



本科毕业论文（设计）

UNDERGRADUATE THESIS (PROJECT)

题 目 无监督特征选择的非凸
建模和优化

学 院 机电工程与自动化
专 业 自动化
学 号 20121834
学生姓名 武乃琼
指导教师 修贤超
起讫日期 2024.1.15 ~ 2024.5.31



姓名: 武乃琼
论文题目: 无监督特征选择的非凸建模和优化

学号: 20121834

原创性声明

本人声明: 所呈交的论文是本人在指导教师指导下进行的研究工作。除了文中特别加以标注和致谢的地方外, 论文中不包含其他人已发表或撰写过的研究成果。参与同一工作的其他同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

签名: 武乃琼 日期: 2024.5.28

本论文使用授权说明

本人完全了解上海大学有关保留、使用学位论文的规定。即: 学校有权保留论文及送交论文复印件, 允许论文被查阅和借阅; 学校可以公布论文的全部或部分内容。

(保密的论文在解密后应遵守此规定)

签名: 武乃琼 指导教师签名: 何振东 日期: 2024.5.28

摘要

无监督特征选择是一项重要的数据处理技术，广泛应用于机器学习、模式识别、人工智能等领域。传统的无监督特征选择方法多是基于凸函数建模和优化的，然而随着大数据时代的到来，这类方法往往难以取得令人满意的性能。本文以主成分分析为例，通过引入不同的非凸函数探讨新模型和新方法，最终形成快速有效的无监督特征选择新范式。具体内容如下：

- (1) 针对无监督特征选择易出现过度拟合和缺乏结构性先验的问题，提出了一种结构稀疏非凸主成分分析框架，旨在将非凸函数（如对数、指数项比值、硬阈值）与结构稀疏相结合，在较大限度减少模型误差的同时保持局部结构性。进一步，利用交替方向乘子法设计了简单易用的一阶优化算法，并保障所产生的子问题均有闭式解。通过数值实验与现有无监督特征选择方法进行比较，验证了所提出方法的优越性，分析了不同非凸函数的性能表现。
- (2) 针对上述方法因数据分布不同可能会产生错误特征选择的问题，进一步提出了一种混合稀疏非凸主成分分析框架，旨在将非凸函数与结构稀疏、元素稀疏相结合，在改善模型泛化能力的同时提高鲁棒性。在算法方面，利用近端交替极小化法设计了优化策略，得到了行之有效的求解器。通过数值实验证了所提出方法的有效性以及在噪声数据集上的鲁棒性。

关键词：无监督特征选择；主成分分析；非凸建模；优化算法

ABSTRACT

As an important data processing technique, unsupervised feature selection is widely used in machine learning, pattern recognition, artificial intelligence and many other fields. Most of the traditional unsupervised feature selection methods are developed based on models with convex functions, however, with the approach of the big data era, it is often unsatisfactory for such methods to achieve acceptable performance. In this paper, we take the example of principal component analysis to explore new models and methods. By introducing various non-convex functions, a new paradigm of unsupervised feature selection has been formed in the end. The specifically contents are as follows:

- (1) To overcome the problem of unsupervised feature selection, which can be prone to overfitting and lack of structural prior, a structurally sparse non-convex principal component analysis framework has been proposed. This framework aims to combine non-convex functions (logarithmic, exponential term ratios and hard thresholds) with structural sparsity, preserving structural properties while minimising the model error to a large extent. The superiority of the proposed method is verified through numerical experiments comparing with existing unsupervised feature selection methods, and the performance of various non-convex functions is analysed.
- (2) To overcome the problem of the above methods, which may result in wrong feature selection due to different data distributions, a hybrid sparse non-convex principal component analysis framework has been further proposed, aiming at combining non-convex functions with structural sparsity and elemental sparsity to improve the robustness while improving the generalisation ability of the model. In terms of the algorithm, an optimisation strategy is designed using the proximal alternating minimisation method, and a well-established solver is obtained. Through numerical experiments, the effectiveness of the proposed method and its robustness on noisy datasets are verified.

Keywords: Unsupervised feature selection; Principal component analysis; Non-convex modeling; Optimisation algorithms

目 录

摘要	I
ABSTRACT	i
第一章 绪论	1
1.1 研究背景与意义	1
1.2 研究现状	3
1.3 内容结构安排	5
第二章 相关原理介绍	7
2.1 主成分分析法	7
2.1.1 主成分分析法概述	7
2.1.2 稀疏主成分分析法	8
2.2 优化算法	10
2.2.1 交替方向乘子法	10
2.2.2 近端交替极小化法	11
第三章 基于结构稀疏的主成分分析方法研究	12
3.1 数学建模	12
3.2 优化算法	13
3.3 数值实验	16
3.3.1 数据集介绍	16
3.3.2 实验数据配置	18
3.3.3 性能指标测试	19
3.3.4 噪声干扰测试	22
3.4 本章小结	28
第四章 基于混合稀疏的主成分分析方法研究	29
4.1 数学模型	29
4.2 优化算法	30
4.3 数值实验	32
4.3.1 实验数据配置	32
4.3.2 性能指标测试	32
4.3.3 噪声干扰测试	36

4.4 本章小结	41
第五章 总结与展望	42
5.1 总结	42
5.2 展望	43
插图索引	44
表格索引	45
参考文献	46
致 谢	49

第一章 绪论

1.1 研究背景与意义

特征选择，又称属性选择 (Feature Selection)，是由 Liu 和 Motoda 等人^[1]在 1998 年提出的，是数据处理和机器学习中非常重要的步骤之一。其目的是从原始数据集中选择出对模型训练和预测最有用的一部分特征，或称为变量、属性等，来提高模型的性能和效率。特征选择流程示例如图1.1所示。

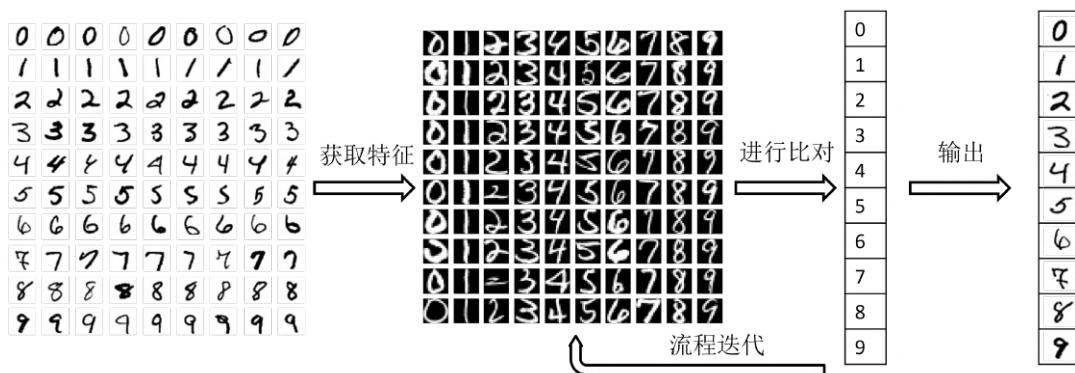


图 1.1 特征选择示例

通常来说，特征选择技术的目标包括简化模型使得研究人员或用户能更轻松的解释它们，缩短数据的培训时间，通过减少过度拟合来增强泛化等。特征选择不仅能降低数据的维度，促进数据的可视化与理解，还能使模型泛化能力加强，这些优势使得其成为了一个备受关注的研究方向。在过去几十年中，研究人员提出了许多特征选择方法^[2]。其中包括根据数据集中的信息可分由 Kotsiantis 等人^[3]提出的有监督式特征选择、由 Sheikhpour 等人^[4]提出的半监督式特征选择以及由 Alelyani 等人^[5]提出的无监督式特征选择。

有监督特征选择需要一组标注数据或称其为有监督数据集，来识别和选择相关的特征。常见的有监督特征选择方法包括有皮尔逊相关系数，该算法计算每个特征与目标变量之间的线性相关性，选择相关性较高的特征。递归特征消除使用一个基础学习模型训练数据，评估各特征的重要性，递归地删除重要性最低的特征。前向选择方法从空特征集开始，逐步添加对模型效果提升最大的特征。这些方法基本各有优缺点，如何对方法进行选择通常需要考虑数据特性、任务需求和计算资源等诸多因素。

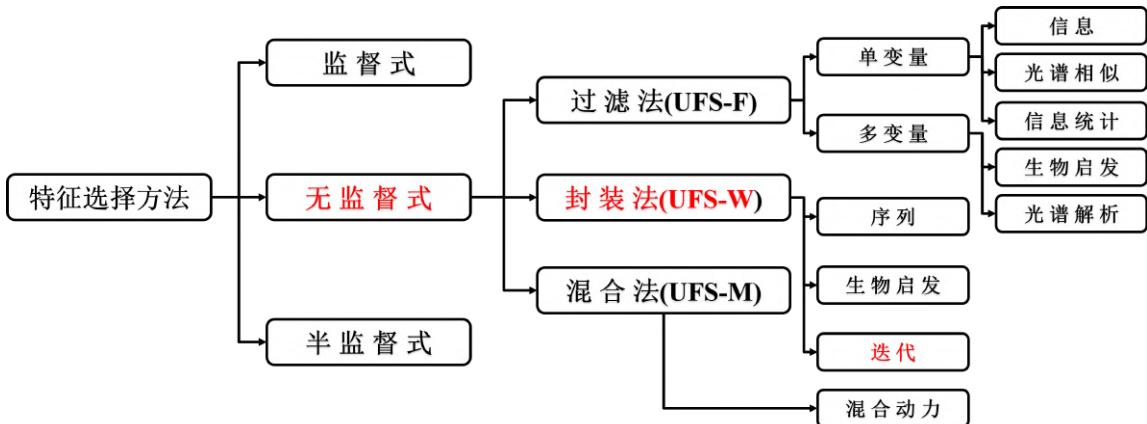


图 1.2 特征选择方法分类

半监督方法则只要求对某些对象进行标记，其结合了有监督特征选择和无监督特征选择的优势，并将其用于提升自身的特征选择的效果。这种方法通过同时利用有标签和无标签数据的统计特性来选择特征。其中，常见的半监督特征选择方法包括半监督评分，该方法将有标签数据的相关性评分与无标签数据的方差评分结合起来，为每个特征分配一个综合评分，然后选择评分最高的特征。半监督支撑向量机(Support Vector Machine, SVM)结合SVM模型，通过引入无标签数据的图正则化或流形正则化，优化特征选择。协同训练方法利用有标签数据训练两个分类器，然后用每个分类器为无标签数据打标签，更新特征选择结果。这些半监督特征选择方法通过结合有标签和无标签数据，可以在标签稀缺的情况下仍然有效地选择特征，提高模型的性能和泛化能力。

根据Guyon等人^[6,7]提出的无监督特征选择(Unsupervised Feature Selection, UFS)在数据预处理中具有两个显著优势。首先，该方法即使在没有先验知识的情况下也能表现良好，称其为具有无偏性。这意味着即使不依赖于任何预先标注的数据，该方法也能够在各种未知和复杂的数据集上有效地应用。其次，无监督特征选择方法相较于有监督特征选择方法显著降低了数据过拟合的风险。过拟合是指模型在训练数据上表现良好，却在新数据上表现较差的现象。由于无监督特征选择方法不依赖于标签信息，它们能够更好地避免这种情况，从而提升模型的泛化能力^[8]。

在大数据时代，由于数据量开始呈指数级增长的同时复杂性和多样性也在不断增加。在庞大的数据集面前如何提取关键特征对于实现有效的数据分析和建模变得至关重要。由于传统的特征选择方法通常依赖于数据标签，在实际应用会较为受限，其数据标记的获取成本高昂或难以获得。而无监督特征选择技术无需依赖数据标签，它通过分析数据内部的结构和特性，自动寻找数据的内在关联，从而达到降维和简化数据的目的。事实上，关于无监督特征选择的研究是当今数据科学中一项重要的课题，极具学术意义和实用价值。

