

Parkinson's Disease Classification and Clinical Score Regression via United Embedding and Sparse Learning from Longitudinal Data

Huang Z, Lei H, Chen G

摘要

帕金森病（PD）是一种不可逆的退行性疾病，会影响人类的运行系统，帕金森病的早期分类对于研究患者退行性过程以及给予治疗至关重要。本文提出了一种基于自适应无监督特征选择的新方法，可以用于联合分类和预测阿尔兹海默病的临床得分。该方法同时利用自适应嵌入式学习以及稀疏回归技术，用来动态更新相似度矩阵，以获取数据的内在结构属性。此外，提出了一种有效的迭代优化算法来解决这个问题，它可以有效地提高回归和分类的性能。我们在帕金森病标记物 PPMI 数据集上进行了大量的实验，实验结果表明，所提出的方法优于对照组方法。

关键词：临床评分；嵌入式学习；稀疏回归；帕金森病

1 引言

帕金森病（PD）是一种常见的神经退行性疾病，多发于老年人，其可能会导致致死事件的发生。患有 PD 的患者表现出四个主要的运动症状：肌肉僵硬、静态震颤、不稳定的姿势和缓慢运动，以及伴随症状（如抑郁症、嗅觉障碍和认知障碍）。病情的恶化可能会导致吞咽困难，进而导致食物被吸入肺部，从而引起严重的肺炎或肺部疾病。尽管早期 PD 的分类对于减缓疾病的进程至关重要，但由于症状的复杂性，以及常规的多巴胺扫描（SWEDD）发现约 15% 的 PD 患者在扫描水平上正常，因此确定症状是否表明或模拟 PD 的任务仍然具有挑战性。为了改善 PD 的分类性能，多模态数据被用于 PD 的分类，例如磁共振成像（MRI）的灰质（GM）、扩散张量成像（DTI）的第一个特征值（L1）和第一个特征向量（V1）。深度学习在医学影像领域也被广泛应用，但是在少量受试者上很难获得良好的泛化性能。因此，本研究建议结合多模态数据，包括 GM、L1 和 V1，以实现 PD 分类，并采取有效的预处理技术以减轻过拟合问题。同时，大多数现有的方法使用规范来控制稀疏性。这种规范不能根据不同的情况实现自适应的稀疏性。大多数现有的研究只进行 PD 分类。很少有研究考虑到临床评分回归的另一个基本任务。由于临床评分回归（如抑郁、睡眠、嗅觉和认知评分）可以帮助医生对疾病进行分期和治疗，这两项任务必须同时进行。此外，在大多数现有的研究中，分类和回归只基于基线数据进行，而纵向数据（即多时间点数据）则被忽略了。由于疾病的持续加重，必须学习可靠的分类和预测模型，以满足多时间点。特征选择是发现与疾病相关的特征的一种有效方法，它可以通过监督或无监督的方法来完成。多任务学习是一种常用的监督方法，它结合了多个相关的任务来共同选择特征，并且能够更好地发挥作用。但是，现有的多任务方法存在两个主要限制：一是选择的特征往往与多任务目标呈线性关系，忽略了学习数据中固有的结构信息；二是这些多任务方法需要额外的分数和标签信息来学习模型。无监督方法更注重学习数据中固有的结构信息。大多数无监督的方法是基

于过滤或嵌入开发的，嵌入方法在很多方面都有优势。然而，现有的嵌入方法存在三个主要方面需要改进：首先，它们计算主体间的相似性矩阵并选择特征，但原始高维空间中的准确数据有噪声和冗余，降低了相似性矩阵的准确性；其次，在计算邻居图时，传统方法产生的主体之间的相似性矩阵很少代表一个适当的邻居分布；第三，许多嵌入方法需要考虑特征之间的相似性。近年来，基于深度学习的精神病学研究取得了长足的进步。然而，大多数现有的深度学习模型忽略了稀疏性，这可能会导致过拟合。为了解决这个问题，研究者们提出了各种稀疏深度学习模型。然而，大多数现有的方法使用规范来控制稀疏性，而不能根据不同的情况实现自适应的稀疏性。此外，大多数现有的研究只进行 PD 分类，而很少有研究考虑到临床评分回归的另一个基本任务。由于临床评分回归（如抑郁、睡眠、嗅觉和认知评分）可以帮助医生对疾病进行分期和治疗，因此，这两项任务必须同时进行。此外，大多数现有的研究中，分类和回归只基于基线数据进行，而纵向数据（即多时间点数据）则被忽略了。由于疾病的持续加重，必须学习可靠的分类和预测模型，以满足多时间点的需求。因此，研究者们需要开发一种新的稀疏深度学习模型，既能控制稀疏性，又能实现自适应的稀疏性，同时考虑到 PD 分类和临床评分回归，以及基线和纵向数据。

