,203(1):1–18. [doi: 10.1016/j.artint.2013.07.002]

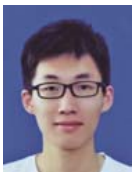
- [19] Zhang T, Zhou Z. Large margin distribution machine. In: Proc. of the 20th ACM SIGKDD Conf. on Knowledge Discovery and Data Mining. 2014. 313–322. [doi: 10.1145/2623330.2623710]
- [20] Zhou Y, Zhou Z. Large margin distribution learning with cost interval and unlabeled data. IEEE Trans. on Knowledge and Data Engineering, 2016,28(7):1749–1763. [doi: 10.1109/TKDE.2016.2535283]
- [21] Geng X. Label distribution learning. IEEE Trans. on Knowledge and Data Engineering, 2016,28(7):1734–1748. [doi: 10.1109/TKDE.2016.2545658]
- [22] Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans. on Medical Imaging, 1998,17(1):87–97. [doi: 10.1109/42.668698]
- [23] Wang Y, Nie J, Yap PT, Shi F, Guo L, Shen D. Robust deformable-surface-based skull-stripping for large-scale studies. In: Proc. of the Conf. on Medical Image Computing and Computer-Assisted Intervention. 2011. 635–642. [doi: 10.1007/978-3-642-23626-6_78]
- [24] Zhang YY, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans. on Medical Imaging, 2001,20(1):45–57. [doi: 10.1109/42.906424]
- [25] Shen D, Davatzikos C. HAMMER: Hierarchical attribute matching mechanism for elastic registration. IEEE Trans. on Medical Imaging, 2002,21(11):1421–1439. [doi: 10.1109/TMI.2002.803111]
- [26] Vapnik VN. An overview of statistical learning theory. IEEE Trans. on Neural Networks, 1999,10(5):988–999. [doi: 10.1109/72.788640]
- [27] Chen X, Pan W, Kwok JT, Carbonell JG. Accelerated gradient method for multi-task sparse learning problem. In: Proc. of the 9th IEEE Int'l Conf. on Data Mining and Knowledge Discovery. 2009. 746–751. [doi: 10.1109/ICDM.2009.128]
- [28] Nemirovski A. Efficient methods in convex programming. 1995. <http://www2.isye.gatech.edu/~nemirovs/>
- [29] Chang CC, Lin CJ. LIBSVM: A library for support vector machines. 2001. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>
- [30] Park MY, Hastie T. L_1 -regularization path algorithm for generalized linear models. Journal of Royal Statistical Society, 2007,69(4): 659–677.
- [31] Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. Journal of Statist Software, 2010,33(1):1–22.
- [32] Davatzikos C, Bhatt P, Shaw LM, Batmanghelich KN, Trojanowski JQ. Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. Neurobiology of Aging, 2011,32(12):19–27. [doi: 10.1016/j.neurobiolaging.2010.05.023]
- [33] Eskildsen SF, Coupé P, García-Lorenzo D, Fonov V, Pruessner JC, Collins DL. Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning. NeuroImage, 2013,65(2):511–521. [doi: 10.1016/j.neuroimage.2012.09.058]



程波(1982—),男,重庆人,博士,副教授,主要研究领域为机器学习,模式识别,医学图像分析.



张道强(1978—),男,博士,教授,博士生导师,CCF 专业会员,主要研究领域为机器学习,模式识别,数据挖掘,医学图像分析.



丁毅(1993—),男,硕士,CCF 学生会会员,主要研究领域为机器学习,强化学习,计算金融.