1.2 研究现状

无监督特征选择流程如图1.4所示。根据文献^[9]，无监督特征选择通常可以分为以下三种主要方法：

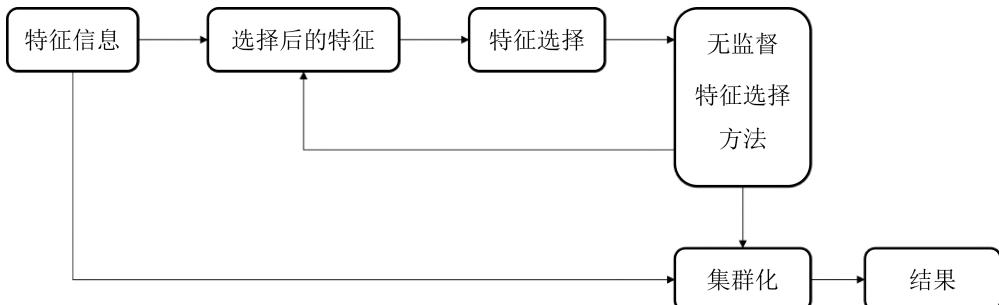


图 1.3 无监督特征选择流程

- (1) **过滤法^[5]**: 通过数据本身选取最相关的特征，即根据数据的内在属性对特征进行评估，其不使用任何聚类算法来引导相关特征的搜索。该方法的优势是速度快、可扩展性强。其中，He 等人^[10]提出的拉普拉斯分数方法（Laplacian Score of Feature Selection）是最为经典的过滤方法之一。其能够根据保留原始数据固有结构的能力，分别计算每个特征的拉普拉斯分数，特征分数越小便说明其局部保持力就越强，越有代表性。Du 等人^[11]提出的矩阵分解鲁棒无监督特征选择方法（Robust Unsupervised Feature Selection, RUFS）则是在某范式下进行鲁棒判别特征选择和鲁棒聚类的同时保留数据的局部流形结构。
- (2) **封装法^[12]**: 无监督特征选择封装法迭代算法通过反复迭代更新特征子集，从而逐步优化特征选择的性能。这种方法的优势是能够找到有助于提高聚类算法结果质量的特征子集。其中，Guo 等人^[13]提出嵌入式无监督特征选择方法（Embedded Unsupervised Feature Selection, EUFS），该方法拥有相同的目标函数，而不同之处在于最终模型的损失函数用 Frobenius 范数来代替 ℓ_2 范数，并采用 K 均值聚类算法更新参数迭代进行，直到模型收敛。

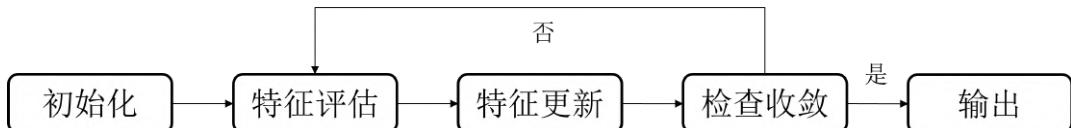


图 1.4 封装迭代算法

- (3) **混合法**: 通过结合过滤和封装方法的优点，在过滤法与封装法的结合之下取得良好的平衡。混合方法在保持效率的同时，尽可能提高相关目标任务的质量。其中，Li 等人^[14]利用非负谱分析获取更准确的聚类标记指标，同时将聚

类标记与特征选择矩阵实现联合迭代学习，最终研究出非负判别特征选择方法 (Nonnegative Discriminant Feature Selection, NDFS)。Yang 等人^[15]则设计出无监督判别的无监督算法 (Unsupervised Discriminative Feature Selection, UDFS) 把判别分析和 $\ell_{2,1}$ 范数最小化结合形成联合框架，优化筛选出在批处理的模式下最具辨别力的特征集合。

由此可见，当下无监督特征选择是各领域的研究热点和难点，因而学术界和工业界对无监督特征选择方法的研究十分广泛，如图像处理^[16,17]、信号处理^[18-20]和机器学习^[21,22]领域等。在这些海量高维数据中存在着大量的冗余数据^[23]和噪声^[24]，这些问题使得进行任务时不仅会大大增加计算机的内存负荷和处理时间，从而这会严重影响数据分析并经常导致偏差，甚至出现特征选择报错。

换言之，无监督特征选择技术在数据分析和机器学习等诸多领域提供了一种有效的优化解决方案。其通过去除数据中的冗余特征并选择出最具代表性的特征，提高了模型的性能和效率，帮助数据科学家和分析师们应用更先进的算法工具来处理大数据和复杂数据集。这些优势为无监督特征选择技术在未来的研究和应用中提供了广阔的前景和发展空间。

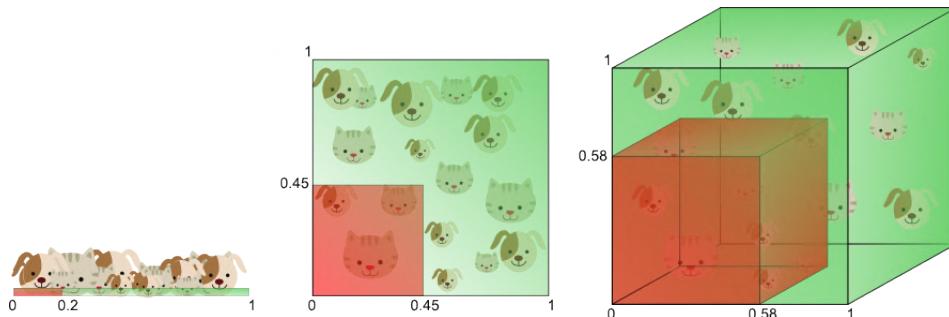


图 1.5 应用场景示意图^[25]

例如，聚类任务的目标是将数据集划分为若干不同的组别，使得同一组别内的数据具有相似性，而不同组别的数据存在显著差异。然而，数据通常包含大量特征，其中有些可能是冗余的、无关紧要的，甚至可能是噪声。这种情况下，直接使用所有特征进行聚类可能会导致算法性能不佳。而无监督特征选择技术不仅可以提高聚类算法的计算效率，还能增强算法对数据内在结构的敏感度。通过选择与聚类目标相关的特征，更加准确地将数据划分为不同的组别，从而揭示数据之间的关系和模式。

在图像识别领域，无监督特征选择能够协助识别和提取图像中最重要的特征，从而提高分类模型的准确性和效率。传统的图像分类模型通常依赖于人工选择的特征或使用预训练的深度学习模型，而这些方法往往都无法充分利用图像数据的多样性和复杂性。无监督特征选择技术则能够通过自动分析图像中的特征，找到最具代表性的特征，有助于减少特征冗余，提高模型的性能。该方法对大规模图像数据集尤

为适用，其有助于加快训练和推理过程。

在无线通信领域，无监督特征选择可以通过分析信号的特征，如频率、相位、幅度、调制等方式，识别出对通信系统性能至关重要的特征。这种自动化分析可以帮助通信系统在信号检测与分离、信号压缩与编码、信道估计与均衡、干扰抑制与滤波等方面取得改进。

在异常检测领域，无监督特征选择技术能够帮助识别数据中的异常点，从而提高异常检测的准确性和效率。通过帮助识别数据中的异常点，能够确定对检测异常最为重要的特征，该技术可以提高异常检测的准确性和效率，为各种领域的异常检测任务提供有力支持。

无监督学习中的特征选择与提取技术正经历着快速发展，如今也正面临着许多挑战和机遇。其未来的发展方向可能包括设计新的算法和框架，以在大规模数据集上提高计算效率和性能等。与此同时，在处理复杂数据集和结构时，非凸优化方法可以提供更大的灵活性和准确性。未来的研究可能更加关注如何在非凸模型中实现对稀疏特征的有效选择，以及如何结合特征之间的结构信息进行更精确的特征选择。这将有助于提高模型的可解释性和泛化能力。随着无监督特征选择在各个领域的深入研究，这一领域有望带来新的理论突破和实际应用创新。

1.3 内容结构安排

本文基于无监督特征选择模型建立及对其性能进行深入探讨，通过对所提出模型进行测试与优化，旨在开发一种在应用层面上性能更优的特征选择新模型。大致的结构安排如下：

第一章为绪论。简要介绍了无监督特征选择研究的背景与意义，概括了目前的研究现状与应用，并阐明了本文的内容安排。

第二章介绍理论基础。本章讲述了无监督特征选择的相关理论基础，详细讨论了交替方向乘子法与近端交替极小化算法的理论基础和算法流程，探究了实验部分所使用的两种优化算法的基本原理，并讨论了非凸优化在无监督特征选择模型中的作用与优势。

第三章通过向主成分分中引入非凸模型，提出了一种基于结构稀疏的无监督特征选择新方法。随后针对建立的新方法的准确率和归一化互信息进行了测试。通过在多个数据集上进行实验，将其与其他特征选择方法进行比较，评估该方法在不同情景下的性能表现并探讨其优缺点以及适用性。此外，对该方法在面对噪声数据时的表现进行了测试。通过引入不同噪声，验证了该方法对噪声的抵抗能力相对较强。

第四章通过向前述模型中引入元素稀疏系数，提出了一种基于混合稀疏的无监

督特征选择新方法。随后对该算法进行了准确率和归一化互信息的测试，评估其在不同数据集上的性能表现，并将其与上一章节以及现有优化算法进行了性能对比。对其在面对噪声数据时的鲁棒性进行了测试，以探究其在复杂环境下的应用潜力和可靠性。

第五章为整篇论文的内容总结与对未来研究的展望。

第二章 相关原理介绍

本章主要介绍了无监督特征选择算法的相关知识，学习了主成分分析以及稀疏主成分分析算法的流程并详细说明了其运行原理。随后讲解了交替方向乘子法与近端交替极小化法的基本运算方法，分析其在不同应用场景中的重要性和实际意义。通过对所应用的理论进行介绍，为后续章节奠定了理论基础。

2.1 主成分分析法

2.1.1 主成分分析法概述

主成分分析方法(Principal Component Analysis, PCA)作为一种由 Pearson 提出^[26]的经典降维方法，其在手写邮政编码分类^[27]、人脸识别^[28]、特征基因分析^[29]等诸多领域都得到了广泛应用。它通过对数据的协方差矩阵进行特征值分解，提取数据中的主要成分，从而达到将高维数据映射到低维空间的目的。

不妨设原始数据集为 $X \in \mathbb{R}^{m \times n}$ ，其中 m 表示样本数量， n 表示特征数量。协方差矩阵 S 的定义为

$$S = \frac{1}{m} X^T X. \quad (2.1)$$

PCA 的目标是找到一个投影矩阵 W ，使得投影后的数据的协方差矩阵 S 的迹达到最大化，通常将该方法写为

$$\begin{aligned} & \max_W \text{Tr}(W^T SW) \\ & \text{s.t. } W^T W = I. \end{aligned} \quad (2.2)$$

其中，最大化 $\text{Tr}(W^T SW)$ 的目的是在投影后的空间中最大化数据的方差，从而保留尽可能多的原始数据信息。约束条件 $W^T W = I$ 表示投影矩阵 W 是正交矩阵，即保持数据之间的正交性。根据极大极小的转换，上式也可以等价地写为

$$\begin{aligned} & \min_W -\text{Tr}(W^T SW) \\ & \text{s.t. } W^T W = I. \end{aligned} \quad (2.3)$$

通过引入拉格朗日乘子法来求解这个问题。不妨设拉格朗日乘子 Λ ，定义拉格朗日函数为

$$L(W, \Lambda) = \text{Tr}(W^T SW) + \text{Tr}(\Lambda(W^T W - I)). \quad (2.4)$$

分别对 W 和 Λ 求偏导并令其为零，能得到

$$\frac{\partial L}{\partial W} = 2SW - 2\Lambda W = 0, \quad (2.5)$$

$$\frac{\partial L}{\partial \Lambda} = W^T W - I = 0, \quad (2.6)$$

由式(2.4)能够解得

$$SW = \Lambda W. \quad (2.7)$$

这意味着 W 是 S 的特征向量， Λ 是相应的特征值矩阵。由于 W 是正交矩阵，所以有 $W^T W = I$ ，满足了约束条件。由此，主成分分析法可以总结为以下流程。

