计 算 机 研 究 与 发 展
Journal of Computer Research and Development
https://crad.ict.ac.cn
63(6): 1548−1562, 2026

基于直觉模糊深度自动编码器的特征选择算法

华启冲　丁卫平　陈悦鹏　周天奕
（南通大学人工智能与计算机学院　江苏南通　226019）
（hqc3246@163.com）

Intuitionistic Fuzzy Deep Autoencoder Algorithm for Feature Selection

Hua Qichong, Ding Weiping, Chen Yuepeng, and Zhou Tianyi
（*School of Artificial Intelligence and Computer Science*, *Nantong University*, *Nantong*, *Jiangsu* 226019）

Abstract　With the continuous development of big data technology, high-dimensional datasets have been increasingly widely applied in various fields. These datasets typically feature high dimensionality and excessive redundancy, which tend to induce model overfitting and increase computational costs. Feature selection helps extracting an informative subset from them, reducing dimensionality and enhancing model interpretability. Although deep neural networks have shown potential in feature extraction, their performance is vulnerable to noise and outliers. To address these challenges, a novel feature selection algorithm based on the intuitionistic fuzzy deep autoencoder (IFDAE) is proposed. First, the algorithm fuses membership and non-membership degree information and applies intuitionistic fuzzy weights to handle uncertainty, thereby suppressing the impact of noise and outliers. Then, it fully learns the highly non-linear latent representations in the deep autoencoder, uses weight transfer to migrate pre-trained knowledge into the target network, and imposes a regularizer on the connection weight matrix between the input layer and the first hidden layer via the structured sparse norm to obtain a structurally sparse weight matrix, thus deriving the weight of each feature. Finally, to verify the effectiveness of the proposed IFDAE algorithm, experiments are conducted on twelve public datasets and one real-world schizophrenia dataset. The results demonstrate that the proposed algorithm exhibits significant superiority in both reconstruction capability and classification performance.

Key words　fuzzy logic; autoencoder; feature selection; high-dimensional data analysis; weight transfer

摘　要　高维数据集在多个领域的应用愈发广泛，其维度高、冗余多的特征易引发模型过拟合并增加计算开销。特征选择有助于从中提取有效特征，降低维度并提升模型的可解释性。尽管深度神经网络在特征提取方面表现出潜力，但在面对噪声和离群值时，其性能易受影响。为了解决这些挑战，提出一种新的基于直觉模糊深度自动编码器的特征选择算法（IFDAE）。该算法首先通过融合隶属度和非隶属度信息，应用直觉模糊权重来处理不确定性，从而抑制噪声和离群值的影响；然后，充分学习深度自动编码器中高度非线性的潜在表示，利用权重迁移将预训练知识迁入目标网络，通过结构化稀疏范数对输入层与首个隐藏层之间的连接权重矩阵施加正则项来获得结构稀疏权值矩阵，从而获得各特征的权值；最后，为了验证IFDAE算法的有效性，在12个公共数据集和1个真实世界的精神分裂症数据集上进行实验。结果表

收稿日期：2025-06-26；**修回日期**：2026-01-19
基金项目：国家重点研发计划项目（2024YFE0202700）；国家自然科学基金重点项目（U2433216）；国家自然科学基金项目（62576178）；江苏省自然科学基金项目（BK20231337）；江苏省研究生实践创新计划项目（SJCX25_2009）
This work was supported by the National Key Research and Development Program of China (2024YFE0202700), the Key Program of the National Natural Science Foundation of China (U2433216), the National Natural Science Foundation of China (62576178), the Natural Science Foundation of Jiangsu Province (BK20231337), and the Practical Innovation Program for Graduate Students in Jiangsu Province (SJCX25_2009).
通信作者：丁卫平（dwp9988@163.com）

明，所提算法在重建能力和分类性能方面具有较高的优越性。

关键词 模糊逻辑；自动编码器；特征选择；高维数据分析；权重迁移

中图法分类号 TP181

DOI: 10.7544/issn1000-1239.202550502 **CSTR:** 32373.14.issn1000-1239.202550502

信息技术的持续进步推动多源异构数据的大规模生成，使得高维数据建模成为研究热点。传统机器学习在该背景下常受维度灾难限制，带来了重大的分析挑战[1-3]。其核心矛盾集中体现于特征空间的冗余性和噪声污染性，这些固有缺陷将直接导致模型计算复杂度呈超线性增长。无用的特征通常会减少与分类算法相关的学习过程，导致数据分析和决策过程的准确性和效率降低[4-5]。各种实际的大数据应用进程可以表述为特征选择问题[6-9]。

特征选择技术通过构建最优特征子集的数学优化模型，依据预设的评价准则实现特征空间的降维重构。其中，基于统计假设检验的过滤式方法、基于启发式搜索策略的包裹式方法以及基于正则化约束的嵌入式方法专注于识别最相关的特征，这些特征对学习任务有显著贡献[10-12]。通过筛选与任务高度相关的特征，目标在于降低噪声干扰和冗余信息，降低过拟合风险，提升模型的可理解性，并优化分类和重构任务性能。基于自编码器的特征选择方法通过构建非线性映射机制，实现了对特征重要性的自适应评估。该方法利用编码器-解码器结构，对所有特征进行加权重构，从而学习特征间的潜在关联模式。近年来，研究者在该领域取得显著进展，提出了多种创新性的自编码器特征选择模型。这些方法基于独特的网络架构，能够学习数据的低维紧凑表示，并在重构过程中捕捉特征间的复杂非线性关系，为特征选择提供了新的思路和实现路径。例如，Han 等人[13]通过融合自编码器与群组 Lasso 技术，在联合优化特征重构误差和稀疏性约束的目标函数中，驱动模型自动识别并保留具有显著判别力的特征子集。Doquet 等人[14]开发 AgnoS，通过将自动编码器与基于松弛变量、权重约束和梯度优化的多层次正则化策略相结合，以有效地恢复原始数据。Abid 等人[15]提出了一种无监督的端到端可微分全局特征选择方法，通过具体选择层和温度退火策略，动态筛选离散特征子集并重建原始数据。Shaham 等人[16]通过图拉普拉斯准则剔除无关特征，并基于自动编码器与 Concrete 层消除特征冗余。在解决特征间的相关性时，Wu 等人[17]提出了特征选择分形自动编码器，它包含一对一分值层和轻量子网络来实现无监督特征选择。Gong 等人[18]将自动编码器与冗余控制相结合，通过双重群组 Lasso 正则化与特征相关性惩罚项，实现无监督特征选择中冗余抑制与紧凑网络联合学习。Zhang 等人[19]通过融合非负约束、正交性约束和稀疏约束，构建可灵活嵌入多样化损失函数与激活函数的统一模型。Ling 等人[20]通过引入自适应权重机制动态抑制异常值对$l_{2,1}$范数自编码器的干扰，同时在特征表示层嵌入k-means 聚类目标函数，增强了所选特征的区分能力。Hassanieh 等人[21]利用深度自动编码器的优势进行特征子集的选择和插补，并且能够内在地学习足够的相关特征子集来表示整个特征空间。

在监督设置中，已经提出了多种模型来改进特征选择。Singh 等人[22]提出了一种面向高维小样本生物数据的深度特征选择网络，它包括一个选择器层和一个有监督的深度神经网络，在训练过程中同时优化特征选择和重建目标，减少重建误差并提高分类性能。Natsumeda 等人[23]通过在分类准确率和选定特征的重建误差之间取得平衡，筛选出对已见和未见故障均有利的特征。Tuan 等人[24]提出了一种基于自编码器的特征排序方法，该方法通过设置不同数量的隐藏单元来训练模型以执行分类任务，并利用连接权重对特征的重要性进行排序。Sun 等人[25]利用具体随机变量优化特征选择过程，并结合信息熵选择最优特征子集，无需依赖类别标签或复杂策略。Ghosh 等人[26]提出了一种稀疏性增强质心编码器进行特征选择，通过 L1 正则化与非线性映射相结合优化特征筛选，并验证了随机优化方法对提升稀疏性的作用。Wang 等人[27]提出了一种基于神经网络和结构化稀疏正则化的非线性特征选择方法，通过捕捉高维数据的非线性结构实现特征选择。

现有研究在基于自编码器的特征选择方法中，尽管通过引入稀疏约束、领域自适应正则化以及动态特征冗余控制等策略，显著提升了模型的判别力与鲁棒性，但这些方法仍存在局限性。首先，多数方法采用随机初始化编码器，特征选择器被迫从零开始学习本应通过编码器参数传递的判别性知识。其

次，有效的特征选择需要在保证重建效果的基础上识别能够增强分类的下游任务性能的特征。然而，现有基于自动编码器的特征选择方法以信息完整性为优化核心，这种设计倾向在追求最大化保留原始数据分布的过程中，模型筛选出的特征子集可能包含大量高重构贡献但低类别可分性的成分。这种倾向损害了这些方法在现实世界应用中的可靠性和有效性。

迁移学习是一种利用先验知识来帮助目标领域的学习过程的技术[28]。参数迁移方法通过充分学习源任务的潜在表示来实现知识复用。在处理数据的不确定性及模糊性问题方面，模糊理论提供了可行的解决路径，模糊隶属度函数是基于从样本到类中心的距离定义的，以解决离群值和噪声的问题[29-30]。直觉模糊隶属度通过引入犹豫度表示与每个隶属度等级相关联的不确定性或犹豫的水平。扩展传统模糊集理论，以更灵活地处理不确定性[31-32]。基于以前的研究[33]，并受到特征迁移和模糊理论优势的启发，本文提出了一种新的基于监督自编码器(autoencoder, AE)的特征选择方法。该方法将模糊理论与自动编码器模型相结合。

总体来说，本研究的主要贡献如下：

1)提出了一种新的非线性特征选择方法，将模糊理论与自动编码器模型结合，有效地处理高维数据的噪声和离群值，增强特征选择过程对分类判别下特征的保留。

2)引入迁移学习中特征迁移概念，学习预训练阶段具有最佳重构效果的编码器参数，将其迁入目标网络，减少模型在新任务上训练开销和引导特征空间的优化，提升模型的重构能力。

3)利用权重迁移的编码器参数进行训练，通过使用$l_{2,p}$范数来提升输入层与隐藏层之间的稀疏连接，更有效地区分相似特征的得分，并降低模型过早收敛至局部最优解的风险。