2 相关工作

近年来，基于机器学习的多模态多任务学习 (M3T)^[1]、稀疏多模态学习 (MMSL)^[2]、灵活的流形嵌入 (FME)^[3]以及互相嵌入学习和稀疏回归 (JELSR)^[4]等方法，都可以用来提高阿尔茨海默病 (PD) 的分类和回归预测性能。M3T 方法学习了一个特征选择模型，从每一种模式中获得多个任务的标准相关特征。MMSL 方法基于统一的多任务特征选择函数，考虑响应矩阵中行和列之间的差异相似性，同时对 PD 进行分类和回归预测。RRDL 方法利用多模态数据来提高 PD 的分类和回归预测性能，通过关系正则化和判别性学习，从多个关系和模式中进行联合学习，以选择用于 PD 分类和预测的判别性特征。MCFS 方法首先计算近邻图，然后选择最能呈现集群信息的判别性特征。FME 方法是一个被许多无监督和半监督嵌入方法所利用的通用模型，用于降低特征维度。RSFS 方法同时利用 FME 和规范来稳健地选择判别性特征。JELSR 方法通过嵌入学习和稀疏回归进行特征选择。然而，这些现有的方法也存在一些限制。有监督的多任务方法忽略了对数据中固有的结构信息的学习。以前的无监督嵌入方法也可能由于原始特征空间的噪声而计算出不准确的相似度矩阵。

3 本文方法

3.1 本文方法概述

首先，我们从 GM、L1 和 V1 中提取特征并直接串联。然后使用本文提出的方法进行特征选择。最后，利用支持向量分类 (SVC) 和支持向量回归 (SVR) 模型对纵向多模态数据进行分类和回归预测。

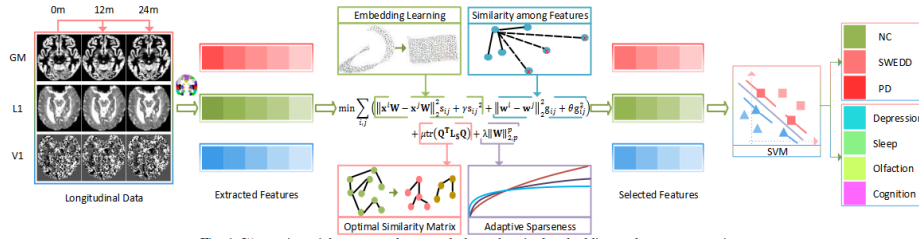


图 1: 方法示意图

3.2 特征提取模块

首先，利用中心质量法进行前段—后段重定向，然后使用统计参数映射（SPM8）工具按照一般流程进行预处理。接着，校正头部运动和几何失真，再使用图割法进行头骨剥离。接着，所有 MRI 图像都与国际脑图谱模板进行注册，以获取相关解剖标签的坐标。然后，将相应的解剖区域分割为 GM，白质和脑脊液（CSF）。同时，这些图像被重新采样为等向性分辨率 1.5mm。使用 60mm 的半高宽高斯核对这些图像进行空间平滑处理，以抑制噪声的干扰。最后，使用脑成像数据处理和分析工具注册自动解剖标记（AAL）图谱，并利用注册的 AAL 图谱从 GM 中提取 116 维特征。每个受试者的数据由 65 个原始格式的图像组成，其中 bo 图像不激活扩散梯度，而其他 64 个图像具有不同的梯度方向。首先，使用 FMRIB 软件库（FSL）校正 bO 图像失真。其次，使用 FSL 工具的 bet 命令来生成一个与校正的 bo 图像对应的掩码图像。然后，使用 dcm2nii 工具将 65 个图像转换为 4D 图像，并生成 b-vector 文件和 b-value 文件，分别指示每个梯度方向及其标量值。第四，使用 FSL 工具的 correct 命令来校正 4D 图像上的涡流失真。第五，将 b-value 文件，b-vector 文件，掩码图像和校正后的 4D 图像导入 FSL 工具的 drift 命令，以计算 L1 和 V1 图像。最后，使用 AAL 图谱计算 L1 和 V1 的每个区域的平均组织密度，然后获得它们的 116 维特征。