Algorithm 1 主成分分析法优化过程

输入：矩阵 X

输出：降维后的数据 $Z = XW$

- 1: 计算数据矩阵的协方差矩阵 $S = \frac{1}{n}X^T X$
 - 2: 进行特征值分解，得到特征值 λ_i 和特征向量 v_i
 - 3: 按照特征值从大到小排序特征向量
 - 4: 选择前 k 个特征向量构建投影矩阵 W
 - 5: 使用投影矩阵 W 对数据进行降维
-

主成分分析的优势包括能够有效地将高维数据降维为低维空间的同时保留数据中的主要信息，有助于简化数据结构，使后续的分析和处理更加高效。将高维数据映射到二维或三维空间，使得数据的内在结构和分布模式能够更直观地呈现出来，有助于发现数据中的规律和异常点。当前主成分分析算法的研究在多个方向上都取得了进展，研究人员致力于开发更高效的数学模型，应对大规模数据处理的挑战。该方法的研究已不仅集中在算法的改进和应用拓展上，还涉及到了对算法理论的深入探索和理解，其能够为主成分分析算法在各个领域的应用提供了更加丰富和强大的支持。

2.1.2 稀疏主成分分析法

稀疏优化是现代数据科学和机器学习领域中的重要技术^[30]，它有助于降低模型的复杂性和存储需求，同时提高了模型的解释性和泛化能力。通过引入稀疏性约束，模型在学习过程中会倾向于选择较少的特征，从而实现更高效的特征选择。这种方

法不仅在一定程度上降低了模型的计算成本和内存占用，还增强了模型的可解释性和稳定性。在处理大规模、高维度的数据时，稀疏优化技术尤其有用。

稀疏优化的核心目标是寻找能够解释或预测数据的模型，这些模型往往都具有较少的参数或特征。通过保持模型的稀疏性，可以实现更好的泛化能力、降低过拟合的风险，并提高模型的可解释性。其中常见的稀疏优化技术^[31]包括 ℓ_1 正则化 (Lasso) 和 ℓ_2 正则化 (Ridge)^[32]。正则化方法通过在损失函数中加入惩罚项，使得参数趋向于零。 ℓ_1 正则化主要针对稀疏性，因为它能够将一些系数压缩到零，而 ℓ_2 正则化则主要用于防止模型过拟合。

基于稀疏优化的思想，稀疏主成分分析^[32] (Sparse Principal Component Analysis, SPCA) 应运而生。传统的主成分分析通过线性变换将数据投影到低维空间，以捕捉数据中的重要特征。然而，主成分分析方法中主成分通常是所有原始变量的线性组合，其难以解释和应用于高维数据。由此，稀疏主成分分析通过在传统主成分分析的基础上引入稀疏性约束，使得每个主成分只依赖于少数几个原始变量，从而提高了结果的可解释性和模型的实用性。稀疏主成分分析结合了稀疏优化和主成分分析的优点，在保留数据的主要信息的同时，又实现了变量选择。在处理高维数据集时，展现出了巨大的应用潜力。

Algorithm 2 稀疏主成分分析优化过程

输入: 数据矩阵 X , 正则化参数 λ

输出: 投影矩阵 W

迭代 $i = 1 \sim k$:

- 1: 计算梯度 $\nabla_W = 2SW - \lambda \text{sign}(W)$
- 2: 使用梯度下降法更新 W : $W \leftarrow W + \eta \nabla_W$
- 3: 应用软阈值以引入稀疏性: $W \leftarrow \text{SoftThreshold}(W, \lambda)$

停止

主成分归一化: $W = \frac{W}{\|W\|_2}$

稀疏主成分分析的目标是找到一个稀疏的投影矩阵 W ，使得投影后的数据具有最大的方差，即最大程度地保留原始数据的信息^[33] 的同时具有尽可能少的非零元素。为了引入稀疏性约束，通常会在求解(2.3)优化目标中加入 ℓ_1 正则化项^[34]，便能够得到下式

$$\begin{aligned} \min_W \quad & -\text{Tr}(W^T SW) + \lambda \|W\|_1 \\ \text{s.t.} \quad & W^T W = I, \end{aligned} \tag{2.8}$$

其中, λ 是正则化参数, $\|W\|_1$ 表示 W 的 ℓ_1 范数。

通过优化上述目标函数, 可以得出稀疏主成分分析 (SPCA) 的结果, 即稀疏的投影矩阵 W 。

2.2 优化算法

2.2.1 交替方向乘子法

近年来, 交替方向乘子法 (Alternating Direction Method of Multipliers, ADMM) 在优化领域崭露头角。该方法^[35] 通过将原始问题分解为多个子问题, 并不断迭代的方式逐步逼近最优解。该方法可以有效地解决约束条件复杂、目标函数非凸等问题, 展现出强大的优化能力。

不妨设其数学模型为

$$\begin{aligned} \min_{X,Z} \quad & F(X) + G(Z) \\ \text{s.t.} \quad & AX + BZ = C, \end{aligned} \tag{2.9}$$

其中 $F(X)$ 和 $G(Z)$ 是凸函数, A 和 B 是矩阵。

Algorithm 3 交替方向乘子法优化过程

输入: 矩阵 X, Z, U , 参数 ρ

初始化: 矩阵 X^0, Z^0, U^0

迭代: $k = 1, 2, 3\dots$

1: 更新 X

$$X^{k+1} = \arg \min_X F(X) + \frac{\rho}{2} \|AX + BZ^k - C + U^k\|_F^2 \tag{2.10}$$

2: 更新 Z

$$Z^{k+1} = \arg \min_Z G(Z) + \frac{\rho}{2} \|AX^k + BZ - C + U^k\|_F^2 \tag{2.11}$$

3: 更新 U

$$U^{k+1} = U^k + \rho(AX^{k+1} + BZ^{k+1} - C) \tag{2.12}$$

引入拉格朗日乘子 U , 构建增强拉格朗日函数如下

$$L_\rho(X, Z, U) = F(X) + G(Z) + U^T(AX + BZ - C) + \frac{\rho}{2} \|AX + BZ - C\|_F^2, \tag{2.13}$$

其中, $\rho > 0$ 是罚参数, $\|\cdot\|_F^2$ 表示矩阵的 Frobenius 范数。通过交替最小化拉格朗日乘子函数来求解问题, 即算法3所示。

交替方向乘子法的优势包括可以将原本难以求解的高维非凸优化问题分解为多个易处理的低维子问题从而使每个子问题都可以高效地求解, 子问题可以并行计算使之显著提高了计算效率, 能够有效地处理实际应用中的噪声和异常值问题等。因此, 交替方向乘子法在处理大规模优化问题时表现出优越的性能, 尤其在模式识别、图像处理、机器视觉等领域。在这些领域中, 稀疏优化问题会涉及到非凸的稀疏正则化项, 导致其直接求解困难, 而交替方向乘子法通过引入辅助变量和分解技术, 将复杂问题分解为易处理的子问题, 从而高效地找到近似解。

2.2.2 近端交替极小化法

近端交替极小化法^[36] (Proximal Alternating Minimization, PAM) 也称为近端迭代算法^[37] 或近端点算法^[38], 其算法核心思路是通过在原目标函数上添加一个二次项使其变得严格凸函数。给定一组变量 (X_1, X_2, \dots, X_n) 和目标函数 $F(X_1, X_2, \dots, X_n)$, 在 PAM 框架下的优化问题为

$$\left\{ \begin{array}{l} X_1^{k+1} = \underset{X_1}{\operatorname{argmin}} F(X_1, X_2^k, \dots, X_n^k) + \frac{\rho_1}{2} \|X_1 - X_1^k\|_F^2, \\ X_2^{k+1} = \underset{X_2}{\operatorname{argmin}} F(X_1^{k+1}, X_2, \dots, X_n^k) + \frac{\rho_2}{2} \|X_2 - X_2^k\|_F^2, \\ \vdots \\ X_n^{k+1} = \underset{X_n}{\operatorname{argmin}} F(X_1^{k+1}, X_2^{k+1}, \dots, X_n) + \frac{\rho_n}{2} \|X_n - X_n^k\|_F^2. \end{array} \right. \quad (2.14)$$

其中 $\rho_i (i = 1, \dots, n)$ 是大于 0 的近端参数。

近端交替极小化法被广泛应用于各种机器学习和优化任务中, 特别是在大规模数据集下或者具有复杂结构的问题中, 如图像处理、信号处理^[39]、机器学习模型训练等领域。但对于非凸优化问题, 其收敛性证明比较复杂, 往往需要用到函数的 KL 性质。

第三章 基于结构稀疏的主成分分析方法研究

本章针对主成分分析模型解释效果差、难于处理高维数据等问题，通过引入非凸函数构建了一种基于结构稀疏的无监督特征选择新方法。该方法的算法流程应用了交替方向乘子法，随后便测试了其在实际数据集上的性能指标。最后通过像数据集中引入不同程度的噪声，验证了该方法在噪声环境下的抗噪性能和稳定性。研究表明，该方法在处理高维复杂数据时具有较强的鲁棒性，为特征选择领域提供了一种有效的新方法。

3.1 数学建模

非凸框架下的结构稀疏算法研究旨在探索非凸优化框架下的全新结构稀疏算法。该研究的重点在于通过引入非凸正则化技术，使特征选择和模型优化过程更具灵活性和适应性，从而更好地适应实际应用中的复杂性和多样性。研究内容包括针对不同非凸函数的优化算法设计、稀疏结构的框架，以及对其在各类常见数据集^[40]中进行性能评估。

回顾第二章中的主成分分析（PCA）模型

$$\begin{aligned} \min_W & -\text{Tr}(W^T SW) \\ \text{s.t. } & W^T W = I, \end{aligned} \tag{3.1}$$

其中， W 为投影矩阵， I 为单位矩阵。为了得到稀疏性，可以通过引入稀疏主成分分析（SPCA）模型

$$\begin{aligned} \min_W & -\text{Tr}(W^T SW) + \lambda \|W\|_{2,1} \\ \text{s.t. } & W^T W = I, \end{aligned} \tag{3.2}$$

其中， $\|W\|_{2,1} = \sum_{i=1}^d \|w_i\|_2$ ， w^i 是矩阵 W 中的第 i 行。通过在式(3.1)中加入 $\|W\|_{2,1}$ ，SPCA 可以得到一个稀疏的投影矩阵，从而增强了方法可解释性并减轻了噪声干扰。

近些年，引入非凸优化的方法得到了迅速发展与研究，与凸优化相比，非凸优化提供了更多的可能性。为此，Li 等人^[41] 通过引入 $\ell_{2,p}$ ($0 < p < 1$) 范数法使投影矩阵更加稀疏，并提出了以下被命名为 SPCAFS 的模型

$$\begin{aligned} \min_W & -\text{Tr}(W^T SW) + \lambda \|W\|_{2,p}^p \\ \text{s.t. } & W^T W = I. \end{aligned} \tag{3.3}$$

显然, 当 $p = 1$ 时, 式(3.3)可被转化为式(3.2)。

不过上述方法有两个缺点: 首先, 其模型没有统一的非凸框架, 从而限制了该模型的应用。其次, 他们的优化流程都是在欧几里得空间进行的, 缺乏对流行结构研究。因此, 在本研究中, 我们提出了以下非凸特征选择 (Non-convex Feature Selection, NOFS) 模型

$$\begin{aligned} \min_W \quad & -\text{Tr}(W^T SW) + \lambda \sum_{i=1}^d \Phi(\|w^i\|_2) \\ \text{s.t.} \quad & W^T W = I. \end{aligned} \quad (3.4)$$