4)对公共数据集和真实世界的精神分裂症数据集进行了实验评估。结果表明，本文方法在分类准确率和重建误差方面基本优于其他6种特征选择技术。

1 预备知识

1.1 直觉模糊集

传统的模糊集由于其处理不确定性和不确定性的能力而被广泛应用于各种领域，包括控制系统、人工智能、数据分析和语言学等[34]。

论域U中的模糊集$\tilde{A}$通过隶属度函数来表征，隶属度函数可以被定义为

$$\tilde{A}=\{\langle x,\theta_{\tilde{A}}(x)\rangle|x\in U\}\,, \tag{1}$$

其中$\theta_{\tilde{A}}(x)$表示x属于$\tilde{A}$的程度，其范围是$\theta_{\tilde{A}}: U\to[0,1]$。

直觉模糊集(intuitionistic fuzzy set, IFS)[35]的定义是为了克服不确定性的困难，它允许使用当前信息和观察结果对情况进行精确模拟。直觉模糊集通过非隶属度的引入，扩展了传统模糊集，实现了更广泛的不确定性表达。

设U为一个论域，直觉模糊集A被定义为[31]

$$A=\left\{\left\langle x,\theta(x),\bar{\theta}(x),h(x)\right\rangle|x\in U\right\}\,, \tag{2}$$

其中$\theta(x)$表示直觉模糊集A中元素x的隶属度，并且$\theta(x): U\to[0,1]$。$\bar{\theta}(x)$表示非隶属度，其反映了元素x不属于集合A的程度，并且$\bar{\theta}(x): U\to[0,1]$。剩下的$h(x)$表示犹豫度，反映了犹豫水平及其不确定性，其与前两者的数量关系为$h(x)=1-\theta(x)-\bar{\theta}(x)$[36]。对于常规模糊集，对每个$x\in U$设置$h(x)=0$，因此$\bar{\theta}(x)=1-\theta(x)$。

直觉模糊集中的得分函数结合了隶属度、非隶属度和犹豫度，对其进行量化分析和排序。该函数用于衡量直觉模糊集合中元素的优先程度或适用性，从而提供对元素偏好的数值表示。对于集合A中的任意元素x，得分函数可用于评估其相对优势，定义为

$$Z(x)=\theta(x)-\bar{\theta}(x)\ 。 \tag{3}$$

当元素x在直觉模糊集合A中的得分较高时，通常意味着其对集合的影响更为显著。该得分函数能够有效衡量元素对集合的贡献程度，使其成为分析不确定性和模糊关系的重要手段。

1.2 深度自编码器的形式化

给定输入数据$\boldsymbol{X}$，自编码器由编码器$f(\cdot)$和解码器$g(\cdot)$组成，其目标是学习非线性映射，最小化编码器-解码器重构数据与原始数据的差异，定义重构数据为

$$\hat{\boldsymbol{X}}=g(f(\boldsymbol{X};\vartheta_{\text{enc}});\vartheta_{\text{dec}})\,, \tag{4}$$

其中编码器和解码器的参数集分别表示为

$$\vartheta_{\text{enc}}=\left\{\boldsymbol{W}^{(i)},\boldsymbol{b}^{i}\right\}_{i=1}^{L_{\text{enc}}}\ \text{和}\ \vartheta_{\text{dec}}=\left\{\boldsymbol{W}^{(j)},\boldsymbol{b}^{j}\right\}_{j=1}^{L_{\text{dec}}}\,, \tag{5}$$

其中$\boldsymbol{W}^{(i)}$和$\boldsymbol{b}^{(i)}$分别表示编码器第i层的权重矩阵和偏置向量。同样地，$\boldsymbol{W}^{(j)}$和$\boldsymbol{b}^{(j)}$分别表示编码器第j层的权重矩阵和偏置向量；L_{enc}为编码器层数；L_{dec}为解码器层数。在训练时针对重构效果进行优化，目标函

数可进一步表示为

$$\min_{\vartheta_{\text{enc}},\vartheta_{\text{dec}}} \|\boldsymbol{X}-\hat{\boldsymbol{X}}\|_{\text{F}}^2, \tag{6}$$

其中$\|\cdot\|_{\text{F}}$表示 Frobenius 范数，用于度量矩阵之间的欧几里得距离。ϑ_{enc}是编码器中连接层的权重集，ϑ_{dec}是解码器中连接层的权重集。在深度自编码器中，编码器和解码器由多个全连接层构成。输入数据$\boldsymbol{X}$被编码为潜在表示$\boldsymbol{V}=f(\boldsymbol{X},\vartheta_{\text{enc}})$，再通过解码器将该潜在表示$\boldsymbol{V}$还原回原始数据空间。潜在表示$\boldsymbol{V}$在语义上并无直接的可解释性，仅作为中间计算结果以重建输入样本。

1.3 特征选择问题形式化

目标是在原始特征空间中筛选出一个最具信息量的特征子集，使其能够最大程度地保留整体特征分布的信息，具体形式为：

$$\min_{S^k,\vartheta}\|\boldsymbol{X}-F(\boldsymbol{X}_{S^k}-\vartheta)\|_{\text{F}}^2, \tag{7}$$

其中S^k表示从原始特征中选出的k个特征子集，$\boldsymbol{X}_{S^k}\in\mathbb{R}^{N\times k}$代表基于$S^k$选择出的数据子集，即仅包含这$k$个特征的样本矩阵，两者之间的数量关系满足$0<|\boldsymbol{X}_{S^k}|\leqslant|\boldsymbol{X}|$，这说明训练过程中至少会选出部分重要特征（即特征选择不会导致所有特征被舍弃），同时网络无法凭空生成额外特征。函数F作为一个深度学习模型，能够将$\boldsymbol{X}_{S^k}$从$\mathbb{R}^{N\times k}$映射回$\mathbb{R}^{N\times n}$，以尽可能地恢复原始特征空间。已有研究[37]表明，该优化问题属于 NP-难问题，因此，本文提出了一种高效的近似算法以求得该特征选择问题的次优解。

1.4 迁移学习

权重迁移（weight transfer, WT）或权重共享（weight sharing, WS）是深度迁移学习（deep transfer learning, DTL）中常用的一种策略[38-39]。在训练深度自动编码器（deep autoencoder, DAE）网络中得到各层的连接的权重$\boldsymbol{W}_i$和偏置向量$\boldsymbol{b}_i$，在迁移学习的架构中，这些训练好的参数可以直接复制并转移到新的深度自动编码器网络中，表示为

$$\boldsymbol{W}_{\text{nn}}^{(i)}=\boldsymbol{W}_{\text{tn}}^{(i)},\ \boldsymbol{b}_{\text{nn}}^{(i)}=\boldsymbol{b}_{\text{tn}}^{(i)},\ i=1,2,\cdots,m, \tag{8}$$

其中m表示训练完成的深度网络以及新网络的隐藏层数量，$\boldsymbol{W}_{\text{tn}}^{(i)}$和$\boldsymbol{b}_{\text{tn}}^{(i)}$分别表示训练好的深度自编码器网络的权重和偏置。$\boldsymbol{W}_{\text{nn}}^{(i)}$和$\boldsymbol{b}_{\text{nn}}^{(i)}$分别表示新的深度网络的权重和偏置。训练好的网络权重和偏置可以直接用于新网络的初始化，以加速新任务的学习过程，提高训练效率。新网络的隐藏层可由如下公式计算：

$$\boldsymbol{H}^{\text{nn}}=f(\boldsymbol{W}^{\text{nn}}x^{\text{nn}}+\boldsymbol{b}^{\text{nn}}), \tag{9}$$

其中$\boldsymbol{H}^{\text{nn}}$是新网络的隐藏层表示，$x^{\text{nn}}$是新应用场景中输入的数据，$\boldsymbol{W}^{\text{nn}}$和$\boldsymbol{b}^{\text{nn}}$是从训练完成的深度编码器网络中迁移并初始化的权重和偏置。

2 直觉模糊深度自动编码器的特征选择算法

IFDAE（intuitionistic fuzzy deep autoencoder）算法是一个结构稀疏特征选择框架，它基于自编码器的编码-解码框架，旨在从高维特征中选择能够重建整个原始特征空间的足够特征。该框架由预处理、预训练和特征选择 3 个阶段构成：预处理阶段抑制低判别性信息的权重，使训练过程聚焦于类别边界敏感的维度；在预训练阶段，模型旨在学习获得整个特征空间良好重建的参数，并迁移到特征选择编码器中；特征选择阶段引入稀疏层，选择维持重建所需和分类效果的足够特征。完整的框架如图 1 所示，下面将进一步详细解释该架构。

2.1 直觉模糊隶属度方案

直觉模糊集的提出旨在应对不确定性带来的挑战，它通过引入隶属度、非隶属度和犹豫度，通过现有信息和观察数据对复杂情况进行更加精准的建模和分析。在本节中，将数据$\boldsymbol{X}\in\mathbb{R}^{N\times d}$投影到核空间来分析和设计直觉模糊集方案，其中$N$表示数据集样本的总数，每个样本具有维度$d$。该步骤的目的是分配一个得分函数，进而对$\boldsymbol{X}$进行模糊加权得到模糊数据$\boldsymbol{X}^{\text{f}}$。

本文将样本x_i的隶属度函数定义为

$$\theta(x_i)=\begin{cases}1-\dfrac{\|\psi(x_i)-C_{\text{T}}\|}{R_{\text{T}}+\sigma}, & c_i=1,\\[2ex] 1-\dfrac{\|\psi(x_i)-C_{\text{F}}\|}{R_{\text{F}}+\sigma}, & c_i\neq 1,\end{cases} \tag{10}$$

其中$C_{\text{T}}(C_{\text{F}})$和$R_{\text{T}}(R_{\text{F}})$分别为样本正（负）类的类中心和类半径，$c_i$表示第$i$个样本的类别，$\sigma$表示微调参数，$\psi$表示内核映射。在隶属度函数中，类中心和类半径分别代表了某类数据在特征空间中的中心位置和分布范围，类中心的定义为

$$C_{\text{T}}=\frac{1}{N_{\text{T}}}\sum_{c_i=1}x_i \text{ 和 } C_{\text{F}}=\frac{1}{N_{\text{F}}}\sum_{c_i\neq 1}x_i, \tag{11}$$

其中$N_{\text{T}}(N_{\text{F}})$是正（负）类的样本总数，类半径定义为

$$R_{\text{T}}=\max_{c_i=1}\|x_i-C_{\text{T}}\| \text{ 和 } R_{\text{F}}=\max_{c_i=0}\|x_i-C_{\text{F}}\|。 \tag{12}$$

