

基于自适应多任务卷积神经网络的脑网络分类方法

邢新颖 冀俊忠 姚垚垚

(北京工业大学信息学部计算机学院 北京 100124)

(xinying.xing@emails.bjut.edu.cn)

Brain Networks Classification Based on an Adaptive Multi-Task Convolutional Neural Networks

Xing Xinying, Ji Junzhong, and Yao Yao

(College of Computer, Faculty of Information Technology, Beijing University of Technology, Beijing 100124)

Abstract Brain networks classification is an important subject in brain science. In recent years, brain networks classification based on convolutional neural networks has become a hot topic. However, it is still difficult to accurately classify brain network data with high dimension and small sample size. Due to the close relationship between different clinical phenotypes and brain networks of different populations, it is highly possible to provide auxiliary information for the brain networks classification. Therefore, we propose a new brain networks classification method based on an adaptive multi-task convolutional neural network in this paper. Firstly, the clinical phenotype predictions are introduced as different auxiliary tasks and the shared representation mechanism of multi-task convolutional neural networks is used to provide general and useful information for brain networks classification. Then, in order to reduce the experimental cost and the error caused by the manual operation, a new adaptive method is proposed to substitute for manual adjustments of the weight of every task in the multi-task learning. The experimental results on the autism brain imaging data exchange I (ABIDE I) dataset show that the multi-task convolutional neural networks which introduce clinical phenotype predictions can achieve better classification results. Moreover, the adaptive multi-task learning method can further improve the performance of brain networks classification.

Key words brain networks classification; convolutional neural network; multi-task learning; clinical phenotype; adaptive method

摘要 脑网络分类是脑科学发展中的一项重要课题。近年来,基于卷积神经网络的脑网络分类方法已经成为一个前沿热点。然而,目前仍难以对数据维度高、样本量小的脑网络数据进行精准分类。由于不同人群的临床表型与其脑网络差异存在着一定的依存关系,极有可能为脑网络分类提供辅助信息,故提出一种新的基于自适应多任务卷积神经网络的脑网络分类方法。该方法引入临床表型预测作为辅助任务,通过多任务卷积神经网络的共享表示机制来为脑网络分类提供有用信息;同时为了降低实验成本和人工操作带来的误差,提出了一种新的自适应方法来代替人工调整多任务学习中各个子任务的权重。在ABIDE I (autism brain imaging data exchange I) 数据集上的实验结果表明:引入临床表型预测任务的

收稿日期:2019-03-21;修回日期:2019-12-26

基金项目:国家自然科学基金项目(61672065)

This work was supported by the National Natural Science Foundation of China (61672065).

通信作者:冀俊忠(jjz01@bjut.edu.cn)

多任务卷积神经网络能够获得更好的脑网络分类结果,而且自适应多任务学习方法能够进一步提升脑网络的分类性能。

关键词 脑网络分类;卷积神经网络;多任务学习;临床表型;自适应方法

中图法分类号 TP18

脑网络是一种利用脑成像数据从结构连接或功能连接角度定量地刻画人脑系统的方式,包括人脑结构网络或人脑功能网络^[1-2],脑网络可以被表示为由节点和边构成的图(graph)模型。图模型中的节点一般为由多个神经元组成的感兴趣区域(region of interest, ROI)或指某一特定脑区,边通常表示相应ROI或脑区对之间的连接强度。为了便于存储,一般使用邻接矩阵表示脑网络数据,矩阵中的一行或一列分别对应于图模型中的节点。自动挖掘和分析受试者脑网络数据的特征,并推断其是否患有某种精神性脑疾病,这种方法被称为脑网络分类。脑网络分类方法的研究和应用有助于人类更好地理解脑疾病发病机理和脑疾病患者的早期诊断,故而具有重要的研究与应用价值。在与脑网络分类任务相关的机器学习方法中,最小绝对收缩和选择算子(least absolute shrinkage and selection operator, LASSO)方法和支持向量机(support vector machine, SVM)方法的应用最为广泛^[3-5]。但这2种传统机器学习方法属于浅层模型,面对高维、小样本的脑网络数据时,其拟合能力往往存在明显不足。最近,基于深度学习的脑网络分类方法因其出色的特征表达能力,逐渐成为该领域中新的研究热点。目前多数研究集中在全连接神经网络(fully connected neural network, FCNN)方法和卷积神经网络(convolutional neural network, CNN)方法这2类脑网络分类方法。

FCNN方法是最简单但最常用的深度学习方法。2015年Kim等人^[6]对FCNN方法中的神经网络参数进行预训练并施加稀疏性限制,然后应用于精神分裂症的分类任务中,首次验证了FCNN方法在脑网络分类研究中的有效性。2017年Guo等人^[7]在栈式稀疏自编码(stacked sparse autoencoders, SSAE)方法的基础上提出了一种特征选择方法,通过将该方法应用于脑网络分类识别出32条与自闭症(autism spectrum disorder, ASD)相关的异常功能连接边。2018年Zeng等人^[8]为了实现精神分裂症的多站点数据分类,提出了一种深度判别自编码器,该自编码器通过引入稀疏约束项来学习不同站点数据间与疾病相关的共性特征,首次实现了基于

深度学习的多站点精神分裂症分类。2019年Ju等人^[9]将SSAE方法应用于对轻度认知障碍患者和健康受试者进行分类,该方法的分类准确率与传统机器学习方法相比有显著提高。此外该研究进一步分析了脑网络数据和原始rs-fMRI(resting-state functional magnetic resonance imaging)数据在相同方法下所得结果的差异,发现脑网络数据在神经性脑疾病的分类诊断任务中更具优势。同年,Kong等人^[10]首先构建受试者的人脑结构网络,然后根据F1-score进行特征选择,最后选择F1-score值排在前3000的特征作为SSAE方法的输入完成分类任务,在小型ABIDE I(autism brain imaging data exchange I)数据集上取得了高达90.93%的分类准确率。为了缓解样本少对分类性能的影响,有研究人员将迁移学习方法应用到脑网络的分类研究中。例如He等人^[11]首先在ABIDE I数据集上用SSAE方法无监督预训练一个模型用于学习ASD患者的脑网络数据蕴涵的潜在特征,然后将其应用于早产儿认知功能缺陷的分类诊断中,该研究首次验证了迁移学习在脑网络分类研究中的可行性。然而,FCNN方法存在2个影响脑网络分类性能的缺陷:一方面其参数量通常非常大,而脑网络数据样本少,容易出现因模型复杂导致的过拟合现象;另一方面,其输入为1维向量,需要将脑网络的邻接矩阵数据重塑为1维,破坏了脑网络中的拓扑结构信息^[4]。

CNN方法具有权值共享和局部连接特性。相比于FCNN方法,CNN方法不仅有效地降低了参数量,还在处理网格型数据方面有着巨大优势。除了在人脸识别等图像分类领域中取得的成功^[12-13],近年来也被成功应用于脑网络分类研究中。2017年Kawahara等人^[14]首次专门为脑网络提出了基于CNN的BrainNetCNN方法,其考虑到脑网络的拓扑局部性,并依次设计了3种卷积层分别用于提取不同层级的脑网络拓扑特征。同年,Meszlenyi等人^[15]提出了CCNN(connectome CNN)方法,该方法充分利用了CNN方法中的多通道结构和多种具有不同信息的脑网络数据以提高脑网络分类效果。然而,上述2个研究均使用1维卷积核对脑网络数

据进行逐行(列)的特征提取,根据 CNN 方法的权值共享机制,在这种逐行(列)的特征提取过程中会为连接至同一节点的每条边赋予相同的权重,显然有悖于神经科学中每条边都具有特定含义的特性^[4].2018 年 12 月考虑到脑网络中每条边的独特意义,我们提出了一种具有独立卷积核的卷积神经网络(convolutional neural network with element-wise filters, CNN-EW)的方法,这种卷积核赋予每个节点和每条边一个唯一的权重,能够更真实地反映脑网络的拓扑结构信息.在 ABIDE I 数据集上的实验表明,该方法不仅能够准确识别不同受试者是否患病,而且能够精确地识别异常脑区,表现出一定的临床应用潜力^[16].