3.3 损失函数定义

在我们提出的方法中，设表示 n 个主题和 d 个特征的数据，表示主题相似矩阵，根据既往经验我们可以用以下公式来计算相似性：

$$\min \sum_{i,j} \left(\| \mathbf{x}^i - \mathbf{x}^j \|_2^2 s_{ij} + \mu s_{ij}^2 \right) \\ \text{s.t. } \mathbf{s}^i \mathbf{1} = 1, 0 \leq s_{ij} \leq 1$$

其中 \mathbf{x}^i 与 \mathbf{x}^j 表示 \mathbf{x} 的第 i 个和第 j 个主题。 μ 是一个正则化参数。为了弥补主题相似矩阵很少具有适当的邻居分布的缺点，可以通过引入拉普拉斯矩阵使得矩阵的秩等于 $n-r$ ，如下公式 2：

$$\min \sum_{i,j} \left(\| \mathbf{x}^i - \mathbf{x}^j \|_2^2 s_{ij} + \mu s_{ij}^2 \right) \\ \text{s.t. } \mathbf{s}^i \mathbf{1} = 1, 0 \leq s_{ij} \leq 1, \text{rank}(\mathbf{L}_s) = n - r$$

其中 \mathbf{L}_s 等于 \mathbf{D} 减相似矩阵 \mathbf{S} 与相似矩阵 \mathbf{S} 的转制矩阵的平均值，其中 \mathbf{D} 为度矩阵。由于秩 $\mathbf{L}_s = n-r$ 依赖于主题相似矩阵 \mathbf{S} ，参考 Ky Fan 定理，将公式再次优化为：

$$\sum_{i=1}^r \psi_i(\mathbf{L}) = \min \text{tr}(\mathbf{Q}^T \mathbf{L}_s \mathbf{Q}), \\ \text{s.t. } \mathbf{Q} \in \mathbb{R}^{n \times r}, \mathbf{Q}^T \mathbf{Q} = \mathbf{E}$$

上述公式我们也可以写为：

$$\begin{aligned} \min \sum_{i,j} \left(\|\mathbf{x}^i - \mathbf{x}^j\|_2^2 s_{ij} + \mu s_{ij}^2 \right) + \sigma \text{tr}(\mathbf{Q}^T \mathbf{L}_S \mathbf{Q}) \\ \text{s.t. } \mathbf{s}^i \mathbf{1} = 1, 0 \leq s_{ij} \leq 1, \mathbf{Q} \in \mathbb{R}^{n \times r}, \mathbf{Q}^T \mathbf{Q} = \mathbf{E} \end{aligned}$$

其中是一个模型参数，可以在每次迭代中减少或者参加。rammer，为了解决原是高维数据有噪音或者冗余。为了解决这个问题，加入自适应稀疏和嵌入学习：

$$\begin{aligned} \min \sum_{i,j} \left(\|\mathbf{x}^i \mathbf{W} - \mathbf{x}^j \mathbf{W}\|_2^2 s_{ij} + \mu s_{ij}^2 \right) + \sigma \text{tr}(\mathbf{Q}^T \mathbf{L}_S \mathbf{Q}) + \lambda \|\mathbf{W}\|_{2,p}^p \\ \text{s.t. } \mathbf{s}^i \mathbf{1} = 1, 0 \leq s_{ij} \leq 1, \mathbf{Q} \in \mathbb{R}^{n \times r}, \mathbf{Q}^T \mathbf{Q} = \mathbf{E}, \mathbf{W}^T \mathbf{W} = \mathbf{E} \end{aligned}$$

其中，表示 m 投影维度的特征权重系数矩阵。为了利用特征之间的关系，提出了一种更新的正则化属于，考虑特征之间的相似性，该公式表示如下：

$$\min \sum_{i,j} \left(\|\mathbf{w}^i - \mathbf{w}^j\|_2^2 g_{ij} + \theta g_{ij}^2 \right), \text{ s.t. } \mathbf{g}^i \mathbf{1} = 1, 0 \leq g_{ij} \leq 1$$