值得注意的是, Φ 表示一类非凸函数, 除了上述的 $\ell_{2,p}$ ($0 < p < 1$), 还包括对数函数、指数项比值函数、硬阈值函数等等。

3.2 优化算法

首先, 引入新变量 V 并将问题(4.1)变换为以下形式

$$\begin{aligned} \min_{W,V} \quad & -\text{Tr}(W^T SW) + \lambda \sum_{i=1}^d \Phi(\|v^i\|_2) \\ \text{s.t.} \quad & W^T W = I, W - V = 0. \end{aligned} \quad (3.5)$$

其对应的增强拉格朗日函数为

$$\begin{aligned} L_\beta(W, V, \Lambda) = & -\text{Tr}(W^T SW) + \lambda \sum_{i=1}^d \Phi(\|v^i\|_2) \\ & - \langle \Lambda, W - V \rangle + \frac{\beta}{2} \|W - V\|_F^2, \end{aligned} \quad (3.6)$$

其中, $\Lambda \in \mathbb{R}^{m \times n}$ 是引入的拉格朗日乘子, $\langle \cdot, \cdot \rangle$ 表示矩阵的内积, $\beta > 0$ 是衡量 W 和 V 之间距离的惩罚参数。

(1) 更新 W

当 V^k 和 Λ^k 固定不变时, 可以通过解决下述的优化问题来对 W 进行更新

$$\begin{aligned} \min_W \quad & -\text{Tr}(W^T SW) + \frac{\beta}{2} \|W - V^k - \Lambda^k / \beta\|_F^2 \\ \text{s.t.} \quad & W^T W = I. \end{aligned} \quad (3.7)$$

该优化问题是一个流形的优化问题。这里通过将目标函数从欧几里得空间转换到黎曼流形, 利用信赖域法进行求解。首先, 令 $\text{Gr}(d, m) = \{W \in \mathbb{R}^{d \times m} \mid W^T W = I\}$,

则式(3.7)在黎曼流形上的优化可以定义为

$$\min_{W \in \text{Gr}(d,m)} -\text{Tr}(W^T SW) + \frac{\beta}{2} \|W - V^k - \Lambda^k/\beta\|_F^2. \quad (3.8)$$

为简化流程，将上述问题的目标函数记作 $f(W)$ 。其梯度为

$$\nabla f(W) = -2SW + \beta(W - V^k - \Lambda^k/\beta). \quad (3.9)$$

Algorithm 4 更新 W 优化过程

输入：矩阵 S, Λ^k, V^k , 参数 $\beta, \xi, 0 < c < 1$

输出：投影矩阵 W

迭代：

1: 使用公式(3.15)计算黎曼梯度 η^k

2: 计算加权值

$$\beta^k = \text{Tr}(\eta^{kT} \cdot \eta^k) / \text{Tr}(\eta^{(k-1)T} \cdot \eta^{k-1}) \quad (3.10)$$

3: 计算传输方向

$$T_{W^{k-1} \rightarrow W^k}(\xi^{k-1}) = P_{W^k}(\xi^{k-1}) \quad (3.11)$$

4: 计算搜索方向

$$\xi^k = -\nabla_R f(W^k) + \beta^k T_{W^{k-1} \rightarrow W^k}(\xi^{k-1}) \quad (3.12)$$

5: 当 $t^k \geq 0$ 时, 使用反向追踪法计算步长

$$f(R_{W^k}(t^k \xi^k)) \geq f(W^k) + ct^k \text{Tr}(\eta^{kT} \xi^k) \quad (3.13)$$

6: 通过回缩计算下一次迭代

$$W^{k+1} = R_{W^k}(t^k \xi^k) \quad (3.14)$$

7: 当 $\|\eta^{k+1}\|_F \leq \varepsilon$ 时收敛, 停止迭代

在黎曼流形优化中, 计算空间可以被分为两个互补的空间, 即切空间和线空间。通过这种方式, 能够将欧几里得梯度投影到黎曼流形的切空间, 从而得到目标函数

的黎曼梯度。其运算可以转化为

$$\text{grad } f(W) = P_W(\nabla f(W)) = \nabla f(W) - W \text{sym}(W^T \nabla f(W)). \quad (3.15)$$

其中, $P_W(\nabla f(W))$ 表示欧几里得梯度在黎曼流形切空间上的投影, $\text{sym}(X) = (X + X^T)/2$ 表示从方阵 X 中提取对称部分。

得到上述黎曼梯度后, 便可以在黎曼流形上应用信赖域法来求式(3.7), 即算法4。在此处, R_W 表示 W 在黎曼流形上的收缩。

(2) 更新 V

当 W^{k+1} 和 Λ^k 固定不变时, 可以通过解决如下的优化问题来对 V 进行更新

$$\min_V \quad \frac{\beta}{2} \|W^{k+1} - \Lambda^k/\beta - V\|_F^2 + \lambda \sum_{i=1}^d \Phi(\|v^i\|_2). \quad (3.16)$$

令 $Y = W^{k+1} - \Lambda^k/\beta$ 。那么式(3.16)可以转化为

$$\min_V \quad \frac{\beta}{2} \|Y - V\|_F^2 + \lambda \sum_{i=1}^d \Phi(\|v^i\|_2), \quad (3.17)$$

进一步分解为一系列向量优化问题, 即

$$\min_{v_i} \quad \frac{\beta}{2} \|y^i - v^i\|_2^2 + \lambda \Phi(\|v^i\|_2). \quad (3.18)$$

与此同时, 问题(3.16)可以通过算法5进行求解。如果函数 Φ 在式 (3.18) 中选择硬阈值进行处理, 可以将相应的近邻算子直接定义为

$$\text{Prox}(\|y^i\|_2, T) = \begin{cases} \|y^i\|_2, & \text{如果 } \|y^i\|_2 \geq T \\ 0, & \text{其他,} \end{cases} \quad (3.19)$$

其中 $T = \sqrt{2\lambda/\beta}$ 。

Algorithm 5 更新 V 优化过程

输入: 矩阵 Y^k, T

输出: 矩阵 V

迭代: 当 $i \leq d$ 时

- 1: 如果 $y_i = 0$, 那么 $v_i = \{0\}$
- 2: 此外 $v_i = \text{Prox}(\|y^i\|_2, T) \cdot \frac{y^i}{\|y^i\|_2}$
- 3: $i = i + 1$

停止迭代

(3) 更新 Λ

当 W^{k+1} 和 V^{k+1} 固定不变时，可以通过解决下述的优化问题来对 Λ 进行更新

$$\Lambda^{k+1} = \Lambda^k - \beta(W^{k+1} - V^{k+1}). \quad (3.20)$$

综上所述，求解问题(4.1)的整个优化方案如算法6所示。在此之后，对选取的 $\|w^i\|^2$ ($i = 1, 2, \dots, d$) 进行降序排序，并选取其中的最优解来进行特征选择。

Algorithm 6 求解 NOFS 的优化算法

输入：矩阵 X , 参数 $\lambda, \beta, m, \varepsilon$

输出：投影矩阵 W

初始化： $H = I_n - \frac{1}{n}11_n^T, S = X^T H X$

迭代

- 1: 通过算法4更新 W^{k+1}
- 2: 通过算法5更新 V^{k+1}
- 3: 通过(3.20)更新 Λ^{k+1}
- 4: 如果满足下述条件则停止

$$\max\left\{\frac{\|W^{k+1} - W^k\|_F}{\|W^{k+1}\|_F}, \frac{\|V^{k+1} - V^k\|_F}{\|V^{k+1}\|_F}\right\} \leq \varepsilon \quad (3.21)$$

3.3 数值实验

3.3.1 数据集介绍

在本研究中，我们选用了三个常见的公开数据集来对该方法的性能进行测试和评估，其每个数据集在各自领域中都具有广泛的应用和代表性，详情见表3.1。

表 3.1 数据详情

	COIL20	Isolet	USPS
特征	1024	617	256
样本	1440	1560	1000
类别	20	26	10
类型	图像	语音	手写数字

- (1) **COIL20 数据集：** COIL20 (Columbia University Image Library) 数据集是计算机视觉和机器学习领域的经典数据集之一。它包含了 20 个不同对象在不同视角

下的图像，这些对象包括家具、玩具等各种日常物品，见图3.13。每个对象从不同角度拍摄了 72 张图像，总共有 1440 张图像。这些图像的特征通常被用来研究对象识别、分类等问题。



图 3.1 COIL20 数据集图像

- (2) **Isolet 数据集**: Isolet (Isolated Letter Speech Recognition) 数据集广泛用于机器学习和语音识别领域。它包含了英语字母表 26 个字母的发音语音数据，总共有 1560 个样本。每个样本由 617 个特征组成，这些特征包括了声音的频率、强度等信息。该数据集常用于训练和测试机器学习模型，尤其是语音识别任务。
- (3) **USPS 数据集**: USPS (United States Postal Service) 数据集是机器学习和图像识别领域的常用数据集之一。它包含手写数字 0-9 的图像数据，共有 1000 个样本，每个数字类别有 100 个样本，见图3.13。每个图像由 256 个特征组成也就是 16x16 像素的灰度图像。该数据集通常用于训练和测试机器学习模型，执行手写数字识别任务。

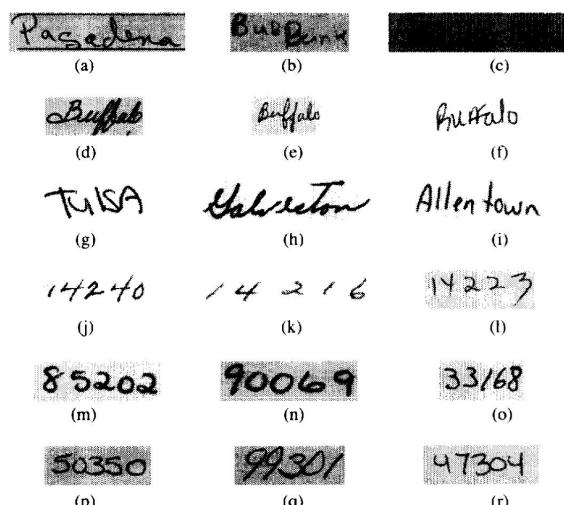


图 3.2 USPS 数据集图像

3.3.2 实验数据配置

为了说明所提方法的有效性，我们与几种基准无监督特征选择方法进行了综合比较，包括 LapScore^[42]、UFSOL^[43]、UDFS^[44]、SOGFS^[45]、SRCFS^[46]、RUFS^[47]、SPEC^[48]。

对于文章提出的新方法，通过选取以下三种不同的非凸函数^[49]:

(1) 对数 (Logarithm)^[50]

$$y_{log} = \frac{1}{\log(\gamma + 1)} \log(\gamma x + 1), \quad (3.22)$$

其中 $\gamma = 10e-5$.

(2) 指数项比值函数 (ETP)^[51]

$$y_{etp} = \frac{1}{1 - \exp(-\gamma)} (1 - \exp(-\gamma x)), \quad (3.23)$$

其中 $\gamma = 10e-6$.