综上,尽管已有研究为脑网络分类提供了一些解决方案,但是其分类性能仍然有待进一步提升.在真实的临床应用中,为了更加准确地对患者进行诊断,专家通常会参考患者的临床表型信息,例如年龄、性别、智商等.近年来,许多研究表明临床表型信息可有助于脑疾病诊断^[17-20].可以说,临床表型差异是脑网络差异的外在表现,而临床表型数据往往获取成本较低,非常适合作为脑网络分类的辅助数据.因此,本文提出一种基于自适应多任务卷积神经网络的脑网络分类方法.该方法首先以脑网络分类任务为主要任务,临床表型预测(回归)任务为辅助任务,通过多任务卷积神经网络并行学习获得分类模型;同时,在多任务学习过程中利用一种新的自适应多任务学习方法来自动确定各个子任务的权重,以降低实验成本并减少人为操作导致的分类误差.具体来说,本文的主要贡献包括 2 个方面:1)将临床表型预测任务引入脑网络分类研究中,通过多任务学习来提高脑网络的分类性能;2)提出了一种自适应多任务学习方法,该方法可以自适应地为辅助任务的赋予不同权重从而进一步提高多任务学习的性能.

1 相关工作

多任务学习(multi-task learning, MTL)方法是指一类同时学习多个子任务的机器学习方法.这类方法具有共享表示机制,能够有效地从多个任务中学习共享特征,进而提高泛化能力.尽管 MTL 方法已经在物体检测、自动驾驶等计算机视觉领域有许多成功的应用,将 MTL 方法应用于其他相对不成熟的领域仍存在许多困难.大多数单任务的机器

学习算法通过最小化一个损失函数来进行特征学习.然而,直接将 MTL 方法的损失函数定义为多个子任务损失函数相加,最终往往会由于不同子任务的学习难度和损失值规模存在差异,导致部分子任务主导了最小化 MTL 方法的损失函数,而其余子任务难以参与的现象.一个简单的解决方法是对不同子任务损失函数施加不同的子任务权重进行加权求和:

$$\mathcal{L}_{\text{all}} = \sum_{t=0}^T \alpha_t \mathcal{L}_t, \quad (1)$$

其中, \mathcal{L}_t 是第 t 个子任务的损失函数, α_t 是相应子任务的权重.通过调整子任务的权重设置,多任务学习可以使各个子任务的损失值对共享表示层的影响大致相同,从而泛化共享特征.目前,各个子任务的权重通常根据研究人员的经验设定或通过实验进行多次调整,需要耗费较高的实验成本.

2018 年文献[21]提出基于同方差不确定性来自动确定各个子任务的权重.该方法将分类、回归任务的输出分别定义一个服从方差为 σ^2 的分布,并通过估计方差值来衡量相应任务的重要性,更具体来说:

首先,重新将分类和回归任务的输出定义为服从方差为 σ^2 的分布.如式(2)所示,回归任务的输出被定义为一个方差为 σ^2 的高斯分布,其中, x 和 y 分别为样本和样本的标签, W 和 $f(x; W)$ 分别为神经网络的权重和输出.

$$p(y | f(x; W); \sigma) = \mathcal{N}(f(x; W), \sigma^2). \quad (2)$$

分类任务的输出是离散的,因此,如式(3)所示,其输出被定义为服从温度系数(即方差)为 σ^2 的 Gibbs 分布.

$$p(y | f_c(x; W); \sigma) = f_{\text{Softmax}}\left(\frac{1}{\sigma^2} f_c(x; W)\right). \quad (3)$$

然后,分别对 2 类任务的方差 σ^2 进行最大似然估计,得到回归任务的对数似然函数:

$$\ln p(y | f(x; W); \sigma) \approx -\frac{1}{2\sigma^2} \mathcal{L}_{\text{Regression}} - \ln \sigma, \quad (4)$$

分类任务的对数似然函数:

$$\ln p(y | f_c(x; W); \sigma) \approx -\frac{1}{\sigma^2} \mathcal{L}_{\text{Softmax}} - \ln \sigma^2, \quad (5)$$

其中, $\mathcal{L}_{\text{Regression}}$ 表示回归任务中常用的损失函数,如最小二乘(least square, LS)损失函数; $\mathcal{L}_{\text{Softmax}}$ 表示分类任务中常用的损失函数,如交叉熵(cross entropy, CE)损失函数.

最后,将上述方法应用在多任务学习过程中,通

过最大化对数似然函数来进行多任务学习。例如一个同时学习回归任务 $p(y_1 | f(\mathbf{x}; \mathbf{W}); \sigma_1)$ 和分类任务 $p(y_2 | f(\mathbf{x}; \mathbf{W}); \sigma_2)$ 的多任务学习方法的对数似然函数:

$$\begin{aligned} \ln p(y_1, y_2 | f(\mathbf{x}; \mathbf{W}); \sigma_1, \sigma_2) = \\ \ln \mathcal{N}(f(\mathbf{x}; \mathbf{W}), \sigma_1^2) \times f_{\text{Softmax}}\left(\frac{1}{\sigma_2^2} f(\mathbf{x}; \mathbf{W})\right) \approx \\ -\frac{1}{2\sigma_1^2} \mathcal{L}_{\text{Regression}} - \frac{1}{\sigma_2^2} \mathcal{L}_{\text{Softmax}} - \ln \sigma_1 - \ln \sigma_2^2, \quad (6) \end{aligned}$$

而最大化式(6)就等价于最小化式(6)的相反数。

最终,多任务学习的损失函数可被定义为

$$\begin{aligned} \mathcal{L}(\mathbf{W}, \sigma_1, \sigma_2) = \alpha_1 \mathcal{L}_{\text{Regression}} + \alpha_2 \mathcal{L}_{\text{Softmax}} + \\ \ln \sigma_1 + \ln \sigma_2^2, \quad (7) \end{aligned}$$

其中, $\alpha_1 = 1/2\sigma_1^2$, $\alpha_2 = 1/\sigma_2^2$ 。

定义好损失函数后,即可利用梯度更新算法学习神经网络权重参数 \mathbf{W} 和子任务方差 σ_1^2 和 σ_2^2 ,并自动确定了相应任务的权重 α_1 和 α_2 。