最后我们把这个正则化加入（5）中得到完整的公式表达：

$$\begin{aligned} \min \sum_{i,j} \left(\|\mathbf{x}^i \mathbf{W} - \mathbf{x}^j \mathbf{W}\|_2^2 s_{ij} + \mu s_{ij}^2 + \|\mathbf{w}^i - \mathbf{w}^j\|_2^2 g_{ij} + \theta g_{ij}^2 \right) + \sigma \text{tr}(\mathbf{Q}^T \mathbf{L}_S \mathbf{Q}) + \lambda \|\mathbf{W}\|_{2,p}^p \\ \text{s.t. } \mathbf{s}^i \mathbf{1} = 1, 0 \leq s_{ij} \leq 1, \mathbf{g}^i \mathbf{1} = 1, 0 \leq g_{ij} \leq 1 \\ \mathbf{Q} \in \mathbb{R}^{n \times r}, \mathbf{Q}^T \mathbf{Q} = \mathbf{E}, \mathbf{W}^T \mathbf{W} = \mathbf{E} \end{aligned}$$

4 复现细节

4.1 与已有开源代码对比

使用了以下参考代码

Algorithm 1: Solution

Input: $\mathbf{X} \in \mathbb{R}^{n \times d}, \mathbf{L}_S \in \mathbb{R}^{n \times n}, \mathbf{L}_G \in \mathbb{R}^{d \times d}, \lambda;$

Output: $\mathbf{W} \in \mathbb{R}^{d \times m}$

Initialize $t = 0, \mathbf{Z}(t)$ by an identity matrix;

Repeat:

Under the current, the optimal solution Algorithm 1 is the m eigenvectors corresponding to the

m minimum eigenvalues of $\mathbf{X}^T \mathbf{L}_S \mathbf{X} + \mathbf{L}_G + \lambda \mathbf{Z}$.

Update matrix $\mathbf{Z}(t+1)$.

Until the convergence or stop condition is satisfied.

代码中通过输入矩阵 \mathbf{X} ，拉普拉斯矩阵 \mathbf{L} ，计算了 \mathbf{W} ，也就是权重矩阵，表示 m 个投影维度的特征权重稀疏矩阵。使用了此参考代码计算出来的 \mathbf{W} 可以带入到文中提出的公式，从而将提取出来的特征进行选择，最后将在此步骤提取并选择的特征放入 SVM 中进行三个二分类和临床评分回归预测。复现过程中在训练模型时经常出现拟合的现象，原因可能是原始高维空间中的真实数据出现噪声或冗余，降低了相似度矩阵的准确性或者是忽略了数据内部的结构信息的学习，因此在利用文中方法选择出特征后，加入了 CNN 来再一次进行特征提取。之所以选择 CNN 是因为由于卷积和池化计算的性质，使得图像中的平移部分对于最后的特征向量是没有影响的。从这一角度说，提取到的特征更不容易过拟合。而且由于平移不变性，所以平移字符进行变造是无意义的，省去了再对样本进行变造的过程，并且可以利用不同的卷积、池化和最后输出的特征向量的大小控制整体模型的拟合能力。在过拟合时可以降低特征向量的维数，在欠拟合时可以提高卷积层的输出维数，相比于其他特征提取方法更加灵活。加入之后就形成了 CNN-SVM 的训练模式，过拟合的情况得到了缓解。

4.2 实验环境搭建

主要使用了 MATLAB 进行复现，并在 Colab 中进行模型运行训练

4.3 创新点

在使用文中提出的基于联合嵌入和稀疏回归的方法提取并选择特征后，加入了 CNN 的结构，对已经选择过的特征再一次进行提取，最后在放入到 SVM 中进行二分类和临床评分回归。使用 CNN 可以利用其平移不变性，位移不变性，尺度不变性在保留原有特征的同时进行降维，减少过拟合风险，在欠拟合时也可以提高输出维数，使模型更具有泛用性。

5 实验结果分析

实现结果如图所示。通过文中方法提取并选择的特征，其中一个方法是直接放入到 SVM，而另一种方法则是先用 CNN 再一次进行特征提取然后在放入到 SVM。可以发现在各种指标中，大部分情况下 CNN-SVM 都要优于 SVM，但两种结果仍然达不到原文所得到的指标。