在直觉模糊隶属度方案中，需要额外计算每个样本的非隶属度值，非隶属度表示在一个样本周围的局部区域内，异类数据相对于所有数据的占比情况，提供了样本不属于某个类别的信息，非隶属度函数$\bar{\theta}$定义为

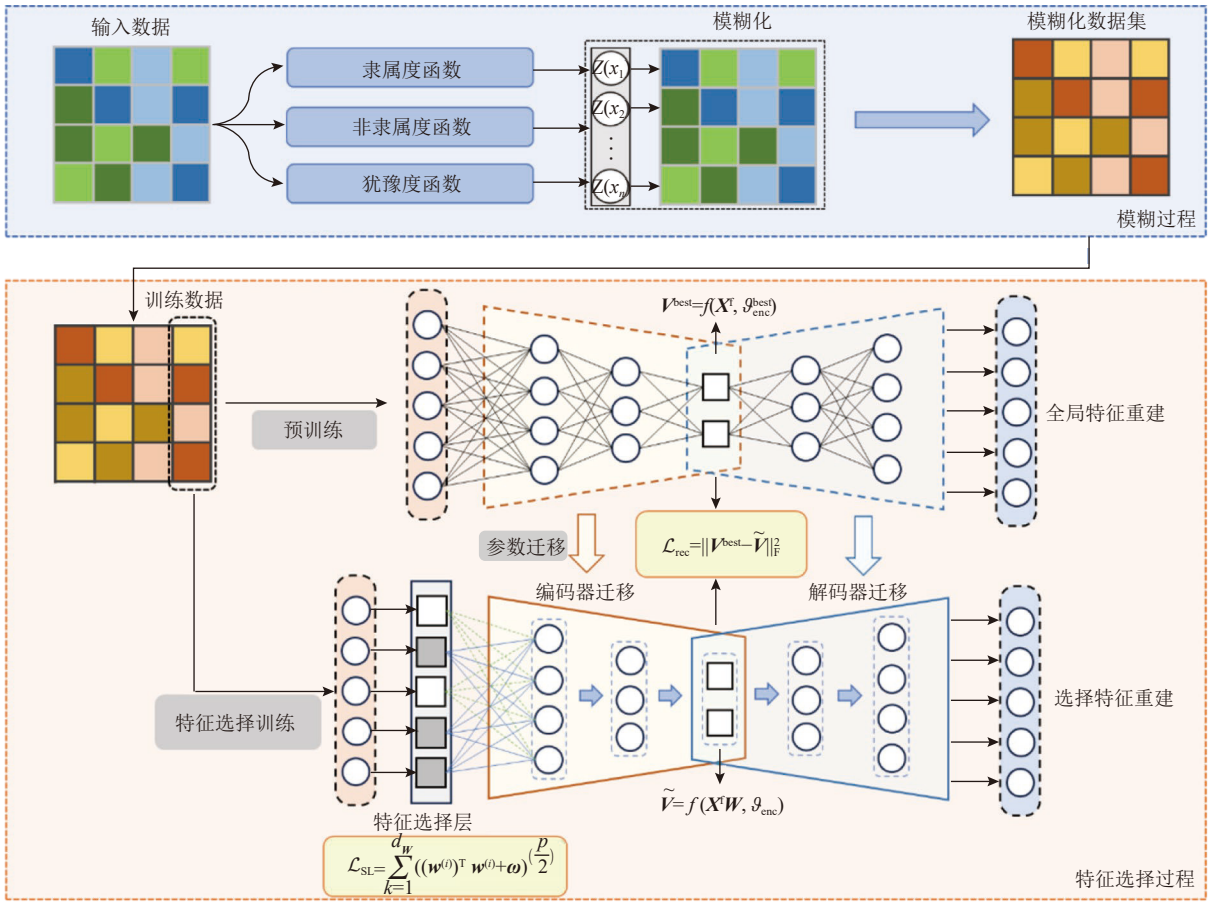


Fig. 1 A visualized illustration of the feature selection method IFDAE based on intuitionistic fuzzy autoencoder

图 1 基于直觉模糊自动编码器的特征选择方法 IFDAE 的可视化示意图

$$\bar{\theta}=(1-\theta(x_i)\Theta(x_i))\,, \tag{13}$$

其中$0\leqslant\theta(x_i)+\bar{\theta}(x_i)\leqslant 1$，该方案的犹豫度计算公式为$h(x)=1-\theta(x)-\bar{\theta}(x)$，邻域异类比例系数$\Theta$量化数据同异质性，取值范围为$[0,1]$，$\Theta$的定义为

$$\Theta(x_i)=\frac{|\{x_k:\|\psi(x_i)-(x_k)\|\leqslant\delta,t_i\neq t_k\}|}{|\{x_k:\|\psi(x_i)-(x_k)\|\leqslant\delta\}|}\,, \tag{14}$$

其中δ为调节参数，$|\cdot|$表示集合的基数。

计算得到每个样本的隶属度值和非隶属度值以后，考虑两者之间的关系，得分函数的定义为：

$$Z(x_i)=\begin{cases}\theta(x_i), & \bar{\theta}(x_i)=0\,,\\ \varepsilon, & \theta(x_i)\leqslant\bar{\theta}(x_i)\,,\\ \dfrac{1-\bar{\theta}(x_i)}{2-\theta(x_i)-\bar{\theta}(x_i)}, & \text{其他}\,。\end{cases} \tag{15}$$

某个样本的非隶属度大于其隶属度时，其与对应类别仍存在潜在的弱关联性，可以选择地对分类效果具有区分能力的特征提供指导。因此将式(15)中的参数设置为$\varepsilon=\min(Z(x_i))$。

得分函数作用于$\boldsymbol{X}$，得到的模糊数据$\boldsymbol{X}^{\mathrm{f}}$定义为

$$\boldsymbol{X}^{\mathrm{f}}=\mathrm{diag}(Z(x_1),Z(x_2),\cdots,Z(x_n))\cdot\boldsymbol{X}\,。 \tag{16}$$

2.2 预训练：潜在表示的迁移学习

在预训练阶段，核心目标是实现高性能的重建，尤其在高维数据中，这是一项具有挑战性的任务。

算法 1. 自动编码器预训练。

输入：$\boldsymbol{X}\in\mathbb{R}^{N\times d}$，第$i$个样本$x_i$及其标签$c_i$，编码器网络$f(\cdot)$，解码器网络$g(\cdot)$，epochs 的数量$\varepsilon$，学习率$\lambda$，损失函数$\mathcal{L}_p$，batches 的数量$B$；

输出：编码器和解码器的参数$\vartheta^{\mathrm{best}}=\{\vartheta_{\mathrm{enc}}^{\mathrm{best}},\vartheta_{\mathrm{dec}}^{\mathrm{best}}\}$。

① 初始化深度自动编码器的参数；

② 用式(10)计算输入数据x_i的隶属度$\theta(x_i)$；

③ 用式(13)计算输入数据x_i的非隶属度$\bar{\theta}(x_i)$；

④ 用式(15)计算直觉模糊得分$Z(x_i)$；

⑤ 用式(16)构造模糊加权矩阵$\boldsymbol{X}^{\mathrm{f}}$；

⑥ 随机初始化参数$\vartheta=\{\vartheta_{\mathrm{enc}},\vartheta_{\mathrm{dec}}\}$；

⑦ for $e\in\{1,2,\cdots,\varepsilon\}$ do

⑧ for $b\in\{1,2,\cdots,B\}$ do

⑨ 计算 $\hat{\boldsymbol{X}}_b^{\mathrm{f}}=g\left(f\left(\boldsymbol{X}_b^{\mathrm{f}};\vartheta_{\mathrm{enc}}\right);\vartheta_{\mathrm{dec}}\right)$；

⑩ 计算损失函数$\mathcal{L}_p=\|\boldsymbol{X}_b^{\mathrm{f}}-\hat{\boldsymbol{X}}_b^{\mathrm{f}}\|_{\mathrm{F}}^2$；

⑪ 更新参数$\vartheta\leftarrow\vartheta+\nabla_\vartheta\mathcal{L}_p$；

⑫ end for

⑬ end for

⑭ return $\vartheta_{\text{enc}}^{\text{best}} \leftarrow \vartheta_{\text{enc}}, \vartheta_{\text{dec}}^{\text{best}} \leftarrow \vartheta_{\text{dec}}$。

为此，本文构建了一个由多层编码器和解码器组成的深度自编码器结构，通过非线性映射，将高维数据压缩至低维潜在表示，并通过重建优化模型参数，确保编码器保留原始特征数据的关键信息。通过最小化损失函数来获得对原始特征空间的全局结构进行有效表示的低维编码：

$$\min_{\vartheta_{\text{enc}},\vartheta_{\text{dec}}} \|\boldsymbol{X}^{\text{f}} - \hat{\boldsymbol{X}}^{\text{f}}\|_{\text{F}}^2 。 \tag{17}$$

本文的特征选择方法在输入层和第 1 个隐藏层之间引入一个稀疏特征选择层，用于保存与输入要素关联的权重，设$\boldsymbol{w} \in \mathbb{R}^n$表示与输入特征$\boldsymbol{X}_i^{\text{f}}$相关联的权重$\boldsymbol{w}_i$的向量，其中$i \in \{1,2,\cdots n\}$，使得乘法层的输出$\boldsymbol{U} \in \mathbb{R}^{N\times n}$可以通过$\boldsymbol{X}^{\text{f}}\boldsymbol{W}$得到，其中$\boldsymbol{W} = \text{diag}(\boldsymbol{w})$表示具有对角线$\boldsymbol{w}$的对角矩阵。此外，特征选择层的权重为非负，因为非负性约束使其解释更有意义，并且能够清楚地反映特征的相关性。其目标是训练网络使冗余特征相关的权重趋于 0。编码器结构通过迁移学习从预训练网络中继承，并且使用预训练网络中的参数$\theta_{\text{enc}}^{\text{best}}$来构建具有良好表示能力的编码。在特征空间到潜在空间的映射$\boldsymbol{X}^{\text{f}}\boldsymbol{A} = \boldsymbol{V}^{\text{best}}$中，$\boldsymbol{A}$是一个$n\times p$的非线性变换矩阵，其中$n > p$。在预训练过程中，编码器和解码器是高度耦合的，若简单地迁移预训练的编码器和解码器网络用于特征选择，可能会使得学习到的潜在表示$\tilde{\boldsymbol{V}}$和具有最佳重建效果的潜在表示$\boldsymbol{V}^{\text{best}}$存在较大偏差。因此，特征选择网络的目标是迁移预训练网络中的潜在表示，同时在网络中忽略解码器，降低了训练成本。

