

## 基于多模态表示学习的阿尔兹海默症诊断算法

樊连玺<sup>1</sup> 刘彦北<sup>2,3</sup> 王雯<sup>2</sup> 耿磊<sup>2</sup> 吴骏<sup>1</sup> 张芳<sup>2</sup> 肖志涛<sup>2</sup>

1 天津工业大学电子与信息工程学院 天津 300387

2 天津工业大学生命科学院 天津 300387

3 天津光电检测技术与系统重点实验室 天津 300387

(1831095395@tiangong.edu.cn)

**摘要** 阿尔兹海默症是一种典型的涉及多种致病因素的神经系统退行性疾病。然而,阿尔兹海默症的病因尚不明确,病程不可逆转,且无治愈方法,因此其早期诊断和治疗一直是人们关注的重点。受试者的神经影像数据对于该疾病的诊断具有重要的辅助作用,而结合多个模态的数据可进一步提高诊断效果。目前,联合该疾病的多模态数据进行辅助诊断逐渐成为一个新兴的研究领域。在此提出了一种基于自编码器的多模态表示学习方法,用于阿尔兹海默症的诊断。首先将多个模态的数据进行初步融合,得到初级的共同表示;然后将其送入自编码器网络,学习隐空间中的共同表示;最后对隐空间中的共同表示进行分类,得到疾病的诊断结果。在国际公开 ADNI 数据集上,所提算法对患病和健康受试者的诊断准确率达到 88.9%,与同类算法相比取得了最好的诊断效果。实验结果验证了所提算法对阿尔兹海默症诊断的有效性。

**关键词:** 阿尔兹海默症;疾病诊断;多模态融合;表示学习;自编码器网络

**中图分类号** TP183

## Multimodal Representation Learning for Alzheimer's Disease Diagnosis

FAN Lian-xi<sup>1</sup>, LIU Yan-bei<sup>2,3</sup>, WANG Wen<sup>2</sup>, GENG Lei<sup>2</sup>, WU Jun<sup>1</sup>, ZHANG Fang<sup>2</sup> and XIAO Zhi-tao<sup>2</sup>

1 School of Electronics and Information Engineering, Tiangong University, Tianjin 300387, China

2 School of Life Sciences, Tiangong University, Tianjin 300387, China

3 Tianjin Key Laboratory of Optoelectronic Detection Technology and Systems, Tianjin 300387, China

**Abstract** Alzheimer's disease (AD) is a complex neurodegenerative disease involving a variety of pathogenic factors. So far, the cause of Alzheimer's disease is not clear, the course of the disease is irreversible, and there is no cure. Its early diagnosis and treatment have always been the focus of attention. The neuroimaging data of subjects has an important auxiliary role in the diagnosis of this disease, and the combination of multimodal data can further improve the diagnostic effect. At present, the multimodal data representation learning of the disease has gradually become an emerging research field, which has attracted wide attention from researchers. An autoencoder based multimodal representation learning method for Alzheimer's disease diagnosis is proposed. Firstly, the multimodal data are initially fused to obtain the primary common representation. Then, it is input into the autoencoder network to learn the final common representation in latent space. Finally, the common representation in latent space is classified to obtain the disease result. The proposed method, which achieves the best diagnostic results compared with comparison algorithms, has an accuracy of 88.9% in the classification of AD and healthy subjects in the ADNI dataset. Extensive experimental results verify its effectiveness.

**Keywords** Alzheimer's disease, Disease diagnosis, Multimodal fusion, Representation learning, Autoencoder network

## 1 引言

阿尔兹海默症 (Alzheimer's Disease, AD) 及其早期阶段——轻度认知障碍 (Mild Cognitive Impairment, MCI), 是

一种强烈的不可逆转的神经系统退行性疾病, 导致了世界各地许多老年人的死亡。它的早期诊断和治疗对提高患者的生活质量具有重要意义。在计算机辅助诊断领域, 近些年有很多研究对该疾病进行了深入的探索<sup>[1]</sup>。Zhang 等<sup>[2]</sup>提出了一

到稿日期:2020-09-25 返修日期:2020-11-26 本文已加入开放科学计划(OSID), 请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金项目(61901297);天津市科学技术重大专项与工程(17ZXSCSY00060)

This work was supported by the National Natural Science Foundation of China (61901297) and Tianjin Science and Technology Major Projects and Engineering(17ZXSCSY00060).