(3) 硬阈值 (HARD)^[52]

$$y_{hard} = x \mathbf{I}\{|x| > T\}. \quad (3.24)$$

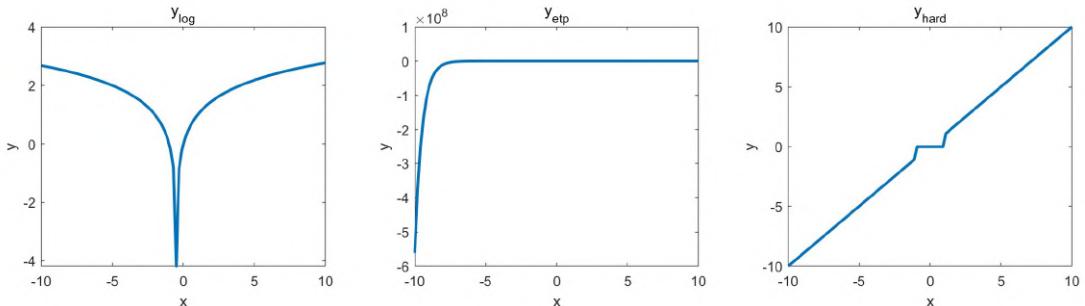


图 3.3 非凸函数图像

进而得到了三种不同的非凸优化模型，分别称为

- (1) 添加对数^[50] (LOG) 的 NOFS 模型 (NOFS-L)
- (2) 添加指数项比值函数^[51] (ETP) 的 NOFS 模型 (NOFS-E)
- (3) 添加硬阈值^[52] (HARD) 的 NOFS 模型 (NOFS-H)

为了公平起见，将聚类迭代次数设置为 50 次。对于不同的数据集， β 的选择略有不同，我们采用网格化方法进行搜索。此外，特征选择数量的范围是 25, 50, 75, 100, 125, 150, 175。

3.3.3 性能指标测试

(1) 准确度

准确率 (Accuracy , ACC) 是一个用于评估分类模型的指标^[53]。可根据正类别与负类别按如下方式计算,

$$ACC = \frac{\text{Correct Predictions}}{\text{Total Predictions}} = \frac{TP + TN}{TP + FP + TN + FN}. \quad (3.25)$$

其中 TP 为真正例, FP 为真负例, TN 为假正例, FN 为假负例。

(2) 归一化互信息

归一化互信息 (Normalized Mutual Information , NMI) 来自于信息论, 是用于图像配准的一种经典相似性测度。最初是在 1995 年^[53] 被引入图像处理领域, 用于配准多模图像。其计算公式如下

$$NMI(A, B) = \frac{I(A; B)}{\sqrt{H(A) \cdot H(B)}}. \quad (3.26)$$

由于其无需预处理、自动化程度高以及鲁棒性强等特点, 利用归一化互信息法进行多模图像配准成为了图像处理领域的研究热点。该性能的计算也成为了多模图像配准的一个关键研究问题, 在工业、生物医学、计算机视觉、遥感等应用领域中将发挥巨大的作用。

表 3.2 ACC 最佳性能

	COIL20	Isolet	USPS	Average
Lapscore	55.73%	54.48%	64.30%	58.17%
UFSOL	58.61%	50.45%	64.76%	57.94%
UDFS	57.80%	46.61%	56.63%	53.68%
SOGFS	52.92%	50.41%	-	51.67%
SRCFS	54.65%	59.87%	65.83%	60.11%
RUFS	57.32%	49.99%	65.25%	57.52%
SPEC	59.80%	52.38%	59.23%	57.13%
NOFS_L	58.95%	56.09%	66.18%	60.41%
NOFS_E	56.29%	53.87%	62.36%	57.51%
NOFS_H	58.98%	56.51%	66.20%	60.56%

接下来, 我们将在选用数据集对不同的方法进行测试。在表格中, 如果该方法需要的时间过长, 则标记为 “-”。其中平均值列 (Average) 表示对应方法在不同数据集中平均最佳性能, 从而更好的避免了由于随机性而造成的波动。如表3.2和表3.3所示的不同算法在准确率 (ACC) 和归一化互信息 (NMI) 结果可见, 本章节提出的 NOFS_L、NOFS_E 和 NOFS_H 在测试的三个数据集中都实现了较为卓越的表现。在 COIL20 和 USPS 数据集中, 三种方法所测试的两项性能指标都远超大部分对照数据。

表 3.3 NMI 最佳性能

	COIL20	Isolet	USPS	Average
Lapscore	68.55%	71.45%	60.11%	66.70%
UFSOL	70.95%	68.95%	58.46%	66.12%
UDFS	71.71%	62.62%	50.06%	61.46%
SOGFS	65.24%	66.46%	-	65.85%
SRCFS	68.93%	73.98%	61.29%	68.07%
RUFS	71.34%	65.30%	59.80%	65.48%
SPEC	72.49%	67.73%	54.55%	64.92%
NOFS_L	74.71%	72.94%	61.22%	69.62%
NOFS_E	73.28%	72.01%	58.18%	67.83%
NOFS_H	74.83%	73.33%	61.51%	69.89%

值得注意的是，随着所选特征数量的增加，部分方法在性能上会呈现先上升后下降的趋势，少部分在性能上甚至产生上下波动的情况。因为现实世界中的数据集往往包含大量冗余特征，过多的特征选择也可能导致性能下降。

下述图片的表头为数据集，横坐标表示特征选择数量（Number of Features），纵坐标分别表示准确率（ACC）与归一化互信息（NMI）性能的图片。

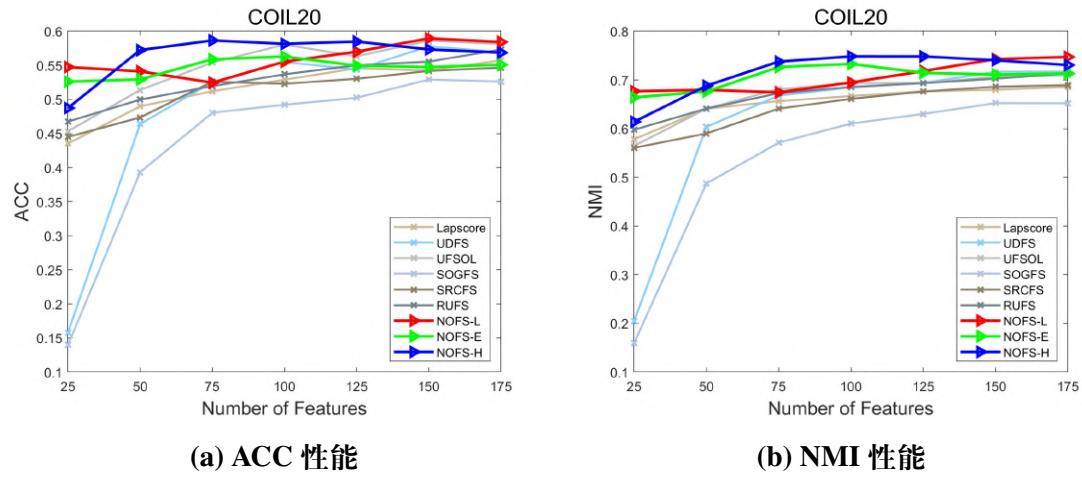


图 3.4 COIL20 数据集性能测试图像

从图3.4可以看出，在 COIL20 数据集中，NOFS 方法的三组数据集都位于图片上部，说明其性能指标较高。其中，NOFS_L 和 NOFS_H 方法分别在特征数为 150 和 75 时，准确率分别高达 58.95% 和 58.98%；而在 USPS 数据集上，NOFS_H 在特征数为 150 时，准确率高达 66.20%，显著优于对照组中表现最好的 SPEC 方法 (59.23%)。而 NOFS_L 和 NOFS_H 更是在 COIL20 数据集上的归一化互信息分别达到 74.71% 和 74.83%，对照组中表现最好的 SPEC 仅为 72.49%。结果表明，NOFS 方法在不同数据集上性能均优于对照组算法，尤其在特征数较少的情况下仍能保持高性能。

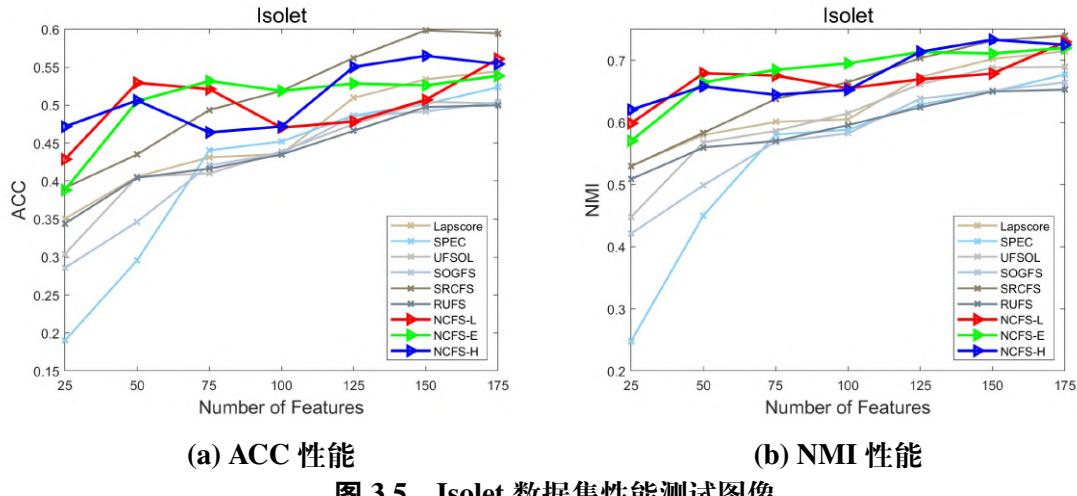


图 3.5 Isolet 数据集性能测试图像

从图3.5可以看出，在 Isolet 数据集中，NOFS_H 在特征数为 150 时准确率高达 56.51%，而 NOFS_E 和 NOFS_L 在特征数为 175 时分别达到了 53.87% 和 56.09%。相比之下，SRCFS 方法在特征数为 150 时表现最佳，准确率高达 59.87%，但此外的方法均未有更优异的性能。NOFS_H 在特征数为 150 时归一化互信息达到了 73.33%，NOFS_E 和 NOFS_L 在特征数为 175 时则分别达到了 72.01% 和 72.94%。尽管 SRCFS 在某些特征选择数量下表现出色，但 NOFS 方法在整体表现上仍然优于大多数对照组方法，尤其是在特征数小于 100 时展现出一骑绝尘的优秀性能。

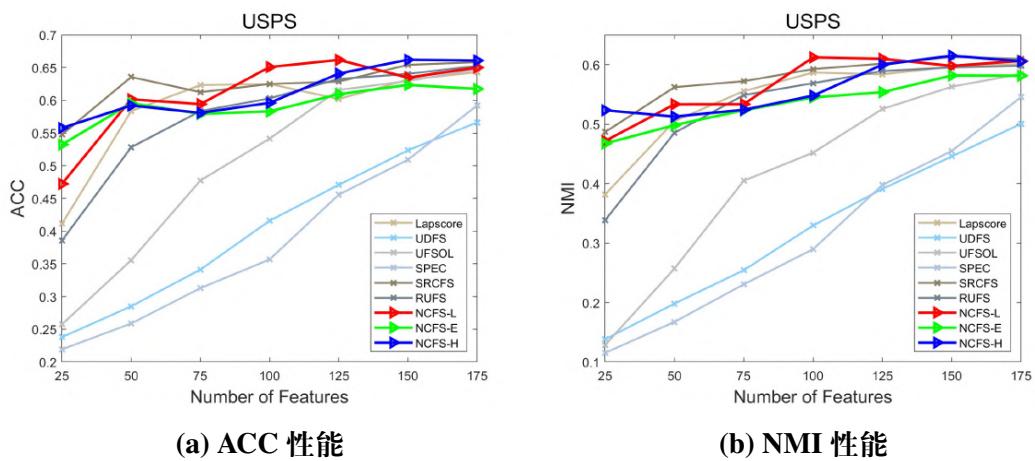


图 3.6 USPS 数据集性能测试图像

从图3.6可以看出，在 USPS 数据集中，NOFS_H 在特征数为 150 时准确率高达 66.20%，而 NOFS_E 和 NOFS_L 分别在特征数为 150 和 125 时达到了 62.36% 和 58.98%，对照组中 SRCFS 和 RUFS 方法在特征数为 150 时的最高数值则分别达到 65.83% 和 65.25%。NOFS_H 在特征数为 150 时归一化互信息达到了 61.51%，NOFS_E 和 NOFS_L 分别在特征数为 150 时分别达到了 58.18% 和 61.22%。相比之下，对照组中的 SRCFS 和 Lapscore 在特征数为 175 时分别为 61.29% 和 60.11%。总的来说，

NOFS 系列方法在整体表现上优于对照组方法，特别是在特征数为 100-150 间展现出极为优越的性能。

3.3.4 噪声干扰测试

不同类型的噪声会对特征选择性能产生不同程度的影响。在设计和选择特征选择算法时，需要考虑噪声类型的影响，以提高算法的性能和稳定性。

高斯噪声是最常见的噪声类型之一，它在数据中呈现为随机扰动，通常服从正态分布，见图3.7。在特征选择过程中，高斯噪声可能导致某些特征的误差增大，从而降低准确度。然而，由于高斯噪声对数据的影响相对均匀，其对归一化互信息的影响通常会略小一些。其中引入如下高斯噪声

$$G(x, \sigma, \mu) = X + \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad (3.27)$$

选取 $\mu = -1$, $\sigma = 0.01$.