2.3 基于权重迁移的特征选择方法

在特征选择过程中，利用预训练编码器得到的最优潜在表示$\boldsymbol{V}^{\text{best}}$作为目标，最小化特征选择后潜在表示$\tilde{\boldsymbol{V}}$与目标表示的距离来获得高性能重建效果。本节的目标是获得最具区别性的特征子集，因此对问题进行正则化处理，最小化以下损失函数为目标：

$$\mathcal{L} = \|\boldsymbol{V}^{\text{best}} - \tilde{\boldsymbol{V}}\|_{\text{F}}^2 + \lambda_1\|\vartheta_{\text{enc}}\|_1 + \lambda_2\|\boldsymbol{W}\|_1 , \tag{18}$$

其中$\boldsymbol{V}^{\text{best}} = f\left(\boldsymbol{X}^{\text{f}}, \vartheta_{\text{enc}}^{\text{best}}\right)$表示预训练网络中得到的重建编码，$\tilde{\boldsymbol{V}} = f\left(\boldsymbol{X}^{\text{f}}\boldsymbol{W}, \vartheta_{\text{enc}}\right)$表示特征选择网络训练过程中得到的重建编码，$\lambda_1$表示编码器中各层正则化项的系数，$\lambda_2$表示特征选择层的正则化项的系数，编码器层的$l_1$范数正则化通过压缩权重值抑制无关特征的影响，防止其被错误地识别为重要特征。特征选择层的l_1范数正则化通过稀疏化权重来强制模型自动选择最具判别力的特征。

然而，l_1范数正则化对特征之间的相关性没有显式处理，可能造成高度相关的特征被同时或交替选择，而其他相关特征则被忽视。此外，在权重接近 0 时，其不连续性也可能导致训练过程不稳定。在时间复杂度上，对于权重向量$\boldsymbol{w} \in \mathbb{R}^n$，$l_1$范数与$l_{2,p}$范数对每个元素$w_i$执行逐步计算，虽然$l_{2,p}$范数单个元素的操作步骤较$l_1$范数稍多，但不会改变整体的复杂度阶，在求解大规模特征维度的 NP-难问题下，$l_{2,p}$范数在时间复杂度上与l_1范数一致，为$O(n)$，因此本文将l_1范数正则化改变为$l_{2,p}$范数正则化$(p > 0)$来进行优化，其通过引入p值对高相关性特征进行“共同惩罚”，从而避免相关性较强的特征被单独剔除。具体地，本文将$l_{2,p}$范数正则化应用到特征选择层的权重矩阵上，损失函数优化为

$$\mathcal{L} = \|\boldsymbol{V}^{\text{best}} - \tilde{\boldsymbol{V}}\|_{\text{F}}^2 + \lambda_1\|\vartheta_{\text{enc}}\|_1 + \lambda_2\|\boldsymbol{W}\|_{2,p}^p 。 \tag{19}$$

因此本文的非线性特征选择模型的最终目标函数可以写为

$$\mathcal{L} = \|\boldsymbol{V}^{\text{best}} - \tilde{\boldsymbol{V}}\|_{\text{F}}^2 + \lambda_1\|\vartheta_{\text{enc}}\|_1 + \lambda_2\sum_{k=1}^{d_W}\|\boldsymbol{w}^{(k)}\|_2^p , \tag{20}$$

其中d_W表示对角权重矩阵$\boldsymbol{W}$的阶数，$\boldsymbol{w}^{(k)}$表示$\boldsymbol{W}$的第k行向量。由于$p > 0$，$\boldsymbol{w}^{(k)}$值可能会趋于$\boldsymbol{0}$。

在数学上，正则化项为 0 将会导致损失函数无法使用梯度下降法进行优化。为了避免这种情况，将$l_{2,p}$范数正则化项进一步优化为$((\boldsymbol{w}^{(i)})^{\text{T}}\boldsymbol{w}^{(i)} + \boldsymbol{\omega})^{(p/2)}$，当$\boldsymbol{w}^{(k)} \to \boldsymbol{0}$时，常数项$\boldsymbol{\omega}$使得正则化项仍为非零值，确保其对于损失函数梯度的贡献。最终损失函数定义为

$$\mathcal{L} = \|\boldsymbol{V}^{\text{best}} - \tilde{\boldsymbol{V}}\|_{\text{F}}^2 + \lambda_1\|\vartheta_{\text{enc}}\|_1 + \lambda_2\sum_{k=1}^{d_W}\left((\boldsymbol{w}^{(i)})^{\text{T}}\boldsymbol{w}^{(i)} + \boldsymbol{\omega}\right)^{\left(\frac{p}{2}\right)} 。 \tag{21}$$

通过最小化$l_{2,p}$范数正则化，可以获得一个结构化稀疏权重矩阵$\tilde{\boldsymbol{W}}$。具体来说，权重矩阵$\tilde{\boldsymbol{W}}$的第k行包含非 0 元素，那么说明数据$\boldsymbol{X}$的第k个特征对模型的学习具有贡献，并在特征选择过程中被保留。此外，该方法对特征权重进行排序以确定每个特征的重要性，并选择h个具有高权重的特征。在此之后，联合加入特征选择层的编码器和预训练的解码器，生成一个能够从所选特征子集重建完整特征空间的网络，如图 2 所示。本文将重建网络的具体参数设置为$\vartheta_{\text{enc}} = \tilde{\vartheta}_{\text{enc}}, \vartheta_{\text{dec}} = \vartheta_{\text{dec}}^{\text{best}}, \boldsymbol{w}_{\text{s}} = \tilde{\boldsymbol{w}}_{\text{s}}$，其中$\tilde{\vartheta}_{\text{enc}}$表示特征选择层编码器的参数。特征选择步骤如算法 2 所示。

算法 2. 特征选择。

输入：训练集$\boldsymbol{X}^{\text{f}} \in \mathbb{R}^{N\times d}$，训练完毕的编码器网络

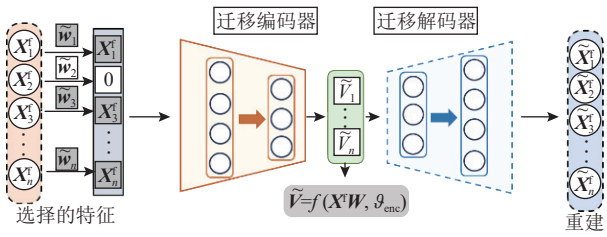


Fig. 2 Reconstruction of network module

图 2 重建网络模块

$f\left(\cdot,\vartheta_{\text{enc}}^{\text{best}}\right)$, epochs 的数量 ε, 学习率 λ, 损失函数 $\mathcal{L}_{\text{FS}}$, batches 的数量 B, 用于编码层的 λ_1 正则化项和用于选择层的 λ_2 正则化项;

输出: 选择层编码器参数 $\tilde{\vartheta}_{\text{enc}}$, 选择层参数 $\tilde{\boldsymbol{w}}_{\text{s}}$, 选出的特征子集 S^k。

① 初始化 $\boldsymbol{w}\in\mathbb{R}^d$, s.t.$\boldsymbol{w}>\boldsymbol{0}$;

② 特征迁移: $\vartheta_{\text{enc}}\leftarrow\vartheta_{\text{enc}}^{\text{best}}$;

③ 计算 $\boldsymbol{V}^{\text{best}}=f\left(\boldsymbol{X}^{\text{f}},\vartheta_{\text{enc}}^{\text{best}}\right)$;

④ for $e\in\{1,2,\cdots,\varepsilon\}$ do

⑤ for $b\in\{1,2,\cdots,B\}$ do

⑥ 计算 $\tilde{\boldsymbol{V}}_b=f\left(\boldsymbol{X}_b^{\text{f}}\boldsymbol{W},\vartheta_{\text{enc}}\right)$;

⑦ 计算重构误差 $\mathcal{L}_{\text{rec}}=\|\boldsymbol{V}^{\text{best}}-\tilde{\boldsymbol{V}}\|_{\text{F}}^2$;

⑧ 计算编码层损失 $\mathcal{L}_{\text{enc}}=\|\vartheta_{\text{enc}}\|_1$;

⑨ 计算选择层损失 $\mathcal{L}_{\text{SL}}=\sum_{k=1}^{d_w}\left((\boldsymbol{w}^{(i)})^{\text{T}}\boldsymbol{w}^{(i)}+\boldsymbol{\omega}\right)^{(p/2)}$;

⑩ 更新参数 $\boldsymbol{w}_{\text{s}},\vartheta_{\text{enc}}\leftarrow\boldsymbol{w}_{\text{s}},\vartheta_{\text{enc}}+\Delta_{\boldsymbol{w}_{\text{s}},\vartheta_{\text{enc}}}(\mathcal{L}_p+\lambda_1\mathcal{L}_{\text{enc}}+\lambda_2\mathcal{L}_{\text{SL}})$;

⑪ end for

⑫ end for

⑬ return $\tilde{\theta}_{\text{enc}}\leftarrow\theta_{\text{enc}},\tilde{\boldsymbol{w}}_{\text{s}}\leftarrow\boldsymbol{w}_{\text{s}}$;

⑭ 将 $\left((\boldsymbol{w}^{(i)})^{\text{T}}\boldsymbol{w}^{(i)}+\boldsymbol{\omega}\right)^{(p/2)}$ 按降序排列, 然后选择排序最高的 h 个特征作为特征子集 S^h。

3 实验结果与分析

在本节中, 我们进行了大量的实验, 以评估所提出的方法在各种数据集上的性能, 并对实验结果进行了分析, 以验证 IFDAE 算法的有效性。

3.1 实验数据集

本文了使用来自 scikit-feature 特征选择代码库[40]和基因表达模型选择系统数据集[41]进行实验。在 12 个不同的高维数据集上评估了所提出的 IFDAE 在不同类型的数据集上的性能, 包括手写数字识别数据集 gisette、文本分类数据集(如 BSAEHOCK, RELATHE, PCMAC)、癌症基因表达数据集(colon)、脑胶质瘤数据集(GLI_85)、经典的微阵列基因表达数据集(leukemia)、前列腺癌基因表达数据集(Prostate_Tumor)、淋巴瘤基因表达数据集(DLBCL)、肺癌检测数据集(SMK_CAN)和蛋白质质谱数据(arcene)等。数据维度从 2 000 到 22 283 不等, 关于这 12 个数据集的更多细节见表 1。