通信作者:刘彦北(liuyanbei@tiangong.edu.cn)

种基于图像块的阿尔兹海默症诊断方法,其在受试者的影像学数据中检测有物理意义的关键点,然后以此为中心提取图像块的形态学特征,最后利用 SVM 进行诊断。在最近的研究中,Lian 等<sup>[3]</sup>提出了一种分层全卷积网络,联合学习和融合多尺度特征表示,以构建诊断模型。此外,Lin 等<sup>[4]</sup>利用磁共振脑图像解剖学特征的变化,提出了基于极限学习机的阿尔兹海默症诊断方法。Zeng 等<sup>[5]</sup>基于感兴趣区域,提出了 3D 卷积神经网络诊断模型。上述工作都是基于单一模态进行诊断,但在实际应用中,我们可以获取到多种模态的数据(如磁共振成像(MRI)和正电子发射断层成像(PET))或多种类型的特征用于阿尔兹海默症的诊断<sup>[6]</sup>。多模态的阿尔兹海默症数据是对同一受试者的多方信息表示,它们具有一定的互补性。因此,可以联合多个模态信息,学习一个共同表示,进而进行疾病诊断,该任务可以建模为基于多模态表示学习的阿尔兹海默症诊断框架。

多模态表示学习旨在学习一个比原始数据更好完成任务的新表示。早期的多模态研究通常基于共同训练的方法探索模态间的最小差异<sup>[7]</sup>。此外,基于典型相关性分析(Canonical Correlation Analysis, CCA)<sup>[8]</sup>的方法(比如 Kernel Canonical Correlation Analysis, KCCA<sup>[9]</sup>, Deep Canonical Correlation Analysis, DCCA<sup>[10]</sup>等)被广泛应用于多模态表示学习。这些方法都采用在不同的模态中学习一个潜在的公共子空间的方式。对于阿尔兹海默症的诊断,最近的工作使用 CCA 将原始特征从多个不同的模态转换到一个共同的空间表示<sup>[11-12]</sup>。在多模态学习中也有一些基于多核的方法<sup>[13-14]</sup>,它们计算每个模态的核矩阵,以获得核的最佳组合,最终学习到多模态数据的共同表示。Zhou 等<sup>[15]</sup>提出了多模态表示学习的诊断方法,将多个模态映射至同一个空间进行预测。Shi 等<sup>[16]</sup>提出了一种基于耦合特征表示的特征级耦合交互捕获方法,用于阿尔兹海默症的检测。此外,受试者年龄、性别、受教育年限等因素也可视为一种模态被引入机器学习模型,同时结合多任务的学习模式以提高诊断效果<sup>[17-19]</sup>。

上述方法都是将多个模态的数据映射至一个共同的特征空间,同时约束各个模态之间的相关性或独立性,由于它们需要将具有较好表征能力的特征向量作为输入,因此均不适用于较为复杂的阿尔兹海默症数据。

结合多模态数据的阿尔兹海默症诊断具有多个挑战:首先,由于目前人类对于阿尔兹海默症的认识还不够深入,提取到的特征向量往往还不具有较好的表征能力,为了更好地融合多模态数据的信息,需要对复杂的阿尔兹海默症数据进行较好的预处理和特征提取;其次,降维后的特征向量往往能够使样本具有更好的区分度,可以提高算法模型的运算效率,因此,如何融合受试者的多模态信息、有效学习共同的低维表示也是一个难点;最后,学习到的共同表示要既能包含多模态数据的共同信息,也能重构各个模态特有的信息,这样的共同表示才能包含更多的数据信息。

为了应对上述问题,本文提出了一种基于自编码器的多模态表示学习方法,用于阿尔茨海默症的诊断。首先,对原始

的多模态数据进行预处理和特征提取;然后将多模态数据进行融合得到初级的共同表示,进而借助自编码器网络将其编码至隐空间,其中解码器可将隐空间的表示进行重构;最后在隐空间可得到融合了多模态信息的最终共同表示。在此共同表示的基础上,可以对其分类进而得到疾病的诊断结果。本文的贡献可总结为:

(1)本文研究了如何有效融合多模态样本信息进行数据表示学习的问题,并将其应用于阿尔兹海默症的诊断。

(2)本文提出了一个基于自编码器的多模态表示学习框架,利用自编码神经网络的优势,能够将多模态数据信息进行有效的非线性融合,从而学习到多模态数据的共同表示。

(3)提出的方法在国际公开数据集 ADNI 中取得了较好的诊断效果,大量的实验结果验证了本文算法的有效性。

## 2 相关工作

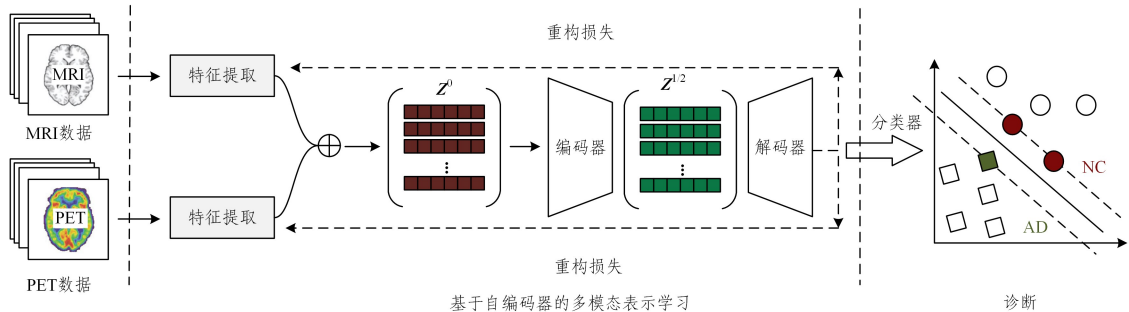
传统的机器学习方法旨在识别单视图数据中的相似行为组<sup>[20-21]</sup>。但是在很多的机器学习问题中,一个样本往往用多个视角或模态的特征集合来表示,每个特征集都可以被视为原始数据集的一个视角。这些特征集可以分为两部分,即不同类型的特征和来自不同数据源的特征<sup>[22]</sup>。例如,在基于内容的网络图像检索中,一个对象可同时由图像和图像周围文本(即不同的数据源)的视觉特征来描述,其中每个特征描述同一样本的不同独立信息。此外,多模态学习的一个值得注意的特点是,即使在缺乏自然特征分割的情况下,制造多个模态也可以提高性能<sup>[23]</sup>。本文研究的阿尔兹海默症数据(磁共振和正电子发射断层成像)来自多个数据源。

近年来,一系列的无监督多模态学习方法被提出以适应更为复杂的无标签环境<sup>[24-25]</sup>。如 Wang 等采用协同正则聚类假设,以利用谱聚类框架内的互补性信息<sup>[26]</sup>。协同训练的方法被用来搜索不同视图之间一致的聚类<sup>[7]</sup>。基于子空间的方法通过自表示来探讨样本之间的关系<sup>[27-28]</sup>。多模态学习要尽可能充分地利用来自不同模态的信息,在各类方法中,在数据映射时采用神经网络将传统映射方法变为非线性映射,可有效提高算法的有效性。神经网络通过模拟生物神经网络,以信号处理技术为基础建立某种简易模型,经常被用来建模输入与输出间的关系。神经网络作为一种模型,由大量的节点(或称作感知器或神经元)相互连接构成,节点之间的连接表示一种加权值(或称为权重)。节点的连接方式、权重、激励函数的不同都将导致神经网络获得不同的输出值。神经网络的本质是对函数的理想化逼近,或是对某种逻辑策略的表达,由于人工神经网络的独特优势,其在 20 世纪 80 年代开始高速发展<sup>[29]</sup>。本文引入自编码器网络作为提出的方法的基础架构,这种典型的无监督非线性的表示学习方法已被研究人员广泛采用,它可有效挖掘阿尔兹海默症的多模态数据信息,在隐空间中学习到具有较好表征能力的共同表示,并用于之后的诊断任务。

## 3 本文方法

本节论述了提出的基于自编码器的多模态表示学习方

法,并具体介绍了所涉及的关键技术以及优化方法。图1展示了本文方法的整体框架图。首先对原始的多模态阿尔兹海默症数据进行预处理和特征提取,其中预处理过程包含特征选择等操作;然后将提取到的各个模态的特征进行拼接,得到



基于自编码器的多模态表示学习

诊断

图1 提出的方法框架图

Fig. 1 Framework of the proposed method

### 3.1 符号定义

首先定义本文用到的各种变量及其符号表示。本文使用来自两种数据源的阿尔兹海默症数据,分别为MRI数据和PET数据,它们的特征向量分别用 $\mathbf{X} = \{\mathbf{X}^{(1)}, \mathbf{X}^{(2)}\}$ 表示,其中 $\mathbf{X}$ 为所有受试者的集合。假设集合 $\mathbf{X} \in \mathbb{R}^{N \times d}$ 有 $N$ 个受试者,每个受试者均由 $d$ 维的特征向量表示。表1给出了本文的符号定义。

表1 本文用到的符号表示

Table 1 Symbolic representation of main notations

符号	意义
$\mathbf{X}^{(1)} \in \mathbb{R}^{N \times d_1}$	MRI数据的特征矩阵
$\mathbf{X}^{(2)} \in \mathbb{R}^{N \times d_2}$	PET数据的特征矩阵
$\mathbf{Z}^0 \in \mathbb{R}^{N \times (d_1 + d_2)}$	两个模态拼接后的特征向量,即自编码器的输入
$\mathbf{Z}^l \in \mathbb{R}^{N \times h}$	自编码器的隐表示,即最终期望得到的共同表示
$\mathbf{Z}^l \in \mathbb{R}^{N \times d}$	自编码器的解码器输出,用于计算重构损失
$\Theta$	自编码器的参数
$\lambda > 0$	平衡各模态贡献的超参数