尽管文献[21]的方法为自动确定子任务权重提供了一个初步解决思路,但是,它存在 3 点不足:

1) 需要为各子任务方差 σ_i^2 设置初始值,而不同的初始值可能会对方差收敛结果和模型性能产生影响。

2) 在相同的学习率下基于同一梯度更新算法同时学习方差 σ_i^2 和神经网络权重 \mathbf{W} 。但是,无论从权重的意义还是求偏导的角度来看,它们是 2 类完全不同的参数,收敛速度存在差异,同时学习这 2 类参数易导致不收敛的风险。

3) 对自动确定的子任务权重没有约束,即 $\alpha_i \in (0, +\infty)$ 且相互独立。这使得在梯度更新过程中容易因梯度过大而导致参数溢出。

因此,我们提出了一种新的自适应多任务学习

方法,并将其应用于脑网络分类的研究中,该方法能够有效地改进文献[21]方法存在的以上 3 点不足。

2 自适应多任务卷积神经网络的脑网络分类

本节我们分 2 部分由浅至深地对基于自适应多任务卷积神经网络的脑网络分类方法进行介绍,分别为多任务卷积神经网络(2.1 节)和自适应多任务卷积神经网络(2.2 节)。

2.1 多任务卷积神经网络

图 1 是基于多任务卷积神经网络(multi-task CNN-EW, MTCNN-EW)的脑网络分类方法的框架。其中,共享表示层 $L_0 \sim L_3$ 为文献[16]提出的 CNN-EW 方法中的 E2Net-EW 模型,其中,输入层 L_0 用于接收脑网络数据, L_1 层用于从脑网络“边”级特征图中提取出“节点”级特征, L_2 层用于从“节点”级特征图中提取出“图”级特征, L_3 层从“图”级特征图中进一步提取出高级特征;任务特定层 L_4 和 L'_4 分别为以脑网络分类作为主要任务和以临床表型预测(回归)为辅助任务的输出层。MTCNN-EW 方法需要人工为各子任务设置权重。我们人工将主要任务权重设置为 α_0 ,同时均分各个辅助任务的权重,使得权重总和为 1,以各个任务的损失函数与相应权重的加权求和来定义总体损失函数:

$$\begin{aligned} \mathcal{L}_{\text{MTL}}(\mathbf{W}, \mathbf{b}) = \alpha_0 \mathcal{L}_0(\mathbf{W}, \mathbf{b}) + \\ \frac{1}{T} (1 - \alpha_0) \sum_{t=1}^T \mathcal{L}_t(\mathbf{W}, \mathbf{b}), \quad (8) \end{aligned}$$

其中, $\mathcal{L}_0(\mathbf{W}, \mathbf{b})$ 是脑网络分类任务的交叉熵损失函数, $\mathcal{L}_1(\mathbf{W}, \mathbf{b})$ 至 $\mathcal{L}_T(\mathbf{W}, \mathbf{b})$ 是各个临床表型辅助任务的最小二乘损失函数, T 是临床表型预测任务的个数。

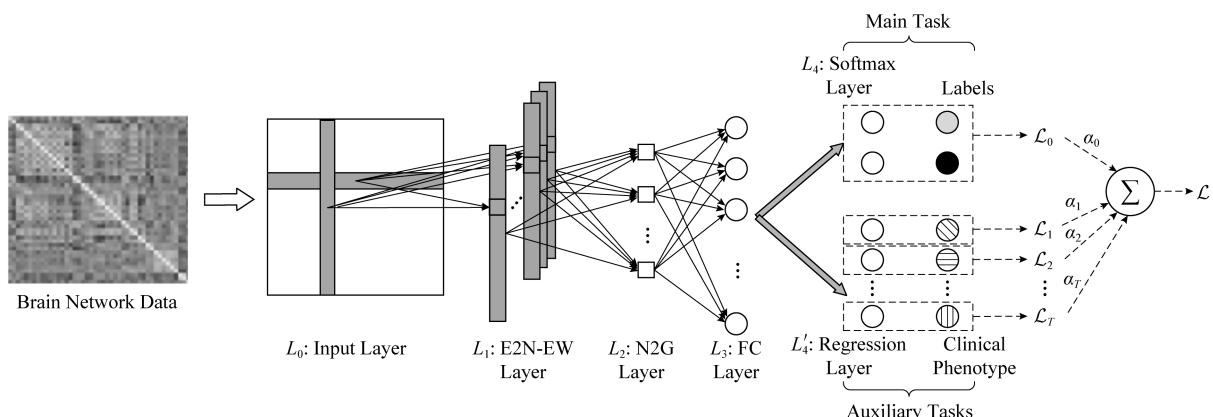


Fig. 1 Framework of brain networks classification based on multi-task CNN-EW (MTCNN-EW)

图 1 基于多任务卷积神经网络(MTCNN-EW)的脑网络分类方法的框架

根据上述介绍,基于 MTCNN-EW 的脑网络分类方法的训练过程如算法 1 所示:

算法 1. 基于多任务卷积神经网络(MTCNN-EW)的脑网络分类方法.

输入:训练集($\mathbf{X}, \mathbf{Y}, \mathbf{P}$),其中脑网络类别标签 $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2)$ 、临床表型数据 $\mathbf{P} = (\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_T)$,脑网络分类任务权重 α_0 、临床表型辅助任务权重 $\{\alpha_1, \alpha_2, \dots, \alpha_T\}$, $\alpha_t = (1 - \alpha_0)/T$;

输出:最优神经网络权重 \mathbf{W}^* 和偏置 \mathbf{b}^* .

- ① 随机初始化神经网络权重 \mathbf{W} 和偏置 \mathbf{b} ;
- ② while 未达到停止条件 do
- ③ 前向传播.如图 1 所示,以脑网络数据 \mathbf{X} 作为 MTCNN-EW 方法的输入,完成前向传播过程,并获得分类任务的输出 $\bar{\mathbf{Y}}_0$ 和临床表型辅助任务的输出($\bar{\mathbf{p}}_1, \bar{\mathbf{p}}_2, \dots, \bar{\mathbf{p}}_T$);
- ④ 计算各个子任务损失值.根据输出 $\bar{\mathbf{Y}}_0$ 和 $(\bar{\mathbf{p}}_1, \bar{\mathbf{p}}_2, \dots, \bar{\mathbf{p}}_T)$ 以及标签 \mathbf{Y}_0 和 $(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_T)$,分别计算脑网络分类任务的交叉熵损失值 \mathcal{L}_0 和临床表型辅助任务的最小二乘损失值 $\{\mathcal{L}_1, \mathcal{L}_2, \dots, \mathcal{L}_T\}$;
- ⑤ 计算整体损失值.根据式(9),计算 MTCNN-EW 方法的总体损失函数值;
- ⑥ 更新参数.使用梯度更新算法学习神经网络权重 \mathbf{W} 和偏置 \mathbf{b} ;
- ⑦ end while
- ⑧ $\mathbf{W}^* = \mathbf{W}, \mathbf{b}^* = \mathbf{b}$.