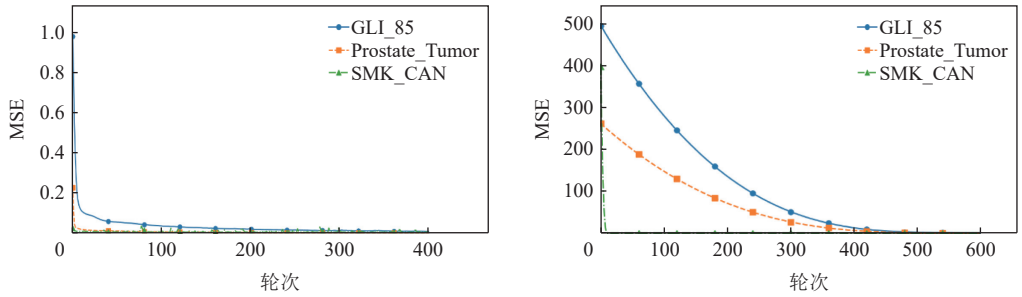
Table 1 Description of Datasets

表 1 数据集描述

序号	数据集	样本数	特征数
1	ALLAML	72	7 129
2	colon	62	2 000
3	GLI_85	85	22 283
4	leukemia	72	7 070
5	Prostate_Tumor	102	10 509
6	RELATHE	1 427	4 322
7	BSAEHOCK	1 993	4 862
8	PCMAC	1 943	3 289
9	DLBCL	77	5 469
10	gisette	7 000	5 000
11	SMK_CAN	187	19 993
12	arcene	200	10 000

3.2 实验设计

在 IFDAE 实验中, 首先对训练轮次和神经网络超参数进行实验分析。根据收敛性预分析, 本文将预训练过程的最大 epoch 数设置为 400, 特征选择过程最大 epoch 数设置为 600, 更多细节如图 3 所示。学习率在{1E−7, 1E−6, 1E−5, 1E−4, 1E−3}范围内改变, 隐藏层数在{1,2,3,4,5}范围内改变, 参数预分析显示两者对模型均方误差(mean squared error, MSE)影响呈非单调趋势, 本文基于分类准确率, 将初始化学习率定为 1E−6, 隐藏层数定为 3, 结果如表 2 和表 3 所示。通过从 $U_{\text{s}}=\{0.999\,999, 0.999\,999\,9\}$ 和其他层均匀采样来初始化选择层的权重。采用 Adam 优化器进行优化, 将式(21)中的 λ_1 和 λ_2 都设置为 1E−6, 确保 $\mathcal{L}_{\text{rec}}$ 和 $\mathcal{L}_{\text{SL}}$ 对 $\mathcal{L}$ 具有相等的影响。对于其他超参数设置, 在验证集上执行网格搜索, 然后选择最优的 1 个。为了简单起见, 该方法只使用 IFDAE 中的线性激活函数。IFDAE 的深度线性版本已经可以实现上级性能。所有实验都在配备有 NVIDIA GeForce RTX 4090 GPU, Intel i9-11900 K CPU@5.80 GHz, 64 GB RAM 的系统上进行。



(a) 不同数据集预训练阶段损失曲线 (b) 不同数据集特征选择阶段损失曲线

Fig. 3 Loss curves during the pre-training and feature selection stages on different datasets

图 3 不同数据集的预训练阶段和特征选择阶段的损失曲线

Table 2 Classification Accuracy Under Different Learning Rates

表 2 不同学习率下的分类准确率

数据集	1E−7	1E−6	1E−5	1E−4	1E−3
GLI_85	0.847±0.090	0.859±0.088	0.824±0.098	0.800±0.121	0.835±0.086
Prostate_Tumor	0.833±0.024	0.874±0.071	0.843±0.046	0.853±0.030	0.844±0.033
SMK_CAN	0.620±0.055	0.631±0.059	0.604±0.073	0.594±0.079	0.615±0.086

注: “±”前数值表示平均值, “±”后数值表示标准差。

Table 3 Classification Accuracy Under Varying Hidden Layers

表 3 不同隐藏层数下的分类准确率

数据集	1 层	2 层	3 层	4 层	5 层
GLI_85	0.859±0.109	0.812±0.146	0.859±0.088	0.835±0.101	0.800±0.080
Prostate_Tumor	0.853±0.029	0.833±0.024	0.874±0.071	0.708±0.113	0.686±0.040
SMK_CAN	0.577±0.102	0.598±0.086	0.631±0.059	0.561±0.121	0.583±0.061

注: “±”前数值表示平均值, “±”后数值表示标准差。

3.3 对比实验设置

为了比较实验结果, 本文选择了一些有代表性的现有算法, 包括: CAE[15], AEFS[13], FAE[17], SDAE[21], LSCAE[16], FSNet[22]。

CAE 是一种无监督的非线性特征选择算法, 使用具体的选择器层作为编码器, 并使用标准的神经网络作为解码器, 用于对所有特征进行排序。

AEFS 通过结合自编码器回归与群组 Lasso 任务, 探索特征间的线性和非线性信息。

FAE 通过添加 1 对 1 的评分层和一个小型子神经网络来扩展自动编码器, 以无监督的方式进行特征选择。

SDAE 充分利用深度自动编码器中高度非线性的潜在表示, 并通过定制的选择层, 以无监督的方式内在地学习相关的全局代表性特征子集。

LSCAE 利用图拉普拉斯准则消除不相关特征, 并通过将自编码器与具体的选择器层相结合来去除特征冗余。

FSNet 包含一个选择层和一个有监督深度神经网络, 在训练过程中同时优化特征选择和重构目标, 以减少重构误差并提升分类性能。

分类准确率是评价特征选择结果分类性能最常用的指标, 而重建误差则用于衡量特征选择过程中所保留信息的完整性, 因此采用 2 个度量用于评估模型: 1)重建误差, 其以 MSE 测量; 2)分类准确性, 其通过将所选特征传递到下游分类器作为可行手段来基准化所选特征子集的质量来测量。为了公平比较, 本文在选定特征集合后训练一个无正则化线性回归模型对原始特征空间进行逆向重建, 并将模型输出的重建误差作为第 1 个度量标准。同时, 本文方法以及对比实验选择的特征子集数量固定为 200 个, 采用 5 折交叉验证进行对比实验, 将数据集随机划分为 5 个大小相同的子集。每次选取其中 1 个作为验证集, 其余 4 个用于训练, 重复 5 次。通过这种方式评估特征选择方法的分类准确率和重构误差, 并计算其平均值与标准差, 确保评估结果的稳健性和客观性。采用极端树作为分类器, 将分类准确率作为第 2 个度量标准。为了进行公平的比较, 对

比算法的参数是根据它们的原始论文设置的。

3.4 公共数据集实验分析

本节中，通过和其他特征选择方法进行比较，采用分类和重建实验来评估本文方法的特征选择性能。本文方法和对比方法均采用 5 重交叉验证进行实验。通过比较表 4 和表 5，得出了以下 3 个结论：

1）IFDAE 算法在分类精度与重建误差方面展现出优越的综合性能，在多数公开数据集上均优于其他 6 种对比方法。其在 8 个数据集上取得了最高的分类准确率，并在 11 个数据集中实现了最优的重建效果。

2）IFDAE 在几乎所有数据集上都展现出出色的重建能力，特别是在 colon 和 GLI_85 数据集上，其他方法普遍存在较严重的信息损失，而 IFDAE 依然保持了较高的重构保真度。

3）不同算法在各个数据集上的表现存在差异，没有哪种方法能在所有数据集上持续占优。这种差异性可能源于样本分布的变化以及各算法采用的特征选择准则不同。尽管如此，IFDAE 在特征选择用于分类与重构的双重任务中始终保持了显著的有效性。

根据分类结果，本研究采用 Friedman 检验[42] 和 Bonferroni-Dunn 检验[43] 对实验结果进行统计分析，以评估各特征选择算法性能差异的显著性。Friedman 检验是一种广泛使用的非参数统计方法，它对每个数据集上的算法进行排名，并比较它们的平均排名，

Table 4 Linear Reconstruction Error After Selecting Features by Different Algorithms

表 4 不同算法选择特征后的线性重建误差

数据集	CAE	AEFS	FAE	SDAE	LSCAE	FSNet	IFDAE（本文）
ALLAML	1.087±0.072	0.973±0.115	0.910±0.163	0.914±0.144	0.912±0.075	1.138±0.201	**0.089±0.048**
colon	/	0.016±0.005	/	/	/	/	**0.001±0.000**
GLI_85	/	**0.040±0.003**	/	/	/	/	0.052±0.006
leukemia	2.667±0.186	2.463±0.108	2.265±0.129	2.288±0.087	2.273±0.047	2.332±0.046	**0.172±0.035**
Prostate_Tumor	0.022±0.009	0.089±0.122	0.030±0.010	0.020±0.008	0.025±0.007	0.018±0.005	**0.005±0.003**
RELATHE	3.627±6.891	0.098±0.046	0.091±0.034	0.071±0.029	0.092±0.038	0.112±0.043	**0.022±0.004**
BASEHOCK	/	0.059±0.019	0.048±0.009	0.032±0.007	0.053±0.020	0.062±0.016	**0.016±0.003**
PCMAC	0.170±0.097	0.051±0.010	0.063±0.011	0.044±0.008	0.060±0.016	0.087±0.022	**0.016±0.003**
DLBCL	0.041±0.000	0.044±0.046	0.025±0.011	0.027±0.011	0.024±0.008	0.028±0.006	**0.004±0.001**
gisette	0.025±0.009	0.043±0.000	0.041±0.000	0.056±0.002	0.046±0.002	0.061±0.000	**0.008±0.001**
SMK_CAN	0.041±0.002	0.042±0.005	0.043±0.003	0.050±0.021	0.051±0.012	0.074±0.010	**0.008±0.002**
arcene	0.087±0.019	0.027±0.003	0.247±0.035	0.088±0.010	0.274±0.067	0.028±0.003	**0.022±0.008**