### 3.2 基于自编码器的多模态表示学习框架

自编码器网络由编码器和解码器两部分组成,是一种能够将原始数据映射为隐表示,并能通过解码器重构原始数据的人工神经网络。自编码器网络属于典型的无监督表示学习方法,它可作为强大的特征检测器降低特征向量维度和去除数据噪声。此外,它还可作为生成模型随机生成与训练数据类似的数据表示。假设输入为 $\mathbf{X}$ ,则一个 $L$ 层的自编码器网络可以表示为 $f(\mathbf{X}; \Theta)$ ,其中 $\Theta = \{\mathbf{W}^l, \mathbf{b}^l\}_{l=1}^L$ 为所有层的参数。进一步设定前 $L/2$ 层为编码器,特征维数逐层下降,可将输入数据映射至隐空间;后 $L/2$ 层为解码器,特征维数逐层上升,可重构输入数据。假设 $\mathbf{x}_i^0$ 为输入数据的第 $i$ 个样本,则自编码器的输出为:

$$\mathbf{x}_i^l = \text{Relu}(\mathbf{W}^l \mathbf{z}_i^{l-1} + \mathbf{b}^l) \quad (1)$$

其中, $l=1,2,\dots,L$ , $\mathbf{W}^l$ 和 $\mathbf{b}^l$ 分别代表第 $l$ 层的网络权重和偏置, $\text{Relu}(\ast)$ 表示Relu非线性激活函数(本文以Rule函数为例,在应用中可替换为其他非线性激活函数)。根据上述公式, $\mathbf{z}_i^{l/2}$ 为编码器的输出即原始特征向量的隐表示, $\mathbf{x}_i^l$ 为解码

初级的共同表示 $\mathbf{Z}^0$ ,作为自编码器网络的输入;编码器可将输入数据映射至隐空间,解码器可将隐空间中的数据重构至每个模态;得到的隐表示 $\mathbf{Z}^{l/2}$ 即为具有较好表征能力的最终共同表示。

器的输出。为了获得更具有表征能力且更为低噪的隐表示,需要最小化以下重构损失:

$$\min_{\Theta} \sum_{i=1}^N \|\mathbf{x}_i^0 - \mathbf{x}_i^l\|_F^2 \quad (2)$$

传统的自编码器网络只处理单一模态的数据,本文将其延伸至多模态数据的表示学习用于阿尔兹海默症的诊断。在以往的方法中,自编码器网络作为无监督的特征提取器也被应用于多模态表示学习(如DCCA<sup>[30]</sup>等方法)。不同的是,本文将自编码器网络作为多模态特征融合的工具,可以隐式地学习融合后的表示,即通过将多模态数据初步融合的共同表示作为输入,解码器重构各个模态数据的方式学习最终的共同表示。

首先将两个模态的特征向量 $\mathbf{X}^{(1)}$ 和 $\mathbf{X}^{(2)}$ 拼接为一个新的初级的共同表示 $\mathbf{Z}^0 \in \mathbb{R}^{N \times (d_1 + d_2)}$ ,将其作为自编码器网络的输入可以得到解码器的重构表示 $\mathbf{Z}^l$ 。为了使学习到的隐表示 $\mathbf{Z}^{l/2}$ 能够尽可能地融合多个模态的信息,可约束重构表示 $\mathbf{Z}^l$ ,使其尽可能地接近原始的各个模态的数据表示。则需要优化的重构损失可表示为:

$$\min_{\Theta} \sum_{i=1}^N \|\mathbf{x}_i^{(1,0)} - \mathbf{z}_i^l\|_F^2 + \lambda \sum_{i=1}^N \|\mathbf{x}_i^{(2,0)} - \mathbf{z}_i^l\|_F^2 \quad (3)$$

其中, $\mathbf{x}_i^{(1,0)}$ 和 $\mathbf{x}_i^{(2,0)}$ 分别代表第 $i$ 个受试者的两个模态的原始特征向量, $\mathbf{z}_i^l$ 为自编码器的解码器输出, $\lambda$ 为平衡目标函数中各项的超参数,也可视为平衡各个模态贡献的权重。具体来说,有两种方法可以确定 $\lambda$ 的取值,一种方法为通过先验知识确定取值,如本文的阿尔兹海默症数据,根据研究人员对于此类疾病的前期研究,可以认为两种模态对于诊断任务来说具有同样的贡献,因此将 $\lambda$ 设置为1;另一种方法适用于未知数据,可以先采用K-Means算法对每个模态的数据进行聚类,然后计算它们的模块度的值,通过模块度来分配 $\lambda$ 的取值。

通过优化以上重构损失,可以得到更加低维且表征能力更强的多模态数据的共同表示 $\mathbf{Z}^{l/2}$ 。由于 $\mathbf{Z}^{l/2}$ 由多模态数据拼接后非线性编码而来,而且通过解码器可重构各个模态信息,因此它具有各个模态的完备信息并且具有低噪声和表征能力强等特点。