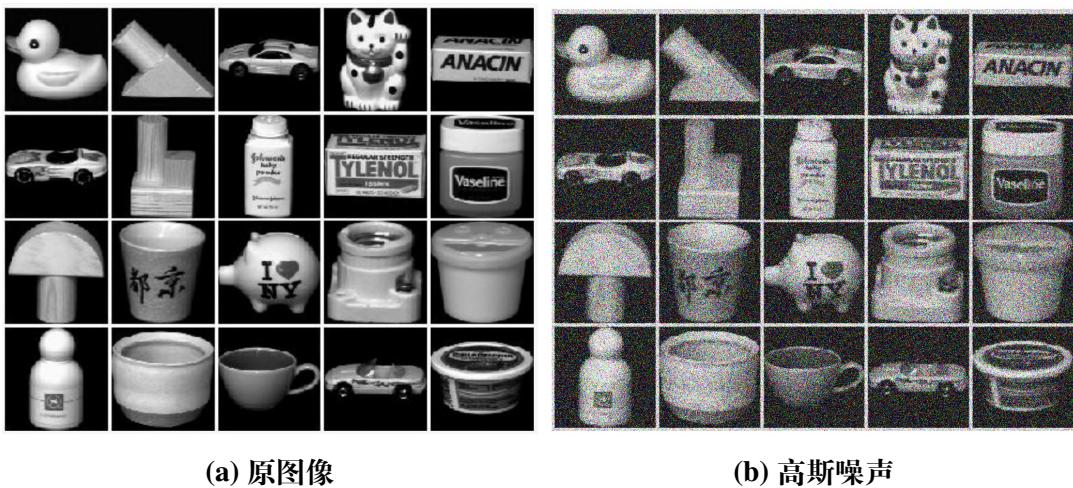


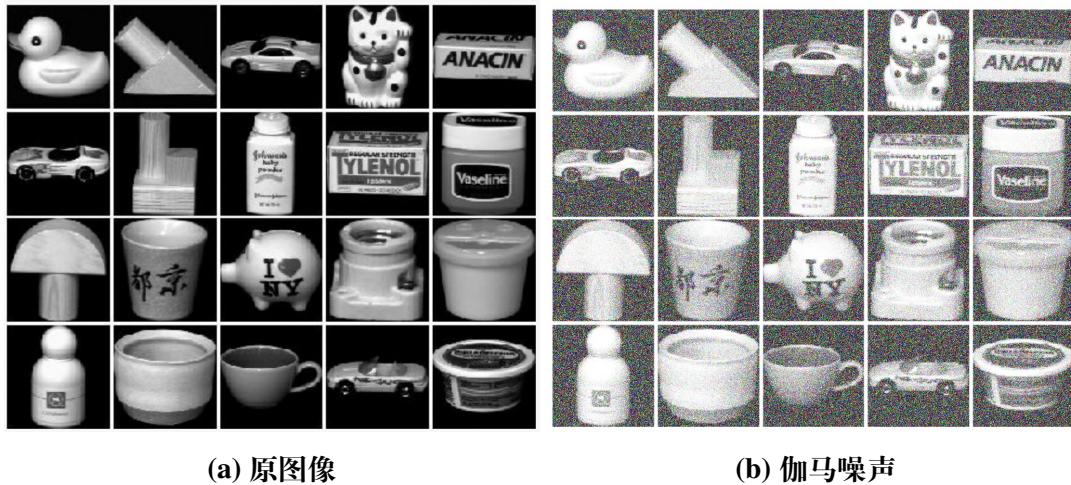
图 3.7 高斯噪声干扰图像

伽马噪声是一种偏斜的噪声类型，通常表现为数据中少量较大值，见图3.8。这种噪声在特征选择中可能对准确度产生的负面影响，可能掩盖或放大某些特征。此外，伽马噪声会导致数据之间的相关性降低，从而对归一化互信息产生不利影响，降低模型对特征的有效识别。

$$Ga(x; \alpha, \beta) = X + \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad (3.28)$$

其中，选取 $\alpha = 0.1$, $\beta = 1$ 。

均匀噪声在数据中呈现为在一定范围内的均匀分布，对所有数据点的影响相似，见图3.9。在特征选择过程中，均匀噪声可能导致特征的区分度降低，从而降低准确



(a) 原图像

(b) 伽马噪声

图 3.8 伽马噪声干扰图形

度。然而，由于均匀噪声对数据的影响一致，它对归一化互信息的影响相对较小，因为数据之间的相关性仍会保持一定的稳定性。

$$U(x, a, b) = \begin{cases} X + \frac{1}{b-a} & \text{if } a \leq x \leq b \\ X & \text{otherwise} \end{cases} \quad (3.29)$$

其中，选取 $a = -0.1$ ， $b = 0.1$.



(a) 原图像

(b) 均匀噪声

图 3.9 均匀噪声干扰图像

在不同类型噪声条件下，对比各个方法的表现（除 NOFS 以外的参考数据集）可以得出以下数据分析结果：

(1) COIL20 数据集的性能表现

在噪声的干扰下，大部分方法所测试出的准确率性能都有所下滑，由于该数据集中噪声对其归一化互信息影响较小，使得其数据均呈现了无噪声时的趋势。而其中 NOFS 方法对高斯噪声均表现出较强的抗噪声干扰能力。相对而言，

SOGFS 和 SPEC 方法在各类噪声条件下的表现较为不稳定，对噪声敏感性较高。受伽马噪声与均匀噪声的影响下，NOFS 方法在特征值 75-125 之间性能指标下滑较为严重，对照方法中 Lapscore 以及 UDFS 则相对更为稳定。

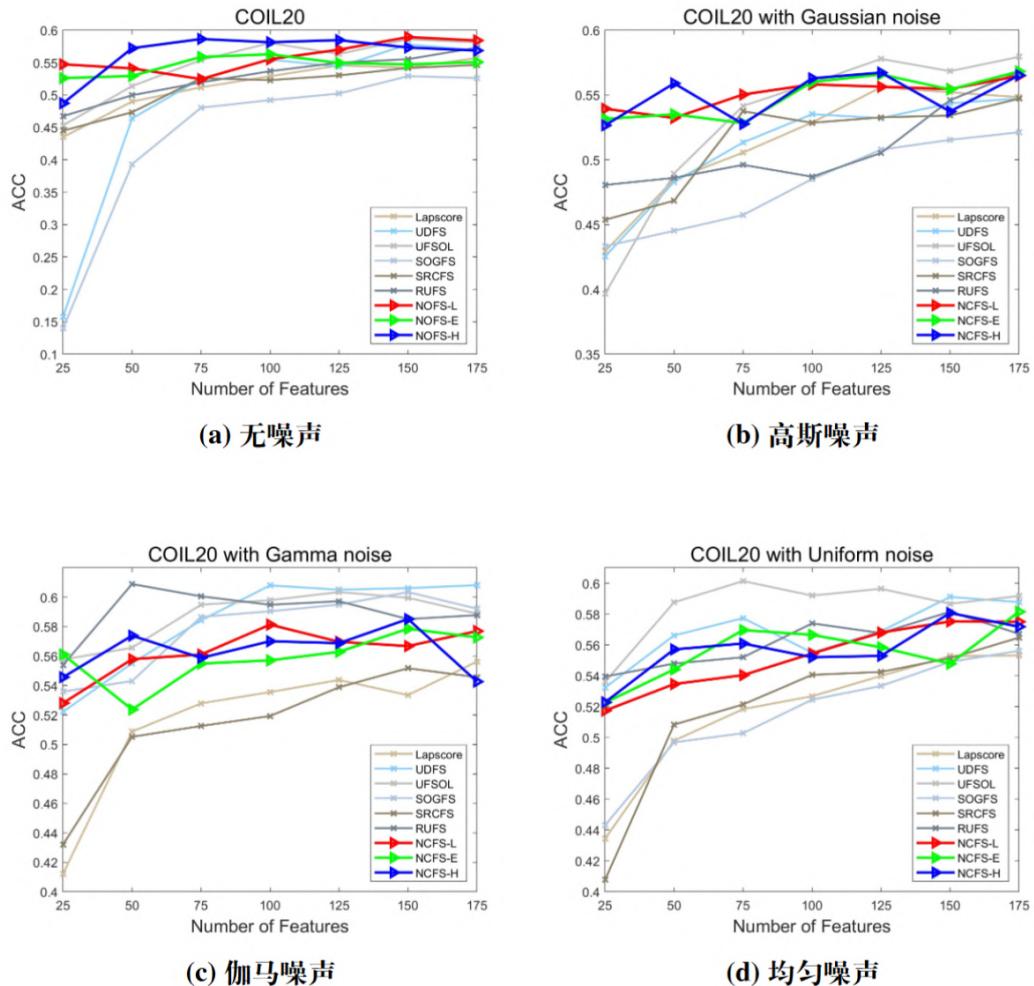
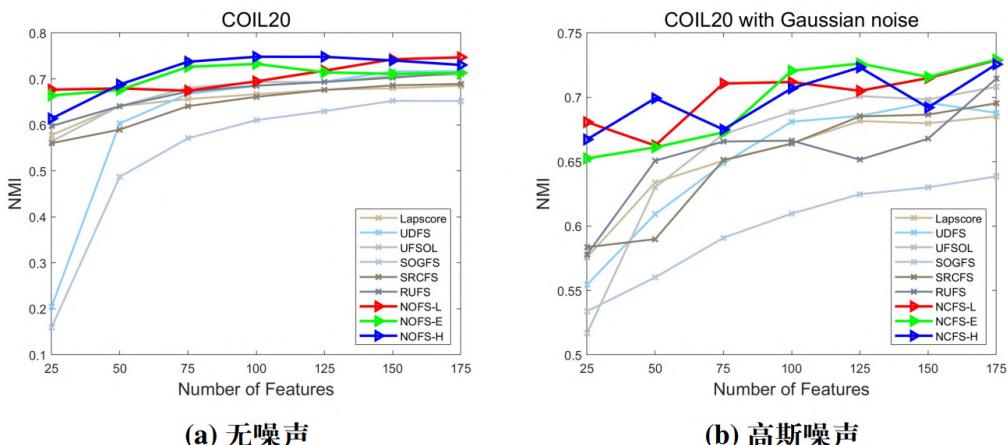