注：最高分类性能和最佳重建性能以黑体数值突出表示，这些值表示为“平均值±标准差”，“/”表示重构误差过大的情况，即重构误差超过 10.000。

Table 5 Classification Accuracy After Selecting Features by Different Algorithms

表 5 不同算法选择特征后的分类准确率

数据集	CAE	AEFS	FAE	SDAE	LSCAE	FSNet	IFDAE（本文）
ALLAML	0.887±0.073	0.862±0.046	0.751±0.065	0.751±0.088	0.820±0.052	0.888±0.087	**0.890±0.091**
colon	0.837±0.055	0.677±0.081	0.696±0.118	0.869±0.069	0.826±0.087	0.659±0.014	**0.888±0.063**
GLI_85	0.694±0.136	0.788±0.089	0.765±0.037	0.788±0.047	0.776±0.120	0.624±0.047	**0.859±0.088**
leukemia	**0.973±0.053**	0.877±0.128	0.751±0.076	0.850±0.096	0.875±0.069	0.931±0.073	0.765±0.126
Prostate_Tumor	0.823±0.042	0.784±0.026	0.718±0.151	0.853±0.053	0.774±0.074	0.844±0.089	**0.874±0.071**
RELATHE	0.650±0.018	0.797±0.042	0.745±0.011	0.844±0.014	0.812±0.029	0.671±0.041	**0.853±0.011**
BASEHOCK	0.799±0.020	0.862±0.065	0.783±0.053	0.957±0.007	0.877±0.031	0.633±0.031	**0.958±0.013**
PCMAC	0.601±0.021	0.851±0.015	0.776±0.023	0.869±0.024	0.825±0.047	0.569±0.011	**0.883±0.036**
DLBCL	**0.975±0.020**	0.831±0.037	0.780±0.121	0.898±0.108	0.818±0.102	0.870±0.058	0.687±0.090
gisette	0.883±0.097	0.942±0.006	**0.960±0.008**	0.768±0.051	0.934±0.020	0.883±0.010	0.915±0.016
SMK_CAN	0.694±0.097	0.696±0.069	**0.707±0.065**	0.652±0.024	0.630±0.069	0.653±0.120	0.631±0.059
arcene	0.765±0.037	0.750±0.079	0.775±0.032	0.775±0.091	0.775±0.076	0.730±0.071	**0.790±0.090**

注：最高分类性能和最佳重建性能以黑体数值突出表示，这些值表示为“平均值±标准差”。

从而评估 7 种算法在 12 个数据集上的整体分类准确率。假设对m种特征选择算法在n个数据集上进行评估；第i种算法的平均排名记为r_i，Friedman 统计量及其改进版本[44]定义为：

$$\tau_{\chi^2}=\frac{12n}{m(m+1)}\left(\sum_{i=1}^{m}r_i^2-\frac{m(m+1)^2}{4}\right)。\tag{22}$$

在 Friedman 检验之后的事后分析中，临界差（critical difference, CD）按照以下公式计算[45]：

$$CD_\alpha=q_\alpha\sqrt{\frac{m(m+1)}{6n}},\tag{23}$$

其中q_α表示对应于特定显著性水平α的 Studentized 极差分布所确定的阈值。Friedman 检验用于判断在多个数据集上评估的多种特征选择算法之间是否存在统计上显著的性能差异。若等效性能的原假设被拒绝，则进一步采用 Bonferroni-Dunn 事后检验计算的临界差来进行后续比较。随后应用该算法来识别哪些特定的算法在性能上表现出统计学上显著的差异[46-47]。在显著性水平$\alpha=0.1$下，Friedman 统计量的临界值为$F(6,66)=1.87$。在极端树分类器下，结果呈现统计学上的显著差异，其 Friedman 检验统计量为$\tau_F=1.93$，超过了给定显著性水平下的临界阈值。因此，原假设（各算法之间不存在显著性能差异）被拒绝。该结果确认，在 7 种评估的特征选择算法中观察到的性能差异具有统计学显著性。此外 Bonferroni-Dunn 检验进一步支持了这一结论，显示所提出的 IFDAE 算法较其他竞争算法具有更好的性能，从而与 Friedman 检验的结果相一致。为了直观展示比较分类性能，图 4 给出了临界差图。显示所评估方法的平均排名，并在顶部标注了临界差阈值。水平线连接那些性能差异未达到统计显著性的算法。图 4 显示，IFDAE 方法表现优于其他 6 种算法。尽管 Bonferroni-Dunn 检验未显示 IFDAE 与某些竞争算法的差异具有统计显著性，IFDAE 在排名上仍优于所有其他比较方法。

为验证 IFDAE 算法的计算成本，本文对其进行时间复杂度分析。模糊化模块将数据$\boldsymbol{X}\in\mathbb{R}^{N\times d}$投影到核空间进行分析，其时间复杂度由全样本核矩阵计算主导，对每个元素进行向量内积计算，其时间复杂度为$O\left(N^2d\right)$。特征选择模块使用自动编码器筛选关键特征，其主要计算成本来源于神经网络的训练过程，时间复杂度为$O\left(EN\left(dh+L_{\mathrm{h}}h^2+hE\right)\right)$，其中$E$为训练轮次，$h$为隐藏层神经元数，$L_{\mathrm{h}}$为隐藏层层数。IFDAE 算法实际应用于高维特征场景，通常满足$d\gg N,L_{\mathrm{h}},E$，同时$O\left(N^2d\right)\ll O\left(EN\left(dh+L_{\mathrm{h}}h^2+hE\right)\right)$，最终复杂度受特征选择模块主导，因此本算法整体复杂度可进一步计算为$O(ENdh)$。尽管 IFDAE 算法与对比算法同属自编码器衍生的特征学习方法，其核心时间复杂度阶数均由全连接层矩阵乘法与训练迭代决定，统一表现为$O(ENdh)$，但由于 IFDAE 引入模糊化模块对噪声数据进行有效预处理，在后续网络训练中具有显著的时间优势。为了进一步验证 IFDAE 的高效性，本研究选用万维以上数据集 GLI_85, Prostate_Tumor 和 SMK_CAN 进行对比实验，以平均训练耗时作为度量标准，表 6 显示，IFDAE 算法优于其他 6 种对比算法，在保证特征选择性能的同时，获得了更高的训练效率。

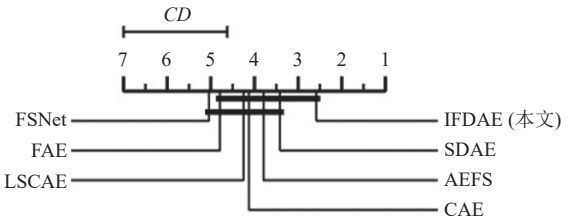


Fig. 4 CD diagrams of feature selection algorithms

图 4 特征选择算法的临界差图

Table 6 Average Training Time Across Three Datasets

表 6 3 种数据集平均训练耗时 s/epoch

算法	GLI_85	Prostate_Tumor	SMK_CAN	平均值
CAE	0.303 9	0.139 9	0.486 1	0.310 0
FAE	0.268 8	0.158 0	0.335 3	0.254 0
AEFS	0.661 4	0.299 0	0.979 9	0.646 8
SDAE	0.086 4	0.088 1	0.182 9	0.119 1
LSCAE	0.242 4	0.091 7	0.352 6	0.228 9
FSNet	0.119 7	0.115 1	**0.165 1**	0.133 3
IFDAE（本文）	**0.029 3**	**0.027 1**	0.295 4	**0.117 3**

注：黑体数值表示最优值。

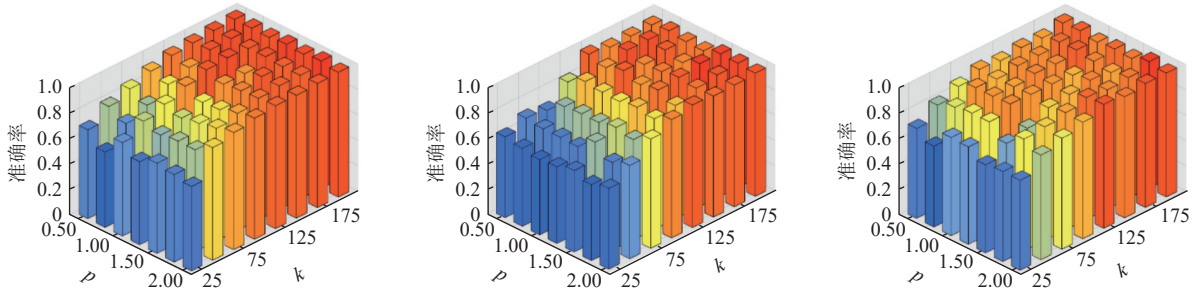
3.5 参数敏感度

本节将进一步研究 IFDAE 中不同参数对实验结果的影响。具体地说，特征选择数k直接界定了最终筛选出的特征子集维度，而正则化参数p的取值决定了正则化项对高相关特征的协同惩罚强度，本节探究上述 2 个核心参数对分类准确率结果的影响规律。选用高冗余度数据集 BSAEHOCK、强特征相关性数据集 PCMAC、平衡性数据集 RELATHE 进行实验。当我们将p的值从$\{0.5,0.75,1,1.25,1.5,1.75,2\}$范围内改变，选择的特征数量$k$的值从$\{25,50,75,100,125,150,175,200\}$范围内改变时，图 5 展示了不同数据集上不同$k$和$p$在极端树分类器下的分类准确率。通过计算

结果可知,该算法的分类性能受所选特征的数量的极大影响,但是对正则化参数p是鲁棒的,随着参数p值的变化,性能仍然稳定。尽管如此,由于每个数据集都有其独特的分布特征,因此最合适的p值因数据集不同而不同,本文还是建议对 IFDAE 执行分层网格搜索,以便在实际应用程序中获得良好的性能。

3.6 消融实验

为了评估 IFDAE 中每个模块的贡献,通过删除直觉模糊隶属度模块以及将损失函数中的$l_{2,p}$范数改变为l_1范数来进行消融研究。本节使用极端树分类器来评估特征选择算法中每个模块的贡献,并验证直觉模糊隶属度和$l_{2,p}$范数选择对模型性能的影响。这项研究的结果如图 6 所示。在本节的消融实验中 IFDAE-1 表示消除直觉模糊隶属度模块的设置,而 IFDAE-2 表示损失函数中的$l_{2,p}$范数改变为l_1范数来进行稀疏化处理。实验结果清楚地说明了每个模块的重要性。排除直觉模糊隶属度模块后降低了特征的辨别能力,突出了对噪声和离群值的有效处理。类似地,改变$l_{2,p}$范数后导致了模型分类准确率的显著下降,强调了结构化稀疏性的优势。



(a) BSAEHOCK数据集参数实验 (b) PCMAC数据集参数实验 (c) RELATHE数据集参数实验

Fig. 5 Accuracy results of parameter sensitivity experiments on three types of datasets