### 3.3 优化

更新自编码器网络的参数,需要最小化以下损失函数:

$$\mathcal{L}(\Theta) = \sum_{i=1}^N \|x_i^{(1,0)} - z_i^l\|^2 + \lambda \sum_{i=1}^N \|x_i^{(2,0)} - z_i^l\|^2 \quad (4)$$

分别计算自编码器网络的权重  $\mathbf{W}$  和偏置  $\mathbf{b}$  的梯度:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{W}^l} = \Delta^l (\mathbf{z}_i^l)^T$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{b}^l} = \Delta^l \quad (5)$$

其中,  $\Delta^l$  具体表示为:

$$\Delta^l = \begin{cases} -(x_i^{(1,0)} + x_i^{(2,0)} - z_i^l) \odot \text{Relu}'(k_i^l), & l=L \\ (\mathbf{W}^{l+1})^T \Delta^{l+1} \odot \text{Relu}'(k_i^l), & \text{otherwise} \end{cases} \quad (6)$$

其中,  $\text{Relu}'(*)$  表示 Relu 非线性激活函数的导数,  $\odot$  表示矩阵的点乘,  $k_i^l = \mathbf{W}^l z_i^{l-1} + \mathbf{b}^l$ . 给定一个学习率  $\eta > 0$ , 通过随机梯度下降法, 可以对自编码器网络的每一层参数进行优化, 具体为:

$$\begin{cases} \mathbf{W}^l = \mathbf{W}^l - \eta \frac{\partial \mathcal{L}}{\partial \mathbf{W}^l} \\ \mathbf{b}^l = \mathbf{b}^l - \eta \frac{\partial \mathcal{L}}{\partial \mathbf{b}^l} \end{cases} \quad (7)$$

通过多次迭代优化, 最终学习到最优的网络参数, 进而可得到多模态数据的共同表示  $\mathbf{Z}^{\frac{1}{2}}$ .

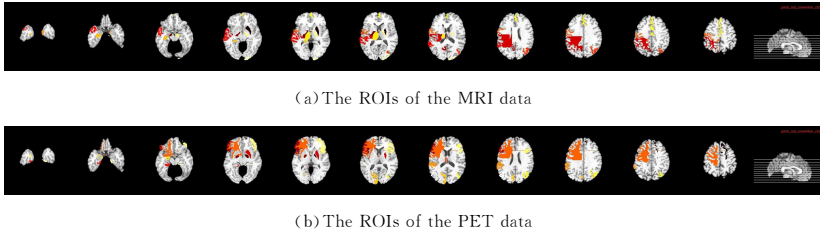


图2 被选择的感兴趣区域可视化

Fig. 2 Visualization of selected ROIs

### 4.2 对比方法

本文将提出的用于阿尔兹海默症诊断的多模态表示学习方法与以下几种方法进行了对比。

Single: 原始的单一模态数据。

FeaCon: 特征拼接方法, 将多模态特征简单拼接作为共同表示。

CCA<sup>[8]</sup>: 典型相关分析, 将多模态数据映射至同一特征空间, 同时约束模态间相关性。

DCCA<sup>[10]</sup>: 基于深度神经网络的 CCA, 将 CCA 作为神经网络的目标函数寻求最佳的数据映射, 需要迭代优化。

DCCA<sup>E</sup><sup>[29]</sup>: 基于自编码器的 CCA, 将 CCA 作为编码器网络的目标函数约束潜在表示以寻求最佳的数据映射。

MDcR<sup>[31]</sup>: 一种多视图学习方法, 采用核方法将多模态数据映射至低维空间, 同时约束模态间的独立性。

K-Com<sup>[6]</sup>: 多模态的阿尔兹海默症诊断算法, 在核空间融合多模态信息, 然后采用 SVM 作为分类器进行诊断。

### 4.3 实验设置

本文进行了两大类的对比实验以验证所提方法的有效

## 4 实验

本节基于真实的阿尔兹海默症数据集, 将本文提出的用于阿尔兹海默症诊断的多模态表示学习方法与其他方法进行了对比, 分析了实验结果并展示了该算法实现的关键细节。

### 4.1 数据集和预处理

本文选取两种模态的阿尔兹海默症数据进行实验, 分别为 MRI 数据和 PET 数据。本文所用数据均从阿尔兹海默症的公开数据集 ADNI<sup>1)</sup> 获取, 共包含 85 名阿尔兹海默症患者 (AD)、185 名早期认知障碍患者 (MCI) 以及 90 名健康受试者 (NC)。

为了尽可能好地提取两个模态数据的特征, 本文采用人工标注的方式选取感兴趣区域, 并提取特征向量。原始的数据经过平均、空间对齐、标准体素大小的插值、强度归一化和一般分辨率平滑处理后, 对每个影像数据均提取了 93 维基于感兴趣区域的特征向量, 然后用卡方检验算法对其进行特征选择。最终在应用中, 选取了前 40 维最具表征能力的向量, 图 2 在标准模板中展示了其中最具有代表性的 5 个感兴趣区域。