图 3.10 COIL20 数据集噪声前后的 ACC 图像



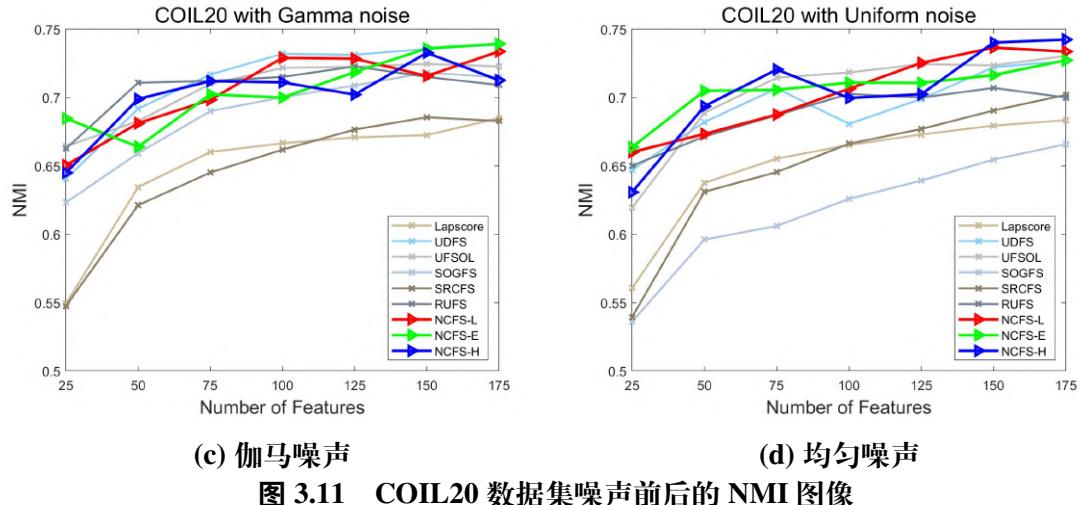
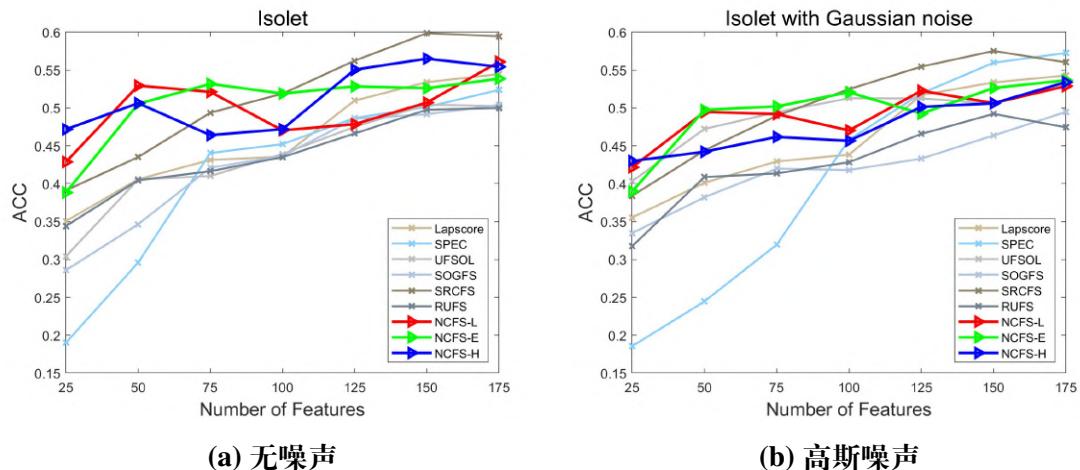


图 3.11 COIL20 数据集噪声前后的 NMI 图像

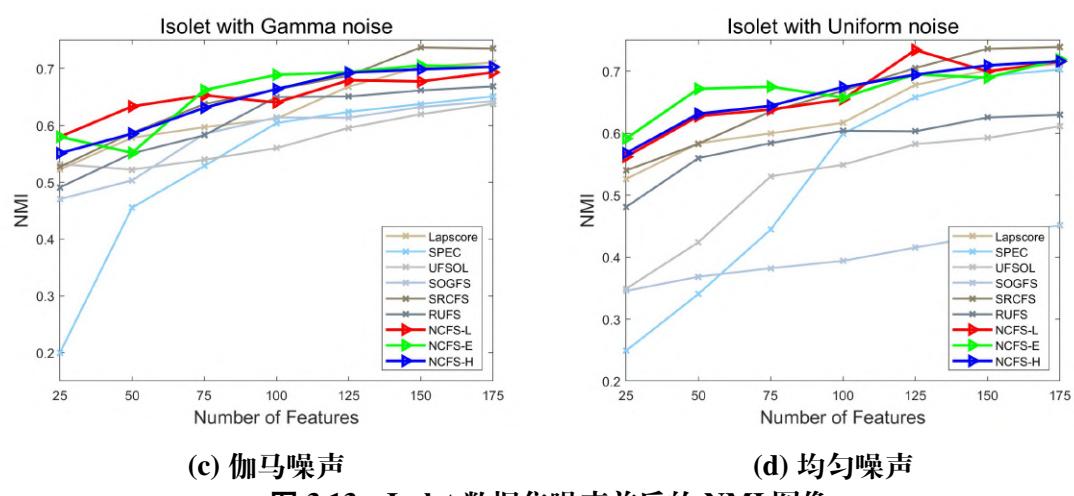
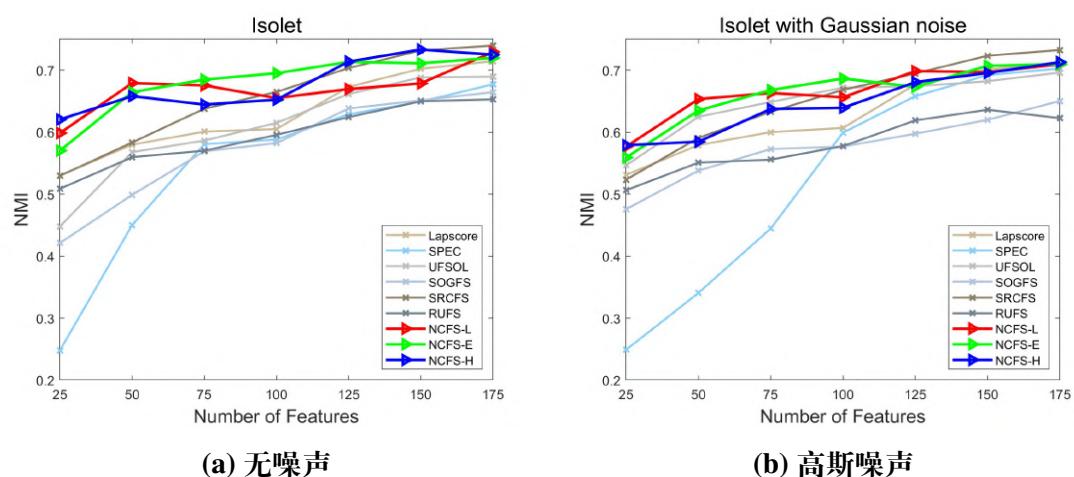
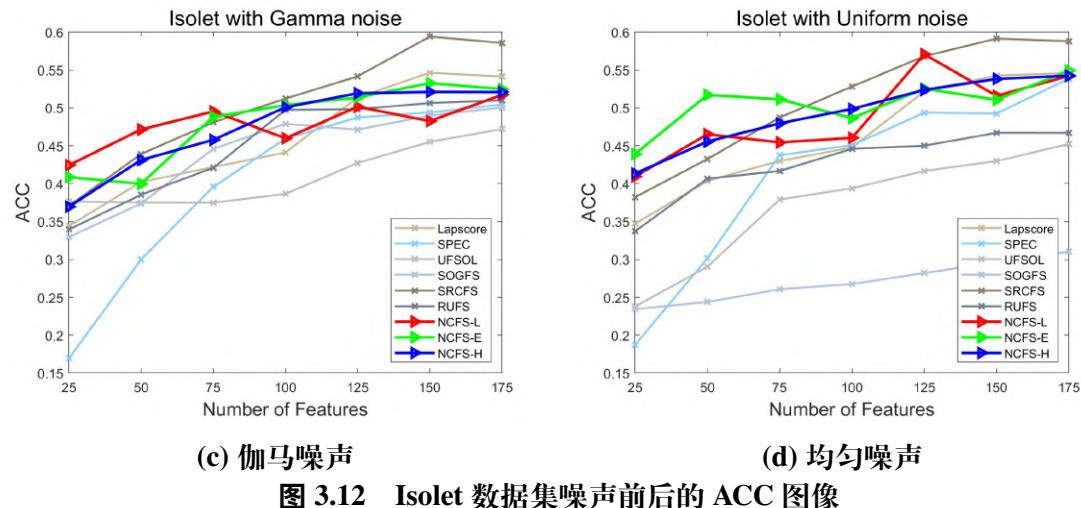
(2) Isolet 数据集的性能表现

在噪声的干扰下，大部分方法所测试出的性能结果都出现明显下滑趋势。由于该数据集为声音数据集，在数据集中各噪声都对归一化互信息的影响都较小，曲线变化与无噪声时相差无几。其中 NOFS 系列方法仍在特征选择数较小时性能更为优越，尤其在 100 以下，但与此同时在特征选择数较小时各方法的性能数值下滑都相对更为严重。



(a) 无噪声

(b) 高斯噪声



(3) USPS 数据集的性能表现

该数据集在噪声的干扰下对归一化互信息的影响相对较大，特征数较低的整体减小趋势接近 5%-10% 上下。由于噪声的随机性，也有部分数据集性能呈现增长趋势，如 UFSOL、UDFS 等。在准确率与归一化互信息上，NOFS 方法添加噪声后与其余对照算法的性能差异并不明显，其准确率数据都集中在 60% 上下，归一化互信息都集中在 55% 上下。同时，特征选择数量的差异变化也变得更为不明显。而 NOFS 整体仍在特征选择数量较少时性能更有优势。