我们在算法 1 的损失函数中添加 L2 正则化项以缓解可能遇到的过拟合的问题:

$$\mathcal{L}(\mathbf{W}, \mathbf{b}) = \mathcal{L}_{\text{MTL}}(\mathbf{W}, \mathbf{b}) + \frac{\lambda}{2} \|\mathbf{W}\|_2^2, \quad (9)$$

其中,参数 λ 用于控制式(9)中 2 项的相对重要程度.同时,算法 1 在训练过程中采取提前停止策略以进一步避免过拟合,即当验证集上主要任务的损失值 $\mathcal{L}_0(\mathbf{X}_{\text{valid}}, \mathbf{Y}_{\text{valid}}; \mathbf{W}, \mathbf{b}) + (\lambda/2) \|\mathbf{W}\|_2^2$ 在达到最小值之后的 N_{patience} 次迭代内不再降低时提前停止,否则在完成最大迭代次数(maximum iterations) N_{max} 后停止.

在 MTCNN-EW 方法中,由于多任务学习具有共享表示机制,各个临床表型预测任务与脑网络分类任务之间共享脑网络特征,间接为脑网络类别决策提供辅助信息.

2.2 自适应多任务卷积神经网络

与大多数多任务学习一样, MTCNN-EW 方法

需要人工调整各个子任务的权重,造成较高的实验成本.因此,为了降低人工实验成本,并且精确地进行子任务权重设置,我们提出一种基于自适应多任务卷积神经网络(adaptive multi-task CNN-EW, AMTCNN-EW)的脑网络分类方法.该方法利用自适应多任务学习(adaptive multi-task learning, AMTL)方法,在训练过程中根据各个子任务的损失自适应地调整相应任务的权重 α_t .AMTCNN-EW 方法的损失函数定义为

$$\mathcal{L}_{\text{AMTL}} = \sum_{t=0}^T \alpha_t \mathcal{L}_t, \quad (10)$$

其中各个子任务权重无需人工手动调整.

在 AMTCNN-EW 方法中,AMTL 方法的基本思想是:首先,沿用文献[21]中回归和分类任务输出方差的估计方法;然后,直接利用微积分学中最大值求解方法解得各个子任务的输出方差估计值;最后,对每个任务的方差估计值进行了归一化,在获取相应任务的权重的同时确保不会因为梯度过大而导致参数溢出的问题.详细步骤为:

1) 子任务似然方程求解

首先,AMTL 方法基于文献[21]方法在式(4)和式(5)的基础上进行多任务的输出方差 σ^2 的估计.由于 $S = \ln \sigma^2$ 比 σ^2 更稳定,可以利用 $S = \ln \sigma^2$ 来代替直接估计 σ^2 ,所以回归和分类任务的对数似然函数分别为

$$\ln p(y | f(x; \mathbf{W}); S) \approx -\frac{1}{2} \exp(-S) \mathcal{L}_{\text{Regression}} - \frac{1}{2} S, \quad (11)$$

$$\ln p(y | f(x; \mathbf{W}); S) \approx -\exp(-S) \mathcal{L}_{\text{Softmax}} - S. \quad (12)$$

这种代替方式还有一个额外的好处,即可以避免式(4)和式(5)中出现除零的可能性.然后,由于式(11)和式(12)关于 σ^2, S 可微,因此可以直接利用微积分学中的最大值求解方法,将确定最大似然估计值 \hat{S} 的问题转化为对数似然方程求解的问题.2 类任务关于 S 的对数似然方程为

$$\frac{d}{dS} (\ln p(y | f(x; \mathbf{W}); S)) = 0. \quad (13)$$

解得回归和分类任务的最大似然估计值相同, \mathcal{L} 表示回归或分类任务的损失函数值:

$$\hat{S} = -\ln \frac{1}{\mathcal{L}}. \quad (14)$$

2) 自适应多任务权重调节

在自适应多任务学习过程中,我们将子任务的权重 α_t 设置为与其估计值 \hat{S}_t 呈反比,同时对其归一化以控制各个子任务权重 α_t 之和为 1.

$$\alpha_t = \frac{\exp(-\hat{S}_t)}{\sum_{t'=0}^T \exp(-\hat{S}_{t'})}. \quad (15)$$

显然,根据上述步骤求解子任务权重的 AMTL 方法能够有效避免文献[21]方法的不足,具体来说:

- 1) 直接求解各个子任务的估计值 \hat{S} , 避免了初始值设置对模型可能产生的影响;
- 2) 直接求解各个子任务的估计值 \hat{S} 使得在仅利用梯度更新算法学习神经网络权重 \mathbf{W} 即可, 避免了模型不收敛的风险;
- 3) 通过归一化使子任务权重 α_t 之和为 1, 在子任务间建立了相对的约束关系, 避免了梯度更新过程中因梯度过大而导致参数溢出风险。

综上, 基于 AMTCNN-EW 的脑网络分类方法的训练过程如算法 2 所示:

算法 2. 基于自适应多任务卷积神经网络(AMTCNN-EW)的脑网络分类方法。

输入: 训练集 $(\mathbf{X}, \mathbf{Y}, \mathbf{P})$, 其中, 脑网络类别标签 $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2)$ 、临床表型数据 $\mathbf{P} = (\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_T)$;

输出: 最优神经网络权重 \mathbf{W}^* 和偏置 \mathbf{b}^* ; 脑网络分类任务的最佳权重 α_0^* , 临床表型辅助任务的最佳权重 $\{\alpha_1^*, \alpha_2^*, \dots, \alpha_T^*\}$.

- ① 随机初始化神经网络权重 \mathbf{W} 和偏置 \mathbf{b} ;
- ② while 未达到停止条件 do
- ③ 前向传播. 如图 1 所示, 以脑网络数据 \mathbf{X} 作为 AMTCNN-EW 方法的输入, 完成前向传播过程, 并获得分类任务的输出 $\bar{\mathbf{Y}}_0$ 和临床表型辅助任务的输出 $(\bar{\mathbf{p}}_1, \bar{\mathbf{p}}_2, \dots, \bar{\mathbf{p}}_T)$;
- ④ 计算各个子任务损失值. 根据输出 $\bar{\mathbf{Y}}_0$ 和 $(\bar{\mathbf{p}}_1, \bar{\mathbf{p}}_2, \dots, \bar{\mathbf{p}}_T)$ 以及标签 \mathbf{Y}_0 和 $(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_T)$, 分别计算脑网络分类任务的交叉熵损失值 \mathcal{L}_0 和临床表型辅助任务的最小二乘损失值 $\{\mathcal{L}_1, \mathcal{L}_2, \dots, \mathcal{L}_T\}$;
- ⑤ 子任务似然方程求解. 利用式(14), 分别求解脑网络分类任务最大似然估计值 \hat{S}_0 和临床表型辅助任务最大似然估计值 $\{\hat{S}_1, \hat{S}_2, \dots, \hat{S}_T\}$;
- ⑥ 自适应多任务权重调节. 利用式(15), 自适应地为各个子任务计算权重 α_0 和 $\{\alpha_1, \alpha_2, \dots, \alpha_T\}$;
- ⑦ 计算整体损失值. 计算 AMTCNN-EW 方法的总体损失函数值;

- ⑧ 更新参数. 使用梯度更新算法学习神经网络权重 \mathbf{W} 和偏置 \mathbf{b} ;
- ⑨ end while
- ⑩ $\mathbf{W}^* = \mathbf{W}, \mathbf{b}^* = \mathbf{b}$;
- ⑪ for $t = 0, 1, \dots, T$ do
- $\alpha_t^* = \alpha_t$;
- ⑫ end for

算法 2 同样在损失函数中增加 L2 正则化项(式(16))并且采取提前停止策略.

$$\mathcal{L}(\mathbf{W}, \mathbf{b}) = \mathcal{L}_{\text{AMTL}}(\mathbf{W}, \mathbf{b}) + \frac{\lambda}{2} \|\mathbf{W}\|_2^2. \quad (16)$$

相比于 MTCNN-EW 方法, AMTCNN-EW 方法无需人工调整各个子任务的权重 α_t , 而是利用 AMTL 方法直接求解各个子任务的 \hat{S}_t (算法 2 行⑤)进而确定权重 α_t (算法 2 行⑥).