Time	Method	ACC(%)	SEN(%)	SPEC(%)	PREC(%)	UAR(%)	F1-score (%)	AUC
Baseline	CNN-SVM	75.45±7.89	60.05±12.36	79.43±8.25	59.23±11.69	71.35±8.62	66.28±7.86	0.70±0.36
	SVM	74.33±6.88	58.33±16.91	82.10±6.95	65.52±12.11	68.26±9.77	59.26±4.36	0.65±0.52
12m	CNN-SVM	73.85±5.66	68.01±3.34	78.25±9.61	62.42±5.27	72.36±5.66	62.23±6.20	0.72±0.12
	SVM	72.65±9.59	65.18±7.62	76.59±6.78	65.24±12.33	68.56±6.39	60.49±9.35	0.70±0.09
24m	CNN-SVM	90.35±6.28	93.28±10.37	91.62±5.12	73.23±16.46	86.25±13.14	70.39±23.11	0.90±0.23
	SVM	88.26±8.26	88.36±6.23	88.33±6.97	66.85±9.28	80.39±6.41	67.34±16.87	0.87±0.13

图 2: 实验结果示意

Time	Method	ACC(%)	SEN(%)	SPEC(%)	PREC(%)	UAR(%)	F1-score (%)	AUC
Baseline	CNN-SVM	82.69±10.25	95.87±8.05	61.88±19.26	73.42±6.34	75.23±8.77	86.12±7.31	0.71±0.34
	SVM	80.11±9.65	90.16±1.96	59.57±23.68	69.69±10.22	74.42±12.69	87.78±5.13	0.70±0.09
12m	CNN-SVM	90.89±3.21	89.32±6.12	89.96±18.77	91.63±4.05	91.07±5.34	91.95±1.09	0.89±0.33
	SVM	82.28±8.76	80.66±5.39	82.36±8.79	85.77±6.32	87.67±2.36	84.33±5.78	0.83±0.36
24m	CNN-SVM	70.26±1.58	95.36±2.33	56.28±11.33	53.96±20.73	76.32±3.28	64.12±8.36	0.85±0.12
	SVM	65.43±15.99	90.47±3.28	51.27±10.24	48.36±10.23	68.85±10.26	60.67±7.12	0.75±0.31

图 3: 实验结果示意

Time	Method	ACC(%)	SEN(%)	SPEC(%)	PREC(%)	UAR(%)	F1-score (%)	AUC
Baseline	CNN-SVM	78.22±6.33	96.31±2.16	32.69±20.17	78.24±6.13	66.12±9.25	89.21±3.12	0.61±0.13
	SVM	75.12±9.22	91.32±3.12	30.24±26.72	71.36±5.12	59.82±14.36	83.16±4.66	0.58±0.09
12m	CNN-SVM	91.22±4.63	96.68±2.08	80.47±6.75	87.26±8.27	83.28±8.66	92.33±3.31	0.83±0.16
	SVM	85.87±6.28	90.33±3.64	76.13±10.41	81.46±6.15	74.19±9.26	84.13±2.67	0.77±0.33
24m	CNN-SVM	87.28±4.56	92.36±2.44	69.23±17.28	89.63±4.22	80.34±6.82	91.37±3.26	0.86±0.21
	SVM	80.35±6.18	86.39±3.49	63.00±21.38	82.77±6.56	75.86±8.48	88.43±2.19	0.79±0.28

图 4: 实验结果示意

6 总结与展望

本文档通过介绍需要复现文章的引言，相关工作以及文章方法中特征提取模块以及损失函数的同时概述了复现的一些细节，如与开源代码对比，实验环境搭建，创新点，最后通过分析实验结果对实验进行了总结分析。只使用与文中相同的 SVM 达不到文中的指标，使用 CNN-SVM 结构后相比于直接使用 SVM 各项指标有所提升，但仍然达不到原文中的效果，原因可能是网络结构过于简单。未来进一步的研究可以从残差网络或者加深网络进行，因此它更容易优化，并且能够通过增加相当的深度来提高准确率，且其内部的可以跳跃链接的残差块，缓解了在深度神经网络中增加深度带来的梯度消失的问题。

参考文献

- [1] ZHANG D A D N I, Shen D. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease[J]. NeuroImage, 2012, 59(02): 895-907.
- [2] LEI H Z J, Huang Z. Joint detection and clinical score prediction in Parkinson's disease via multi-modal sparse learning[J]. Expert Systems with Applications, 2017, 80: 284-296.
- [3] NIE F T I W H, Xu D. Flexible manifold embedding: A framework for semi-supervised and unsupervised dimension reduction[J]. IEEE Transactions on Image Processing, 2010, 19(07): 1921-1932.
- [4] HOU C L X, Nie F. Joint Embedding Learning and Sparse Regression: A Framework for Unsupervised Feature Selection[J]. TEEE transactions on cybernetics, 2013, 44(06): 793-804.