性。一方面对于学习到的共同表示进行了多种类型的分类实验, 分别为 AD/MCI/NC 的三分类实验以及 3 种二分类实验 (AD/NC, MCI/NC, AD/MCI)。对于每个方法均采用 LIBSVM<sup>2)</sup> 工具包中参数  $C=1$  的支持向量机 (SVM) 作为分类器。同时本文采用 30 次的 10 折交叉验证以减小随机误差, 对于数据集中所有样本, 首先将其随机拆分成 10 个子集, 其中 9 个子集作为训练样本, 剩余 1 个子集作为测试样本。另一方面, 为了进一步评价方法的有效性, 本文对学习到的共同表示进行了不同类型的聚类实验, 对各个方法学到的表示均采用 K-Means 算法进行聚类。同时在本文中, 两个模态被视为具有同等重要性, 因此设置超参数  $\lambda=1$ , 学习率  $\eta$  为 0.001。对于上述的每个实验, 均引入准确率和标准差作为评价指标。需要详细指出的是, 聚类准确率的计算方法如下:

$$ACC = \frac{\sum_{i=1}^n \Gamma(s_i, \text{map}(r_i))}{n} \quad (8)$$

其中,  $r_i$  和  $s_i$  分别表示第  $i$  个样本的簇标签和真实标签。如果  $x=y$ , 则  $\Gamma(x, y)=1$ , 否则  $\Gamma(x, y)=0$ 。  $\text{map}(\ast)$  为置换映射

<sup>1)</sup> <http://adni.loni.usc.edu/>.

<sup>2)</sup> <https://www.csie.ntu.edu.tw/~cjlin/libsvm/>.

函数,可将簇标签映射至类标签。最优的映射可以通过 Kuhn-Munkres 算法得到。

#### 4.4 实验分析

图 3 给出了聚类实验的所有结果,其中红色柱体代表本文提出的方法,蓝色柱体代表其他对比方法,柱体顶端的竖线表示标准差,可反映准确率的波动大小。聚类实验的好坏从一定程度反映了特征向量的表征能力,质量更好的特征能取

得更好的聚类效果。通过对比其他方法,可观察到本文方法具有更高的聚类准确率,验证了其所学到的共同表示具有更好的表征能力。尤其在 AD/MCI 的聚类实验中,本文方法得到了相对更高和更稳定的准确率,体现了算法的有效性和稳定性。

需要指出的是,由于 K-Com 不是表示学习的方法,故本文没有对其进行聚类实验。

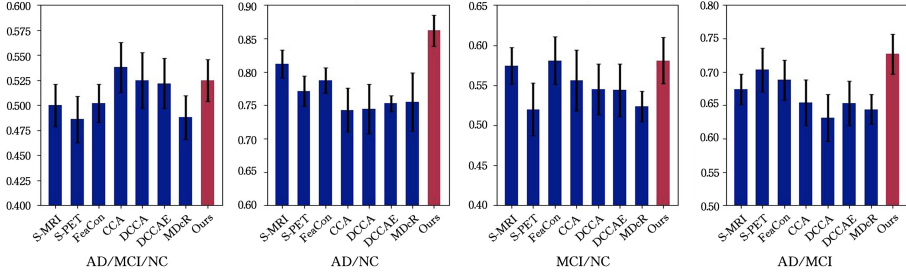


图 3 聚类实验结果(电子版为彩色)

Fig. 3 Clustering results

表 2 列出了分类实验结果,即阿尔兹海默症诊断结果。可以观察到在每个类型的分类实验中,本文提出的方法的诊断效果均好于其他对比方法。特别是观察表中数据可知,在多个类型的分类实验中,用单一模态数据进行诊断并不能达到最好的诊断效果。比如在 AD/NC 实验中,MRI 和 PET 的单一模态数据所达到的分类准确率分别为 0.864 和 0.828,而其他多模态学习方法大多取得了更好的效果,其中本文提出的方法达到了最高的准确率。与其他多模态学习的方法相比,本文方法能更有效地融合多个模态的信息。例如,基于 CCA 的方法都需要学习特定于不同模态的映射向量,然后将映射后的向量进行拼接或相加,而本文提出的方法直接借助强大的人工神经网络学习融合后的共同表示,这个过程可视为一种隐式的映射,再结合重构损失的约束,使其保持各个模态的共同信息,因此本文方法取得了更好的结果。

表 2 分类实验结果

Table 2 Classification results

Algorithm	AD/MCI/NC	AD/NC	MCI/NC	AD/MCI
Single-MRI	0.511±0.075	0.864±0.077	0.708±0.070	0.683±0.077
Single-PET	0.550±0.061	0.828±0.087	0.660±0.071	0.745±0.065
FeaCon	0.525±0.074	0.758±0.081	0.677±0.083	0.723±0.088
CCA	0.538±0.075	0.740±0.088	0.665±0.093	0.692±0.090
DCCA	0.525±0.080	0.738±0.078	0.679±0.074	0.688±0.077
DCCA	0.522±0.092	0.880±0.054	0.665±0.047	0.722±0.089
MDcR	0.513±0.047	0.868±0.093	0.672±0.098	0.685±0.078
K-Com	0.533±0.057	0.813±0.077	0.667±0.067	0.707±0.071
Ours	0.566±0.077	0.889±0.060	0.754±0.081	0.768±0.069