3 实验与结果

3.1 实验数据与预处理

- 1) 人脑功能网络数据

实验部分使用的数据集全部来自于 ABIDE I^[22]. ABIDE I 收集了 17 个国际站点, 共计 1112 名受试者的 rs-fMRI 数据和 sMRI(structural magnetic resonance imaging)数据, 并记录了相应受试者的临床表型.

原始 rs-fMRI 数据需要经过一系列预处理操作后得到人脑功能网络数据. 首先, 出于可复现性考量, 我们使用了 PCP(preprocessed connectomes project)项目^[22]提供的利用 DPARSF 软件进行预处理后的数据, 其中有 16 名受试者的预处理数据异常, 我们使用了数据正常的 569 个对照组受试者和 527 个实验组(即 ASD 患者)受试者的预处理数据. 然后, 利用自动解剖标记图谱(automated anatomical labeling, AAL)从预处理数据中提取感兴趣区域(ROI)的平均时间序列, 我们去除了位于小脑的脑区, 仅选取位于大脑皮层的 90 个脑区(即 $|V|=90$)作为 ROI. 最后, 利用皮尔森相关系数(Pearson correlation coefficient, PCC)计算每对 ROI 平均时间序列之间的相似度, 得到的带权邻接矩阵即可表示相应受试者的人脑功能网络.

2) 临床表型数据

由于来自不同站点的临床表型种类存在差异, 我们按照 2 个标准对临床表型进行筛选: 1) 临床表

型不能与脑疾病诊断结果直接相关,如长期用药史;2)缺失的数据占比不应高于30%。

按照上述方法对ABIDE I数据集的临床表型数据进行筛选,最终确定的临床表型(编号1~7)如表1所示。对于这7种临床表型中存在的数据缺失问题,我们根据相应临床表型数据中的非缺失值的数据分布特点填充缺失值,具体方法为:

1)年龄(age).众所周知,神经疾病在不同年龄段中的发病率不同,年龄是分析和诊断神经疾病不可忽视的因素和重要线索。当年龄数据出现缺失问题时,可以将缺失数据用相应数据组中(即对照组或实验组)其他受试者的平均值填充。

2)性别(sex).男性和女性在大脑功能和结构方面有明显差异,许多疾病在不同性别中的发病率、发病机理和症状表现不同,这使得性别成为诊断脑疾病的一大重要因素。考虑到数据采集时一般会尽量保证不同数据组的性别分布一致,因此,当性别数据出现缺失问题时,可以假定缺失数据为样本量较小的性别类型。

Table 1 Details of Clinical Phenotype from the ABIDE I Dataset

表1 从ABIDE I数据集中筛选出的临床表型详细信息

Phenotype	Phenotype Data			Distribution	
	ID	Phenotype	Value	ASD Group(527 Subjects)	Control Group(569 Subjects)
1	age		>0	17.1±8.1	16.8±7.7
2	sex	0: male		463	471
		1: female		64	98
		1: left		40	31
3	handedness	0: ambidextrous		12	9
		-1: right		475	529
4	FIQ		>0	105.0±17.0	110.6±12.9
5	VIQ		>0	103.5±17.6	109.5±14.3
6	PIQ	>0		104.7±16.8	106.1±14.2
		1: closed		163	181
		0: open		364	388
7	eye status		1: free	0	0

3.2 实验设置和超参数

实验过程中,所有基于深度学习的脑网络分类方法均采用如表2所示的超参数设置,以Adam方法作为梯度更新算法来最小化损失函数。实验数据按照3:1:1划分数据集,展示的所有结果均在测试集上由10次5折交叉验证法给出。

3.3 MTCNN-EW方法权重设置对分类准确率的影响

在多任务学习中,每个任务的权重通常根据研究人员的经验和实验结果进行设定,我们依次将

3)利手(handedness),指受试者的常用手。作为大脑结构和功能不对称性的外部表现,利手与许多疾病存在密切的关联。由于大多数人为右利手,所以当利手数据出现缺失问题时,可假定为右利手。

4)韦氏智力量表是由心理学专家编制的一系列智力测验量表。在世界各国的医学领域中,它是最受重视的智力测验量表,能够测量不同年龄段受试者的全量表智商(full scale IQ, FIQ),言语智商(verbal IQ, VIQ)和操作智商(performance IQ, PIQ)。这3项智商评分服从 $\mathcal{N}(100, 15)$ 分布。因此,当其中存在数据缺失问题时,可以用服从 $\mathcal{N}(100, 15)$ 分布的随机数进行填充。我们在使用这3种智商评分时,需将其归一化至[-1,1]。

5)眼睛状态(eye status)。在数据采集过程中,受试者通常会被指定眼睛状态(睁眼或闭眼),眼睛状态不同,大脑活动的区域往往不同。但是,存在一部分研究没有指定并记录眼睛状态导致数据缺失问题,可以将眼睛状态填充为区别于睁眼和闭眼的自由状态。

Table 2 Hyper Parameter Settings

表2 超参数设置

Hyper Parameter	Value
λ Candidate Set	{ $1 \times 10^{-1}, 5 \times 10^{-2}, 1 \times 10^{-2}, 5 \times 10^{-3}, 1 \times 10^{-3}, 5 \times 10^{-4}$ }
Learning Rate	1×10^{-4}
Batch Size	96
Maximum Iterations N_{\max}	1×10^4
Parameter N_{patience}	50

MTCNN-EW 方法中脑网络分类任务的权重 α_0 设置为 $0.1 \sim 0.9$, 并将 $(1 - \alpha_0)$ 均分给各个临床表型辅助任务, 即 $\alpha_i = (1 - \alpha_0)/T$, $T = 7$. MTCNN-EW 方法的权重设置对分类准确率的影响如图 2 所示:

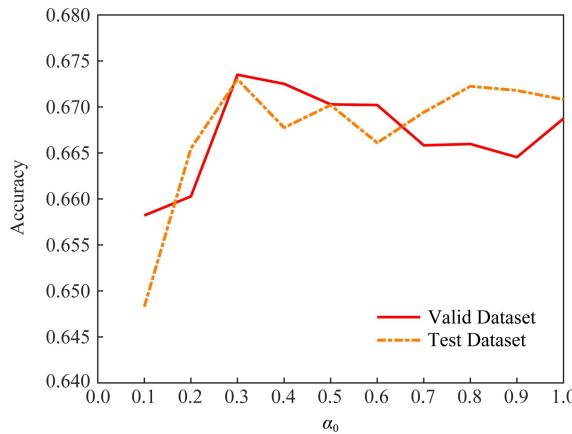


Fig. 2 Accuracy of MTCNN-EW with different α_0

图 2 不同权重 α_0 下 MTCNN-EW 方法的分类准确率

显而易见, 权重设置的合理性对多任务学习方法的分类性能至关重要, 具体来说: $\alpha_0 < 0.3$ 时, MTCNN-EW 方法的分类准确率在验证集和测试集上的表现相对较差; $\alpha_0 > 0.3$ 时, MTCNN-EW 方法的分类准确率在验证集和测试集之间逐渐出现明显的差异, 模型陷入了局部最优. 因此, $\alpha_0 = 0.3$ 为 MTCNN-EW 方法中脑网络分类任务的最佳权重设置, 此时分类准确率最高.

3.4 AMTCNN-EW 方法的性能测试

为了验证并分析本文提出的自适应多任务学习(AMTL)方法的性能, 我们在脑网络分类任务上分别从 2 个角度对 AMTCNN-EW 方法进行实验.

第 1 个角度从损失函数和子任务权重收敛情况出发, 通过对比 AMTLCNN-EW 方法和基于文献[21]的方法, 可验证和分析 AMTL 方法在多任务学习过程中对各个任务的损失函数值和相应任务权重的收敛效果.

基于文献[21]的方法其结果如图 3 所示, 为清晰起见, 图 3(b)中仅展示了主要任务——脑网络分类任务的权重(α_0)曲线, 和 3 项辅助任务的权重(α_2, α_3 和 α_6)曲线, 分别为性别、利手和操作智商. 由图 3 可知, 神经网络权重 \mathbf{W} 和各个子任务的权重 α_i 均出现无法正常收敛的情况, 具体来说, 当迭代次数 $Step = 250$ 左右时, 模型迅速过拟合, 这是由于该方法对各个子任务的权重没有限制, 在执行梯度更新算法过程中梯度过大导致; 当迭代次数 $Step > 500$ 时, 损失函数值下溢, 神经网络训练失败.

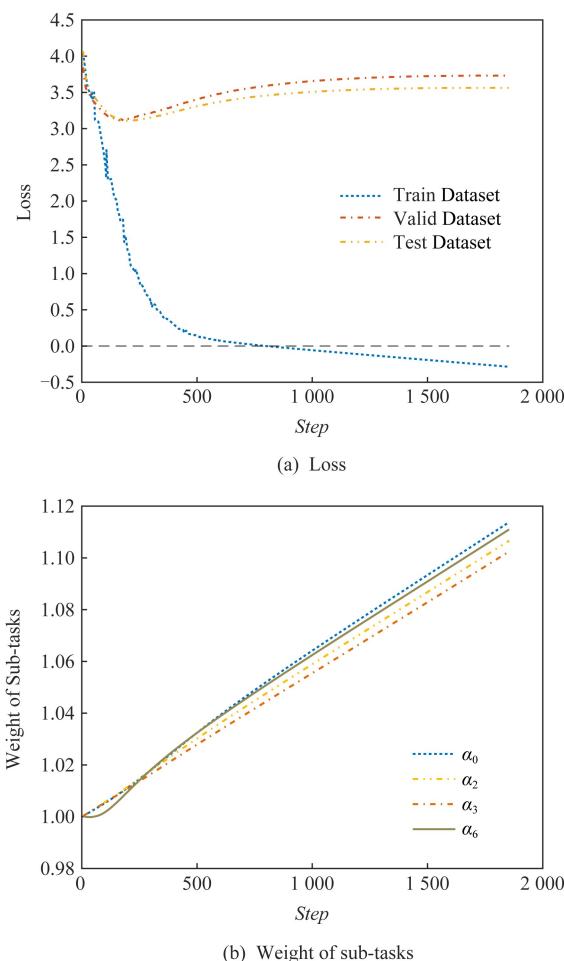


Fig. 3 Train plot of the method of Ref [21]

图 3 文献[21]方法的训练曲线

相反, 基于 AMTL 的 AMTLCNN-EW 方法其结果如图 4 所示, 无论是损失函数值还是子任务权重均如期收敛. 分析图 4(a)可知, $Step > 1000$ 时, AMTLCNN-EW 方法在训练集上损失值仍持续下降, 在验证集和测试集上的损失值能够保持稳定; 直到 $Step = 1800$ 时, 在训练集上的损失值开始趋于稳定, 此时提前结束训练, AMTLCNN-EW 方法没有出现过拟合现象. 由图 4(b)可知, $Step > 1000$ 后, 各个子任务的权重曲线逐渐平缓, 脑网络分类任务和 3 项辅助任务的权重最终收敛至 0.3, 0.1, 0.15 和 0.05. 可见, AMTL 方法可以自适应且准确地求解各个子任务权重.

第 2 个角度主要针对各个子任务的权重值进行分析. 通过对比 AMTCNN-EW 方法和需要人工设置子任务权重的 MTCNN-EW 方法, 可验证和分析 AMTL 方法学习得到的子任务权重是否合理.

首先利用 AMTCNN-EW 方法自适应地确定了各个子任务权重, 脑网络分类任务权重 α_0 最终平均

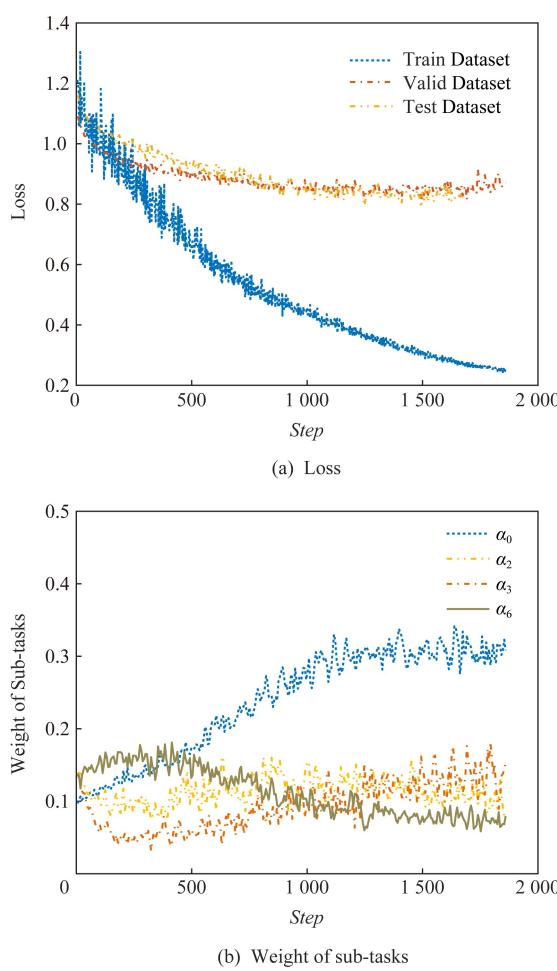


Fig. 4 Train plot of the AMTL

图4 AMTL方法的训练曲线

收敛值为 0.296, 7 项临床表型辅助任务权重 α_1 至 α_7 的平均收敛值分别为: 0.093, 0.114, 0.140, 0.085, 0.093, 0.090, 0.089。通过分析这几项权重可以看出, 利手和性别这 2 类临床表型能够提供相对有效的信息用于辅助脑网络分类。目前, 来自世界各国的许多研究表明男孩 ASD 发病率明显相对较高;也有一些研究发现, ASD 患者人群具有相对高的混合利手率^[19] 和左利手率^[20]。众所周知 ASD 患者的核心症状表现在语言交流方面存在障碍, AMTCNN-EW 方法中语言智商的权重值明显高于操作智商的权重的结果与这一现象不谋而合。另外, 各个临床表型辅助任务的权重值相差不大, 说明不同的临床表型辅助任务都能够为脑网络分类任务提供了一定信息。