图 5 3 类数据集上参数敏感性实验的准确率结果

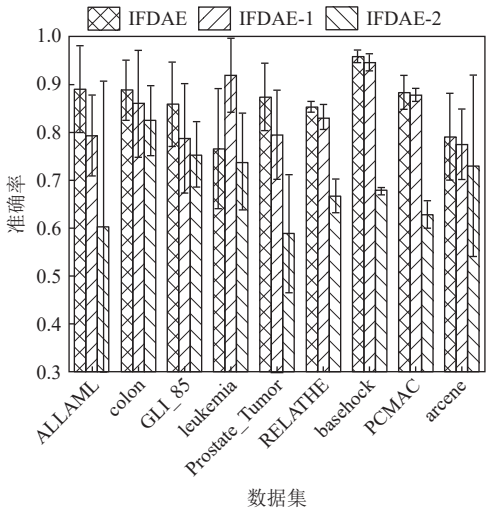


Fig. 6 Accuracy results of ablation experiment on different datasets

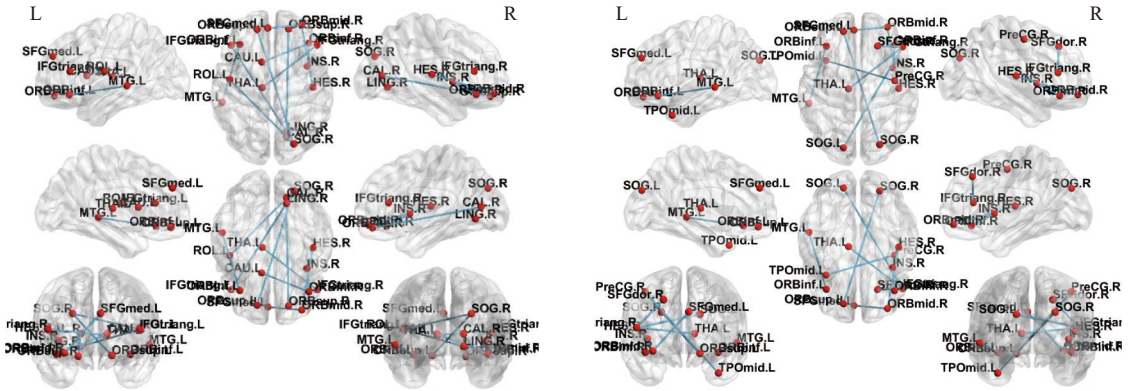
图 6 在不同数据集上消融实验的准确率结果

3.7 精神分裂症数据集的应用

在验证算法在多个公开数据集上的有效性后,本文进一步将其应用于精神分裂症数据集,以测试其在复杂任务中的适应性。精神分裂症因其症状多样、机制复杂,诊断难度较大。传统方法依赖临床观察和主观报告,准确性有限。近年来,研究者尝试将机器学习与临床数据结合,发掘潜在模式和生物标志物,以提升诊断的客观性和准确性,并加深对疾病机制的理解。本文选用了来自四川大学华西医院的精神分裂症数据集[48],该数据集因其复杂性而成为检验 IFDAE 算法在特征识别和分类分析方面能力的理想对象。所用数据为静息态功能磁共振成像(rs-fMRI)通过 3-T 通用电气 MRI 扫描仪获得,其预处理过程包括去除异常数据以确保数据质量,并通过 SPM 8 软件及其辅助工具完成时序校正与空间归一化[49]。随后,采用 AAL 模板将 rs-fMRI 图像划分为 90 个脑区[50],并从每个脑区提取所有体素的平均时间序列作为区域特征。最终,本文构建了一个包含 773 个样本反映 4 005 个脑区之间连接的数据集。

为了评估特征选择方法的有效性及可靠性,本文采用逻辑回归和极端树 2 个分类器,对所选特征子集的分类性能进行了测试。为确保结果的稳定性与可靠性,采用了 5 折交叉验证策略。在逻辑回归(LR)分类器中,IFDAE 方法取得了 69.9% 的最高分类准确率,明显优于其余算法;FAE 和 CAE 分别以 66.2% 和 64.5% 的准确率位列第二和第三。类似地,在极端树分类器中,IFDAE 同样表现出色,达到 63.0% 的准确率,领先于 FAE(61.2%)和 CAE(61.0%)。这些结果充分展示了 IFDAE 在精神分裂症分类任务中的优势与稳健性。此外,对脑网络拓扑连接结构

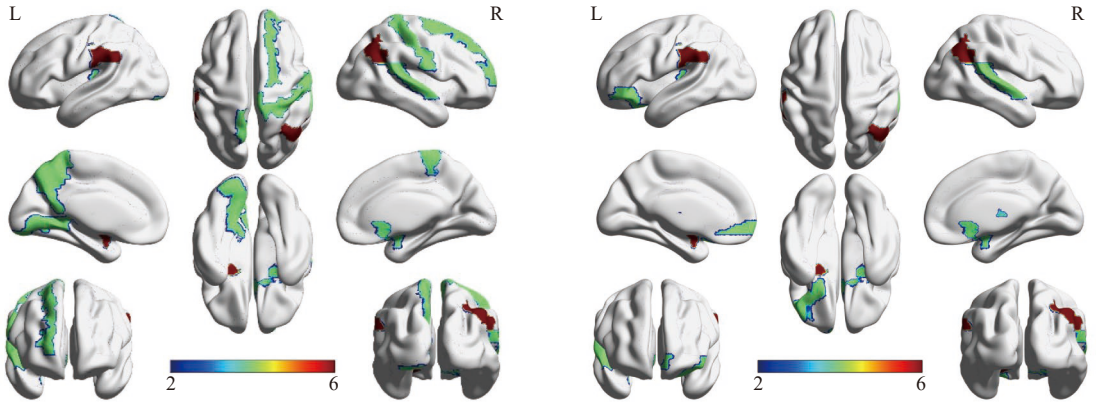
的分析进一步揭示了患者组与健康对照组之间的显著差异。如图 7 所示，展示了最具代表性的 10 条功能连接。对这些显著连接的进一步分析帮助识别出在分辨 2 个群体中起关键作用的高频脑区。为了进一步评估这些脑区的重要性，本文统计了 2 个分类器对各脑区的特征选择，如图 8 所示，该实验有助于识别最常被选择的脑区，突出其在分类过程中的关键作用。强调其在分类过程中的相关性。大量研究[51-54]已证实，本文方法识别出的脑区在区分精神分裂症患者方面具有显著效能，这些研究为本文的发现提供了有力支持，并进一步增强了已识别脑区的可靠性。



(a) 逻辑回归分类器结果　　(b) 极端树分类器结果

Fig. 7　Visualization diagram of the top ten alterations of connectivities differentiating schizophrenia and normal groups under two classifiers

图 7　在 2 个分类器下区分精神分裂症和正常组的前 10 个连接性改变的可视化图



(a) 逻辑回归分类器结果　　(b) 极端树分类器结果

Fig. 8　Visualization diagram of brain regions differentiating schizophrenia and normal groups under two classifiers

图 8　在 2 个分类器下区分精神分裂症与正常组的脑区可视化图

4　总　　结

本文提出了一种结合结构稀疏正则化的非线性特征选择算法(IFDAE)，该算法融合了模糊理论与自动编码器模型，旨在应对高维数据中常见的噪声与离群值问题。IFDAE 通过学习非线性变换来捕捉数据的潜在结构特征，并在输入层与首个隐含层之间引入 $l_{2,p}$ 范数结构稀疏正则化，借此实现基于梯度的高效优化过程。同时，本文引入特征迁移结构以促进知识在不同任务间的迁移与重用。通过计算样本的隶属度和非隶属度，引入的直觉模糊权重机制有效处理了数据中的不确定性因素，进一步提升了特征选择的鲁棒性及分类和重构性能。IFDAE 在多个公开数据集以及真实的精神分裂症数据集上进行了实证评估，验证了其优越性。

尽管实验结果比较理想，但本文的工作仍然存在一些局限性。首先，IFDAE 直觉模糊权重的计算依赖类中心和类半径的计算，这种依赖可能削弱该算法在小样本场景的泛化能力。此外，该算法参数选择依赖网格搜索，其离散性可能导致部分最优参数组合被忽略。解决方案可能涉及使用基于进化算

法的优化策略来更好地学习参数空间的全局结构。在未来工作中，我们计划将该算法拓展至多模态脑影像融合，以进一步提升对复杂脑区连接的建模能力，并探索自适应参数优化机制与特征选择模型的结合方法。

作者贡献声明：华启冲提出了基于直觉模糊深度自动编码器的特征选择算法，进行了实验与验证；丁卫平构建了论文算法的总体架构，并对算法实现提供了指导；陈悦鹏给出了实验的指导，对论文及模型框图进行了修改；周天奕提供了实验必备的硬件和软件设施，并对实验提供指导帮助。

参 考 文 献

[1] Luo Chuan, Wang Sizhao, Li Tianrui, et al. Large-scale meta-heuristic feature selection based on BPSO assisted rough hypercuboid approach[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34(12): 10889−10903

[2] Zheng Zhenxing, An Gaoyun, Cao Shan, et al. Collaborative and multilevel feature selection network for action recognition[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34(3): 1304−1318

[3] Yang Jiaquan, Yang Qite, Du Kejing, et al. Bi-directional feature fixation-based particle swarm optimization for large-scale feature selection[J]. IEEE Transactions on Big Data, 2023, 9(3): 1004−1017

[4] Zhang Yongshan, Wu Jia, Cai Zhihua, et al. Multi-view multi-label learning with sparse feature selection for image annotation[J]. IEEE Transactions on Multimedia, 2020, 22(11): 2844−2857

[5] Gunduz H. An efficient dimensionality reduction method using filter-based feature selection and variational autoencoders on Parkinson's disease classification[J]. Biomedical Signal Processing and Control, 2021, 66: 102452

[6] Ding Weiping, Lin Chin-Teng, Pedrycz W. Multiple relevant feature ensemble selection based on multilayer co-evolutionary consensus MapReduce[J]. IEEE Transactions on Cybernetics, 2020, 50(2): 425−439

[7] Liu Jinghua, Lin Yan, Ding Weiping, et al. Fuzzy mutual information-based multilabel feature selection with label dependency and streaming labels[J]. IEEE Transactions on Fuzzy Systems, 2023, 31(1): 77−91

[8] Zhao Hong, Wang Ping, Hu Qinghua, et al. Fuzzy rough set based feature selection for large-scale hierarchical classification[J]. IEEE Transactions on Fuzzy Systems, 2019, 27(10): 1891−1903