为了更准确地设定实验所需参数,本文对共同表示的维度对实验的影响进行了研究。图 4 给出了在 AD/NC 的分类实验中,不同的共同表示的维度与分类准确率之间的关系。可知,当共同表示的特征向量为 20 维时可达到最好的分类效果,因此将其设定为 20 维。另外,为了验证特征选择的有效性,本文研究了特征选择前后的单一模态数据在 AD/NC 聚类实验中的表现。由图 5 可明显观察到通过特征选择可得到更好的诊断效果。

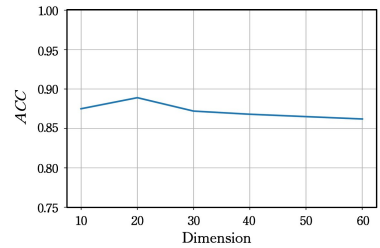


图 4 共同表示的维度与分类准确率的关系曲线

Fig. 4 Relationship between dimensions of common representation and classification accuracy

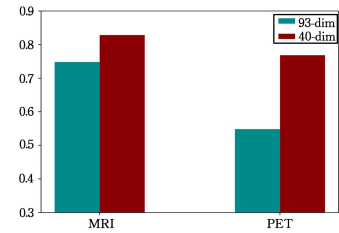


图 5 特征选择前后的单一模态数据聚类表现

Fig. 5 Single modal data clustering performance before and after feature selection

本文进一步展示了提出的算法的收敛曲线,如图 6 所示,其中横坐标为迭代次数,纵坐标为损失值。可以看到,本文算法的损失函数在初始时快速下降,并在第 20 次左右达到收敛,体现出较好的收敛性。

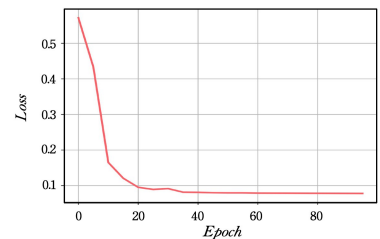


图 6 收敛曲线

Fig. 6 Convergence curve

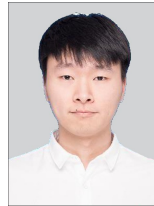
**结束语** 本文对用于阿尔茨海默症诊断的多模态表示学习方法进行了研究,并提出了一种基于自编码器网络的多模态数据表示学习框架。首先对原始的多模态数据进行预处理和特征提取,获得各个受试者的特征向量;然后将多模态数据进行融合,得到初级的共同表示;进而借助自编码器网络将其编码至隐空间,解码器可将隐空间的表示重构至各个模态;最终在隐空间可学习到融合了多模态信息的共同表示。在此共同表示的基础上,引入 SVM 分类器得到诊断结果。本文提出的方法在国际公开数据集 ADNI 上进行了实验,取得了较好的实验结果。在未来工作中,可考虑更为实际的应用需求,比如阿尔茨海默症的数据往往是不完备的,并不是所有受试者均能提供所有模态的数据,因此进一步设计可用于有数据缺失的多模态阿尔茨海默症的诊断算法具有重要意义。