对比 MTCNN-EW 方法和 AMTCNN-EW 方法的结果可知, 以脑网络分类任务作为多任务学习的主要任务, AMTCNN-EW 方法确定其权重为 $\alpha_0=0.296$, 该结果与 MTCNN-EW 方法中人工设置的最佳权重 $\alpha_0=0.3$ 基本一致。

以上结果说明 AMTL 方法可以自适应且准确地评估每个任务的相对重要程度, 利用 AMTL 方法可避免人工调试子任务权重, 节约实验成本。

3.5 2类多任务卷积神经网络对比实验

引入临床表型辅助任务主要是为了更好地完成脑网络分类任务。因此, 我们从文献[16]的 CNN-EW 方法中选取分类性能最佳的 E2Nnet-EW 模型作为基线模型, 分别从准确率(Accuracy)、灵敏度(Sensitivity)和特异性(Specificity)三个指标对比这 2 类引入了临床表型辅助任务的脑网络分类方法, 其结果如表 3 所示:

Table 3 Performance of 2 Types of Multi-Task

CNN-EW and AMTCNN-EW

表 3 2类多任务卷积神经网络与 CNN-EW 方法的评价指标

Methods	λ	Accuracy/%	Sensitivity/%	Specificity/%
CNN-EW	1×10^{-3}	66.88 ± 0.42	63.09 ± 0.58	70.40 ± 0.52
MTCNN-EW	1×10^{-3}	67.30 ± 0.56	62.96 ± 1.30	71.31 ± 1.20
AMTCNN-EW	1×10^{-3}	67.61 ± 0.39	63.27 ± 0.68	71.63 ± 0.48

由表 3 可知, 引入临床表型预测辅助任务, 并通过 MTCNN-EW 方法和 AMTCNN-EW 方法为脑网络分类提供辅助信息是可行的, 2 种方法均取得了比 CNN-EW 方法更好的分类性能。特别地, AMTCNN-EW 方法在 3 个评价指标上均达到最高值。这一结果证明了 AMTCNN-EW 方法能够更好的挖掘临床表型预测任务提供的信息, 使得模型学习到更加完备的脑网络特征, 从而获得更好的分类性能。需说明, 一旦这 2 类多任务卷积神经网络模型完成训练, 无需额外提供临床表型数据, 便可单独完成脑网络分类任务, 这一特点使这 2 类多任务脑网络分类方法自然地适合应用于真实的临床诊断场景中。综上所述, 引入临床表型预测辅助任务的思路是合理的, 且基于自适应多任务卷积神经网络(AMTCNN-EW)的脑网络分类方法不仅能够充分学习到临床表型预测任务提供的辅助特征, 而且能够自动确定各个子任务权重, 避免了人为操作带来的分类误差, 因而具有最好的分类性能。

3.6 多种经典方法的对比实验

如引言所述, 基于传统机器学习的方法和基于深度学习的方法是目前解决脑网络分类问题常用的 2 类方法。为了验证本文提出的 AMTCNN-EW 方法相比于其他方法的优劣, 我们从上述 2 类方法中选取 5 种经典方法与 AMTCNN-EW 方法进行对比实验, 分别实现了文献[3]首次提出并已广泛应用于脑网络分类问题中的基于 RFE_SVM 的方法, 其

中,特征选择的数量为 200;文献[5]提出的基于 LASSO 的方法,其中,L1 正则化项的权重衰减参数 $\lambda_{L1}=5\times10^{-3}$;文献[9]提出的基于 SSAE 的方法,包括 2 个具有 200 个神经元的隐藏层,并且,稀疏性参数 $\rho=0.1$,稀疏性惩罚因子的权重 $\beta=3.0$,L2 正则化项的权重衰减参数 $\lambda_{L2}=1\times10^{-2}$;文献[14]中提出的基于 CNN 的 BrainNetCNN 方法,我们选择脑网络分类性能最好的 E2Nnet 模型,其中 $\lambda_{L2}=1\times10^{-2}$;文献[16]中提出的 CNN-EW 方法,同样选择脑网络分类性能最好的 E2Nnet-EW 模型,其中 $\lambda_{L2}=1\times10^{-3}$.由于这些文献中使用的数据与本文不同,因此,上述所有方法的参数均为在本文数据集上的最佳配置.

由图 5 所示的实验结果可以发现,AMTCNN-EW 方法的各项评价指标均明显高于其他方法.这 6

种方法中,基于深度学习的脑网络分类方法优于传统机器学习方法,这体现了深度学习在处理高维样本时的优势.在 4 种深度学习方法中,AMTCNN-EW 方法优势明显.特别地,AMTCNN-EW 方法的共享表示层部分与 E2Nnet-EW 方法一致,但是经过 T 检验证两者的分类性能,AMTCNN-EW 方法的分类准确率和特异性显著优于前者($p<0.05$),主要有 2 方面原因:1)AMTCNN-EW 方法通过多任务学习的共享表示机制学习临床表型差异与脑网络异常之间共享特征,为脑网络类别决策起到了辅助作用;2)AMTCNN-EW 方法通过 AMTL 方法求解各个子任务的权重,能够比人工调整更准确地设置每个任务对整体的贡献.总而言之,通过 AMTCNN-EW 方法引入并利用多种临床表型辅助任务能够有效提高脑网络分类性能.