[9] Yuan Zhong, Chen Hongmei, Zhang Pengfei, et al. A novel unsupervised approach to heterogeneous feature selection based on fuzzy mutual information[J]. IEEE Transactions on Fuzzy Systems, 2022, 30(9): 3395−3409

[10] Zheng Wei, Chen Shuo, Fu Zhenyong, et al. Feature selection boosted by unselected features[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(9): 4562−4574

[11] Chakraborty R, Pal N R. Feature selection using a neural framework with controlled redundancy[J]. IEEE Transactions on Neural Networks and Learning Systems, 2015, 26(1): 35−50

[12] Zhou Peng, Li Peipei, Zhao Shichao, et al. Feature interaction for streaming feature selection[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(10): 4691−4702

[13] Han Kai, Wang Yunhe, Zhang Chao, et al. Autoencoder inspired unsupervised feature selection[C]//Proc of the 43rd IEEE Int Conf on Acoustics, Speech and Signal Processing (ICASSP). Piscataway, NJ: IEEE, 2018: 2941−2945

[14] Doquet G, Sebag M. Agnostic feature selection[C]//Proc of European Conf on Machine Learning and Knowledge Discovery in Databases. Berlin: Springer, 2020: 343−358

[15] Abid A, Balin M F, Zou J. Concrete autoencoders for differentiable feature selection and reconstruction[J]. arXiv preprint, arXiv: 1901.09346, 2019

[16] Shaham U, Lindenbaum O, Svirsky J, et al. Deep unsupervised feature selection by discarding nuisance and correlated features[J]. Neural Networks, 2022, 152: 34−43

[17] Wu Xiaotong, Cheng Qiang. Fractal autoencoders for feature selection[C]//Proc of the 35th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2021: 10370−10378

[18] Gong Xiaoling, Yu Ling, Wang Jian, et al. Unsupervised feature selection via adaptive autoencoder with redundancy control[J]. Neural Networks, 2022, 150: 87−101

[19] Zhang Yijie, Lu Zhaoyu, Wang Shuihua. Unsupervised feature selection via transformed auto-encoder[J]. Knowledge-Based Systems, 2021, 215: 106748

[20] Ling Yuunzhi, Nie Feiping, Yu Weichang, et al. Discriminative and robust auto-encoders for unsupervised feature selection[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 36(1): 1622−1636

[21] Hassanieh W, Chehade A. Selective deep autoencoder for unsupervised feature selection[C]//Proc of the 38th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2024: 12322−12330

[22] Singh D, Climente-Gonzalez H, Petrovich M, et al. FSNet: Feature selection network on high-dimensional biological data[C]//Proc of the 30th Int Joint Conf on Neural Networks (IJCNN). Piscataway, NJ: IEEE, 2023: 1−9

[23] Natsumeda M, Yairi T. Feature selection with partial auto-encoding for zero-sample fault diagnosis[J]. IEEE Transactions on Industrial Informatics, 2024, 20(2): 2144−2153

[24] Tuan P M, Trung N L, Adel M, et al. Autoencoder-based feature ranking for predicting mild cognitive impairment conversion using FDG-PET images[C]//Proc of the 36th IEEE Statistical Signal Processing Workshop (SSP). Piscataway, NJ: IEEE, 2023: 720−724

[25] Sun He, Ren Jinchang, Zhao Huimin, et al. Novel Gumbel-softmax trick enabled concrete autoencoder with entropy constraints for unsupervised hyperspectral band selection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1−13

[26] Ghosh T, Kirby M. Nonlinear feature selection using sparsity-promoted centroid-encoder[J]. Neural Computing and Applications, 2023, 35: 21883−21902

[27] Wang Rong, Bian Jintang, Nie Feiping, et al. Nonlinear feature selection neural network via structured sparse regularization[J]. IEEE

Transactions on Neural Networks and Learning Systems, 2023, 34(11): 9493−9505

[28] Zhu Zhuangdi, Lin Kaixiang, Jain A K, et al. Transfer learning in deep reinforcement learning: A survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(11): 13344−13362

[29] Malik A K, Ganaie M A, Tanveer M, et al. Alzheimer's disease diagnosis via intuitionistic fuzzy random vector functional link network[J]. IEEE Transactions on Computational Social Systems, 2022, 11(4): 4754−4765

[30] Ganaie M A, Sajid M, Malik A K, et al. Graph embedded intuitionistic fuzzy random vector functional link neural network for class imbalance learning[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(9): 11671−11680

[31] Sajid M, Malik A K, Tanveer M. Intuitionistic fuzzy broad learning system: Enhancing robustness against noise and outliers[J]. IEEE Transactions on Fuzzy Systems, 2024, 32(8): 4460−4469

[32] Chen Yuepeng, Ding Weiping, Huang Jiashuang, et al. Multigranularity fuzzy autoencoder for discriminative feature selection in high-dimensional data[J]. IEEE Transactions on Neural Networks and Learning Systems, 2025, 36(9): 17433−17447

[33] Ceylan T. Intuitionistic fuzzy eigenvalue problem[J]. An International Journal of Optimization and Control: Theories & Applications, 2024, 14(3): 220−228

[34] Chen Yuepeng, Ding Weiping, Ju Hengrong, et al. Cascaded two-stage feature clustering and selection via separability and consistency in fuzzy decision systems[J]. IEEE Transactions on Fuzzy Systems, 2024, 32(9): 5320−5333

[35] Atanassov K T. Intuitionistic Fuzzy Sets[M]. Heidelberg: Springer, 1999

[36] Xiao Fuyuan. A distance measure for intuitionistic fuzzy sets and its application to pattern classification problems[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2021, 51(6): 3980−3992

[37] Hamo Y, Markovitch S. The COMPSET algorithm for subset selection[C]//Proc of the 19th Int Joint Conf on Artificial Intelligence (IJCAI). Edinburgh: Professional Book Center, 2005: 728−733

[38] Wen Long, Gao Liang, Li Xinyu. A new deep transfer learning based on sparse auto-encoder for fault diagnosis[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2017, 49(1): 136−144

[39] Kandaswamy C, Silva L M, Alexandre L A, et al. Improving transfer learning accuracy by reusing stacked denoising autoencoders[C]// Proc of the 47th IEEE Int Conf on Systems, Man, and Cybernetics. Piscataway, NJ: IEEE, 2014: 1380−1387

[40] Li Jundong, Cheng Kewei, Wang Suhang, et al. Feature selection: A data perspective[J]. ACM Computing Surveys, 2017, 50(6): 1−45

[41] Statnikov A, Tsamardinos I, Dosbayev Y, et al. GEMS: A system for automated cancer diagnosis and biomarker discovery from microarray gene expression data[J]. International Journal of Medical Informatics, 2005, 74(7/8): 491−503

[42] Friedman M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance[J]. Journal of the American Statistical Association, 1937, 32(200): 675−701

[43] Dunn O J. Multiple comparisons among means[J]. Journal of the American Statistical Association, 1961, 56(293): 52−64

[44] Sheldon M R, Fillyaw M J, Thompson W D. The use and interpretation of the Friedman test in the analysis of ordinal-scale data in repeated measures designs[J]. Physiotherapy Research International, 1996, 1(4): 221−228

[45] Hu Ming, Tsang E C C, Guo Yuhua, et al. Fast and robust attribute reduction based on the separability in fuzzy decision systems[J]. IEEE Transactions on Cybernetics, 2021, 52(6): 5559−5572

[46] Demšar J. Statistical comparisons of classifiers over multiple data sets[J]. Journal of Machine Learning Research, 2006, 7: 1−30

[47] Sun Lin, Yin Tengyu, Ding Weiping, et al. Feature selection with missing labels using multilabel fuzzy neighborhood rough sets and maximum relevance minimum redundancy[J]. IEEE Transactions on Fuzzy Systems, 2022, 30(5): 1197−1211

[48] Huang Jiashuang, Zhu Qi, Hao Xiaoke, et al. Identifying resting-state multifrequency biomarkers via tree-guided group sparse learning for schizophrenia classification[J]. IEEE Journal of Biomedical and Health Informatics, 2018, 23(1): 342−350

[49] Zhu Qi, Li Hujie, Huang Jiashuang, et al. Hybrid functional brain network with first-order and second-order information for computer-aided diagnosis of schizophrenia[J]. Journal of Neuroscience, 2019, 13: 603

[50] Lynall M E, Bassett D S, Kerwin R, et al. Functional connectivity and brain networks in schizophrenia[J]. Journal of Neuroscience, 2010, 30(28): 9477−9487

[51] Nestor P G, Forte M, Ohtani T, et al. Faulty executive attention and memory interactions in schizophrenia: Prefrontal gray matter volume and neuropsychological impairment[J]. Clinical EEG and Neuroscience, 2020, 51(4): 267−274

[52] Friedman N P, Robbins T W. The role of prefrontal cortex in cognitive control and executive function[J]. Neuropsychopharmacology, 2022, 47(1): 72−89

[53] Fujii M, Maesawa S, Ishiai S, et al. Neural basis of language: An overview of an evolving model[J]. Neurologia Medico-Chirurgica, 2016, 56(7): 379−386

[54] Li Yuntao, Mateos G. Learning to model the relationship between brain structural and functional connectomes[J]. IEEE Transactions on Signal and Information Processing over Networks, 2022, 8: 830−843



Hua Qichong, born in 2001. Master candidate. His main research interests include fuzzy theory and deep learning.

华启冲, 2001 年生。硕士研究生。主要研究方向为模糊理论、深度学习。



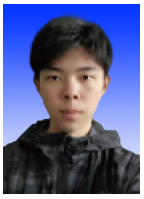
Ding Weiping, born in 1979. PhD, professor, PhD supervisor. Distinguished member of CCF. His main research interests include data mining, multimodal machine learning, multigranulation computing, evolutionary computation, and medical big data analysis.

丁卫平, 1979 年生。博士，教授，博士生导师。CCF 杰出会员。主要研究方向为数据挖掘、多模态机器学习、多粒度计算、演化计算、医学大数据分析。



Chen Yuepeng, born in 1991. PhD candidate. His main research interests include granular computing, machine learning, and data mining.

陈悦鹏, 1991 年生。博士研究生。主要研究方向为粒计算、机器学习、数据挖掘。



Zhou Tianyi, born in 2000. PhD candidate. His main research interests include fuzzy theory, granular computing, and deep learning.

周天奕, 2000 年生。博士研究生。主要研究方向为模糊理论、粒度计算、深度学习。