### 参 考 文 献

- [1] WEINER M W, VEITCH D P, AISEN P S, et al. Recent publications from the alzheimer's disease neuroimaging initiative: reviewing progress toward improved ad clinical trials[J]. *Alzheimer's & Dementia*, 2017, 13(4): 1-85.
- [2] ZHANG J, YUE G, GAO Y, et al. Detecting anatomical landmarks for fast Alzheimer's disease diagnosis[J]. *IEEE Transactions on Medical Imaging*, 2016, 35(12): 2524-2533.
- [3] LIAN C, LIU M, ZHANG J, et al. Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's Disease diagnosis using structural MRI[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(4): 880-893.
- [4] LIN W, YUAN J, FENG C, et al. Computer-Aided Diagnosis of Alzheimer's Disease Based on Extreme Learning Machine[J]. *Chinese Journal of Biomedical Engineering*, 2020, 39(3): 288-294.
- [5] ZENG A, ZOU C, PAN D. Diagnosis of Alzheimer's disease based on 3D convolutional neural network-regions of interest [J]. *Journal of Biomedical Engineering Research*, 2020, 39(2): 133-138, 144.
- [6] ZHANG D, WANG Y, ZHOU L, et al. Multimodal classification of alzheimer's disease and mild cognitive impairment[J]. *Neuroimage*, 2011, 55(3): 856-867.
- [7] KUMAR A, RAI P, DAUME H. Co-regularized multi-view spectral clustering[C]// *Advances in Neural Information Processing Systems*. 2011; 1413-1421.
- [8] HOTELLING H. Relations between two sets of variates[M]// *Breakthroughs in Statistics*. Springer, New York, NY, 1992: 162-190.
- [9] AKAHO S. A kernel method for canonical correlation analysis [J]. arXiv preprint cs/0609071, 2006.
- [10] ANDREW G, ARORA R, BILMES J, et al. Deep canonical correlation analysis [C] // *International Conference on Machine Learning*. 2013; 1247-1255.
- [11] ZHU X, SUK H I, SHEN D. Multi-modality canonical feature selection for alzheimer's disease diagnosis[C]// *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2014: 162-169.
- [12] ZHU X, SUK H I, LEE S W, et al. Canonical feature selection for joint regression and multi-class identification in alzheimer's disease diagnosis[J]. *Brain Imaging and Behavior*, 2016, 10(3): 818-828.
- [13] ZIEN A, ONG C S. Multiclass multiple kernel learning[C]// *International Conference on Machine Learning*. 2007; 1191-1198.
- [14] LIU X, ZHU X, LI M, et al. Multiple kernel k-means with incomplete kernels[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.
- [15] ZHOU T, LIU M, THUNG K H, et al. Latent representation learning for Alzheimer's disease diagnosis with incomplete multi-modality neuroimaging and genetic data[J]. *IEEE Transactions on Medical Imaging*, 2019, 38(10): 2411-2422.
- [16] SHI Y, SUK H I, GAO Y, et al. Leveraging Coupled Interaction for Multimodal Alzheimer's Disease Diagnosis[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2019, 31(1): 186-200.
- [17] LI C, FAN C. Classification and prediction of Alzheimer's disease based on machine learning[J]. *Chinese Journal of Medical Physics*, 2020, 37(3): 379-384.
- [18] LIU M, ZHANG J, ADELI E, et al. Joint classification and regression via deep multi-task multi-channel learning for Alzheimer's disease diagnosis[J]. *IEEE Transactions on Biomedical Engineering*, 2018, 66(5): 1195-1206.
- [19] PAN Z J, LIU N, ZHANG W, et al. MTHAM: Multitask Disease Progression Modeling Based on Hierarchical Attention Mechanism[J]. *Computer Science*, 2020, 47(9): 185-189.
- [20] VON LUXBURG U. A tutorial on spectral clustering[J]. *Statistics and Computing*, 2007, 17(4): 395-416.
- [21] STEINWART I. Fully adaptive density-based clustering[J]. *The Annals of Statistics*, 2015, 43(5): 2132-2167.
- [22] ZHANG C, ADELI E, ZHOU T, et al. Multi-layer multi-view classification for alzheimer's disease diagnosis[C]// *AAAI Conference on Artificial Intelligence*. 2018.
- [23] SUN S. A survey of multi-view machine learning[J]. *Neural Computing and Applications*, 2013, 23(7/8): 2031-2038.
- [24] SA V R D. Spectral clustering with two views [C]// *ICML Workshop on Learning with Multiple Views*. 2005: 20-27.
- [25] XU J, HAN J, NIE F. Discriminatively embedded k-means for multi-view clustering[C]// *IEEE Conference on Computer Vision and Pattern Recognition*. 2016: 5356-5364.
- [26] WANG H, WENG C, YUAN J. Multi-feature spectral clustering with minimax optimization[C]// *IEEE Conference on Computer Vision and Pattern Recognition*. 2014: 4106-4113.
- [27] ELHAMIFAR E, VIDAL R. Sparse subspace clustering: algorithm, theory, and applications[J]. *IEEE Transactions on Pat-*

tern Analysis and Machine Intelligence, 2013, 35(11):2765-2781.

- [28] HU H, LIN Z, FENG J, et al. Smooth representation clustering [C]//IEEE Conference on Computer Vision and Pattern Recognition. 2014:3834-3841.
- [29] JIAO L, YANG S, LIU F, et al. Seventy Years beyond Neural Networks: Retrospect and Prospect[J]. Chinese Journal of Computers, 2016, 8(39):1697-1716.
- [30] WANG W, ARORA R, LIVESCU K, et al. On deep multi-view representation learning[C]// International Conference on Machine Learning. 2015:1083-1092.
- [31] ZHANG C, FU H, HU Q, et al. Flexible multi-view dimensionality co-reduction[J]. IEEE Transactions on Image Processing, 2016, 26(2):648-659.



**FAN Lian-xi**, born in 1995, postgraduate. His main research interests include medical data analysis, multi-view learning, and graph representation learning.



**LIU Yan-bei**, born in 1985, Ph.D, lecturer. His main research interests include network embedding, clustering analysis, machine learning and data mining.