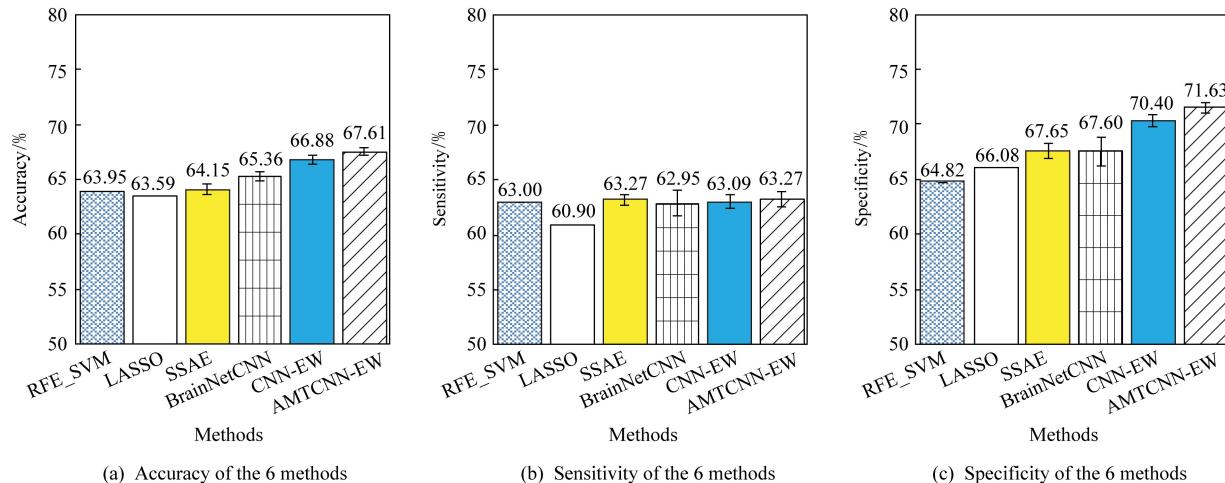


Fig. 5 Comparison of the 6 methods' performance

图 5 6 种方法的分类性能对比

4 总 结

鉴于不同人群的临床表型与其脑网络差异存在着一定的依存关系,有望为脑网络分类提供有用的信息.因此,为了进一步提升脑网络分类方法的性能,我们提出了一种基于自适应多任务卷积神经网络的脑网络分类方法.本文分 2 部分由浅至深地对基于多任务卷积神经网络(MTCNN-EW)、自适应多任务卷积神经网络(AMTCNN-EW)2 类脑网络分类方法进行全面的研究和探索.特别地,本文提出的 AMTCNN-EW 方法不仅能够充分利用临床表型预测任务提供的辅助信息,而且能够在训练过程中自适应且准确地求解各个任务的权重,降低人工操

作带来的实验成本和分类误差.在 ABIDE I 真实数据集上的实验结果表明:1)通过引入临床表型预测任务,并利用 MTCNN-EW 方法和 AMTCNN-EW 方法提取临床表型差异与脑网络异常之间的共享特征,能够有效辅助脑网络分类;2)AMTCNN-EW 方法能够较准确地对各个子任务的权重进行求解,进一步提升了分类性能,而且与多种经典方法相比具有明显的优势.

参 考 文 献

- [1] Glasser M F, Smith S M, Marcus D S, et al. The human connectome project's neuroimaging approach [J]. Nature Neuroscience, 2016, 19(9): 1175–1187

- [2] Sporns O, Tononi G, Kotter R. The human connectome: A structural description of the human brain [J]. *Plos Computational Biology*, 2005, 1(4): 245–251
- [3] Craddock R C, Holtzheimer III P E, Hu X P, et al. Disease state prediction from resting state functional connectivity [J]. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 2009, 62(6): 1619–1628
- [4] Brown C J, Hamarneh G. Machine learning on human connectome data from MRI [J]. *arXiv preprint, arXiv:1611.08699*. 2016
- [5] Meszlenyi R J, Hermann P, Buza K, et al. Resting state fMRI functional connectivity analysis using dynamic time warping [J]. *Frontiers in Neuroscience*, 2017, 11: Article 75
- [6] Kim J, Calhoun V D, Shim E, et al. Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: Evidence from whole-brain resting-state functional connectivity patterns of schizophrenia [J]. *Neuroimage*, 2015, 124: 127–146
- [7] Guo Xinyu, Dominick K C, Minai A A, et al. Diagnosing autism spectrum disorder from brain resting-state functional connectivity patterns using a deep neural network with a novel feature selection method [J]. *Frontiers in Neuroscience*, 2017, 11: Article 460
- [8] Zeng Lingli, Wang Huaning, Hu Panpan, et al. Multi-site diagnostic classification of schizophrenia using discriminant deep learning with functional connectivity MRI [J]. *EBioMedicine*, 2018, 30: 74–85
- [9] Ju Ronghui, Hu Chenhui, Zhou Pan, et al. Early diagnosis of alzheimer's disease based on resting-state brain networks and deep learning [J]. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2019, 16(1): 244–257
- [10] Kong Yazhou, Gao Jianliang, Xu Yunpei, et al. Classification of autism spectrum disorder by combining brain connectivity and deep neural network classifier [J]. *Neurocomputing*, 2019, 324: 63–68
- [11] He Lili, Li Hailong, Holland S K, et al. Early prediction of cognitive deficits in very preterm infants using functional connectome data in an artificial neural network framework [J]. *Neuroimage Clinical*, 2018, 18: 290–297
- [12] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks [C] // Proc of Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, 2012: 1097–1105
- [13] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition [J]. *Proceedings of the IEEE*, 1998, 86(11): 2278–2324
- [14] Kawahara J, Brown C J, Miller S P, et al. BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment [J]. *Neuroimage*, 2017, 146: 1038–1049
- [15] Meszlenyi R J, Buza K, Vidnyanszky Z. Resting state fMRI functional connectivity-based classification using a convolutional neural network architecture [J]. *Frontiers in Neuroinformatics*, 2017, 11: Article 61
- [16] Xing Xinying, Ji Junzhong, Yao Yao. Convolutional neural network with element-wise filters to extract hierarchical topological features for brain networks [C] //Proc of 2018 IEEE Int Conf on Bioinformatics and Biomedicine (BIBM). Piscataway, NJ: IEEE, 2018: 780–783
- [17] Akshoomoff N, Pierce K, Courchesne E. The neurobiological basis of autism from a developmental perspective [J]. *Development and Psychopathology*, 2002, 14(3): 613–634
- [18] Fombonne E. The epidemiology of autism: A review [J]. *Psychological Medicine*, 1999, 29(4): 769–786
- [19] Escalante-Mead P R, Minshew N J, Sweeney J A. Abnormal brain lateralization in high-functioning autism [J]. *Journal of Autism and Developmental Disorders*, 2003, 33(5): 539–543
- [20] McManus I, Murray B, Doyle K, et al. Handedness in childhood autism shows a dissociation of skill and preference [J]. *Cortex*, 1992, 28(3): 373–381
- [21] Kendall A, Gal Y, Cipolla R. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics [C] //Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 7482–7491
- [22] Di M A, Yan C G, Li Qingyang, et al. The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism [J]. *Molecular Psychiatry*, 2014, 19(6): 659–667
- [23] Cameron C, Yassine B, Chu C, et al. The neuro bureau preprocessing initiative: Open sharing of preprocessed neuroimaging data and derivatives [J]. *Frontiers in Neuroinformatics*, 2013, 4: 41



Xing Xinying, born in 1994. Master. Her main research interests include deep learning and bioinformatics.



Ji Junzhong, born in 1969. Professor and PhD supervisor at Faculty of Information Technology. Senior member of CCF. His main research interests include data mining, machine learning, swarm intelligence and bioinformatics.



Yao Yao, born in 1992. PhD candidate. His main research interests include computation intelligence, deep learning and brain science. (yaoyaoyao1314@emails.bjut.edu.cn)