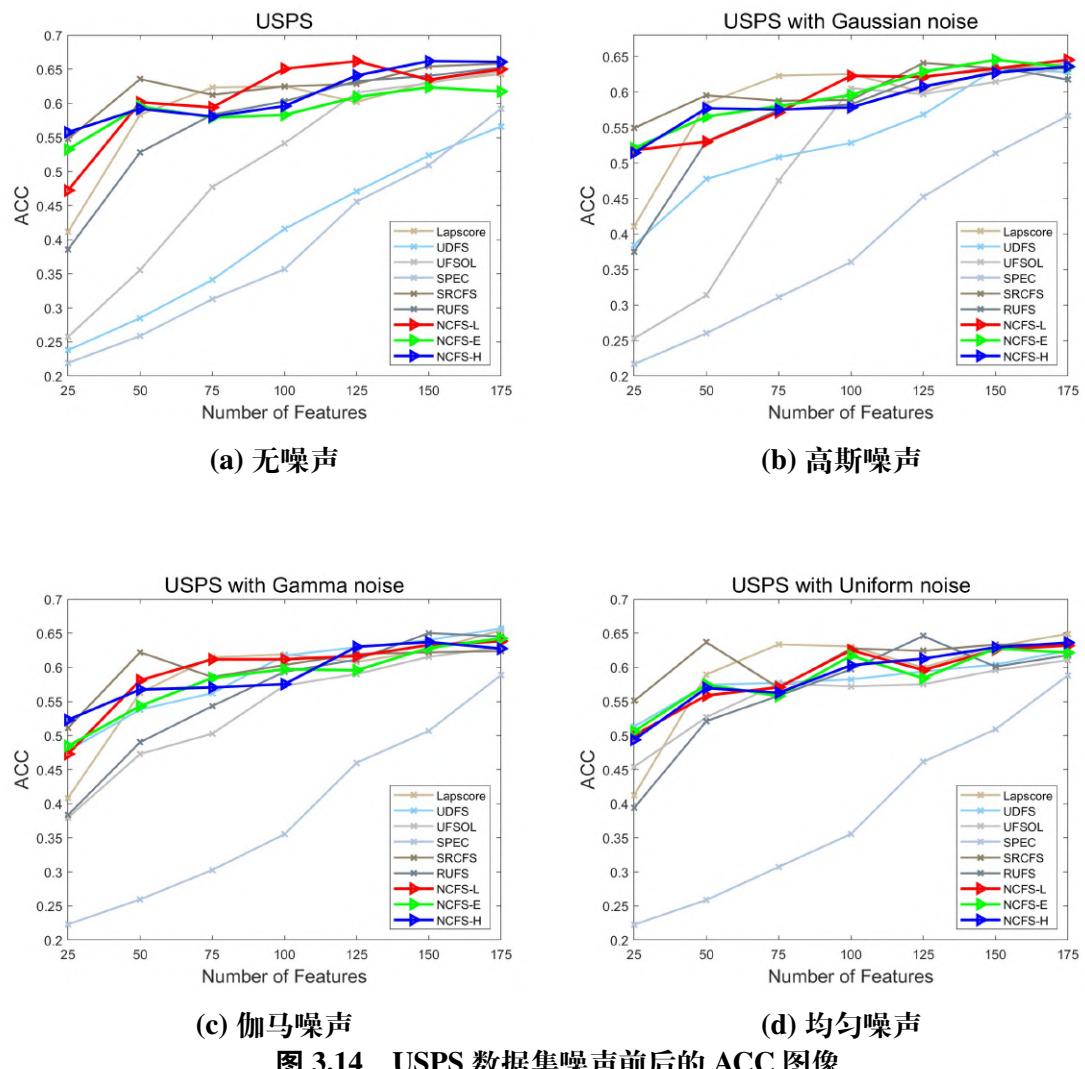


图 3.14 USPS 数据集噪声前后的 ACC 图像

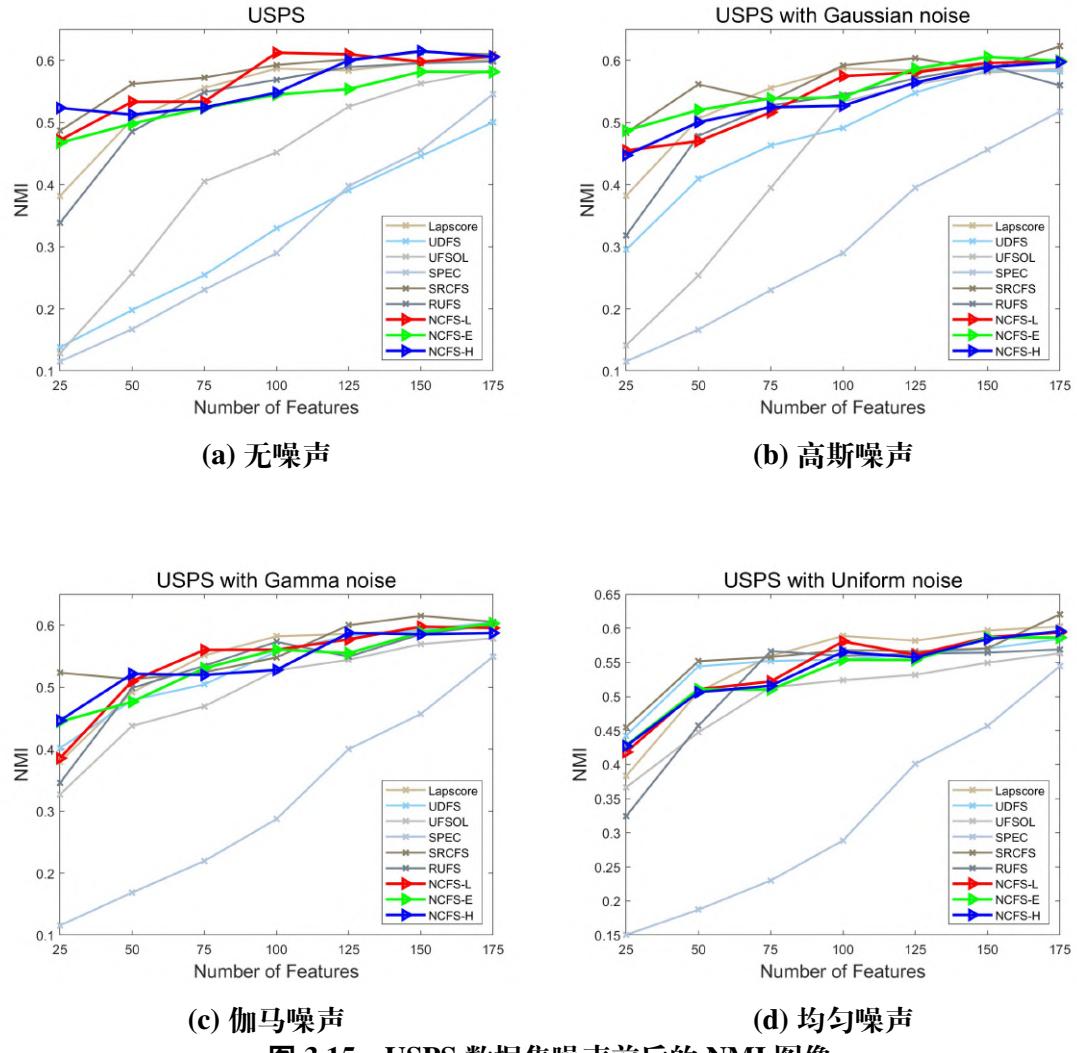


图 3.15 USPS 数据集噪声前后的 NMI 图像

3.4 本章小结

本章提出了一种无监督特征选择新型框架，其结合了非凸正则化和结构化稀疏主成分分析技术。与现有的稀疏主成分分析方法不同，这一框架涵盖了许多非凸函数，增强了无监督特征选择的鲁棒性。同时通过结合黎曼流形优化和增广拉格朗日法，开发了一种迭代算法来解决这个问题。测试结论表明，基于结构稀疏的无监督特征选择新方法在处理复杂数据时，具备较强的抗噪性能和鲁棒性。在与一些其他特征选择方法进行对比时，该方法尤其在特征选择数量较小时在准确度和归一化互信息性能方面表现较为出色。

第四章 基于混合稀疏的主成分分析方法研究

本章针对第三章提出的结构稀疏非凸主成分分析没有考虑元素稀疏的情况，通过向模型中引入元素稀疏，提出了一种基于混合稀疏的无监督特征选择新方法。随后将该方法与上一章节提出的方法以及现有优化算法进行了性能对比，也对该方法在面对噪声数据时的鲁棒性进行了测试。研究结果证明了其的性能优越性，说明其能够为无监督特征选择提供一种在特征选择性能上更有优势的新方案。

4.1 数学模型

回顾第三章中提出的 NOFS 方法。该模型的目标是通过引入稀疏性约束，使得主成分在保持解释性和重要性的同时具有稀疏性，从而提高模型的可解释性。具体的优化问题形式如(4.1)所示

$$\begin{aligned} \min_W \quad & -\text{Tr}(W^T SW) + \lambda \sum_{i=1}^d \Phi(\|w^i\|_2) \\ \text{s.t.} \quad & W^T W = I. \end{aligned} \quad (4.1)$$

该模型通过对 W 的行引入 ℓ_2 范数的稀疏约束，促使生成的主成分具有行结构的稀疏性。这意味着每个主成分仅依赖于部分原始特征，从而提高了模型的可解释性。然而，该模型仅考虑了行结构的稀疏性，而忽略了对元素级别稀疏性的考虑。为了进一步提升模型的稀疏性，尝试在原有模型的基础上引入元素稀疏项 $\Phi(W)$ ，从而提出了如下混合稀疏的主成分分析模型

$$\begin{aligned} \min_W \quad & -\text{Tr}(W^T SW) + \lambda_1 \sum_{i=1}^d \Phi_1(\|w^i\|_2) + \lambda_2 \Phi_2(W) \\ \text{s.t.} \quad & W^T W = I. \end{aligned} \quad (4.2)$$

在本章中，我们将(4.2)记为 NAFS，区别于上一章提出的 NOFS。

4.2 优化算法

首先，引入变量将式(4.2)转化为

$$\begin{aligned} \min_{W,U,V} \quad & -\text{Tr}(W^T SW) + \lambda_1 \sum_{i=1}^d \Phi_1(\|v^i\|_2) + \lambda_2 \Phi_2(u) \\ \text{s.t.} \quad & W^T W = I, \quad W = V, \quad W = U. \end{aligned} \quad (4.3)$$

令

$$\phi(W) = \begin{cases} 0, & \text{如果 } W^T W = I \\ +\infty, & \text{其他,} \end{cases} \quad (4.4)$$

于是式(4.3)便可更改为

$$\begin{aligned} \min_{W,U,V} \quad & f(W, U, V) = -\text{Tr}(W^T SW) + \lambda_1 \sum_{i=1}^d \Phi_1(\|v^i\|_2) + \lambda_2 \Phi_2(U) + \\ & \frac{\beta_1}{2} \|W - U\|_F^2 + \frac{\beta_2}{2} \|W - V\|_F^2 + \phi(W), \end{aligned} \quad (4.5)$$

其中， β_1 和 β_2 是惩罚参数。在 PAM 算法框架下，交替更新每个变量，具体过程如下所示

$$\begin{cases} W^{k+1} = \arg \min_W f(W, U^k, V^k) + \frac{\rho_1}{2} \|W - W^k\|_F^2, \\ U^{k+1} = \arg \min_U f(W^{k+1}, U, V^k) + \frac{\rho_2}{2} \|U - U^k\|_F^2, \\ V^{k+1} = \arg \min_V f(W^{k+1}, U^{k+1}, V) + \frac{\rho_3}{2} \|V - V^k\|_F^2, \end{cases} \quad (4.6)$$

其中， $\rho_1, \rho_2, \rho_3 > 0$ 是正常数， k 是迭代索引，以此为基础推导出 W ， U 和 V 的更新过程。

(1) 更新 W

当 U^k 和 V^k 固定不变时，通过解决下述的优化方案来对 W 进行更新

$$\begin{aligned} \min_W \quad & -\text{Tr}(W^T SW) + \frac{\beta_1}{2} \|W - U^k\|_F^2 + \frac{\beta_2}{2} \|W - V^k\|_F^2 \\ & + \frac{\rho_1}{2} \|W - W^k\|_F^2 + \phi(W), \end{aligned} \quad (4.7)$$

也可以写为另一种形式

$$\begin{aligned} \min_W \quad & -\text{Tr}(W^T SW) + \frac{\beta_1}{2} \|W - U^k\|_F^2 + \frac{\beta_2}{2} \|W - V^k\|_F^2 + \frac{\rho_1}{2} \|W - W^k\|_F^2 \\ \text{s.t.} \quad & W^T W = I. \end{aligned} \quad (4.8)$$

其求解过程类似于算法4，此处不再赘述。

(2) 更新 \mathbf{U}

当 \mathbf{V}^k 和 \mathbf{W}^{k+1} 固定不变时，通过解决下述的优化方案来对 \mathbf{U} 进行更新

$$\min_{\mathbf{U}} \quad \lambda_2 \Phi_2(\mathbf{U}) + \frac{\beta_1}{2} \|\mathbf{W}^{k+1} - \mathbf{U}\|_F^2 + \frac{\rho_2}{2} \|\mathbf{U} - \mathbf{U}^k\|_F^2. \quad (4.9)$$

将算式展开为

$$\min_{\mathbf{U}} \quad \lambda_2 \Phi_2(\mathbf{U}) + \frac{\beta_1 + \rho_2}{2} \|\mathbf{U} - \frac{\rho_2}{\beta_1 + \rho_2} \mathbf{W}^{k+1} - \frac{\rho_2}{\beta_1 + \rho_2} \mathbf{U}^k\|_F^2. \quad (4.10)$$

令 $\mathbf{Y} = \frac{\beta_1}{\beta_1 + \rho_2} \mathbf{W}^{k+1} + \frac{\rho_2}{\beta_1 + \rho_2} \mathbf{U}^k$ ，则可以更改为

$$\min_{\mathbf{U}} \quad \lambda_2 \Phi_2(\mathbf{U}) + \frac{\beta_1 + \rho_2}{2} \|\mathbf{U} - \mathbf{Y}\|_F^2. \quad (4.11)$$

考虑到矩阵每个元素为独立元素，优化问题可以重新表述为

$$\min_{u_{ij}} \quad \lambda_2 \Phi_2(u_{ij}) + \frac{\beta_1 + \rho_2}{2} (u_{ij} - y_{ij})^2. \quad (4.12)$$

其中， u_{ij} 和 y_{ij} 分别是 \mathbf{U} 和 \mathbf{Y} 的第 ij 元素。如果 Φ_2 是硬阈值函数，则 u_{ij} 可应用以下结果

$$u_{ij} = \begin{cases} y_{ij}, & \text{如果 } |y_{ij}| > \sqrt{\frac{2\lambda_2}{\beta_1 + \rho_2}} \\ 0, & \text{其他.} \end{cases} \quad (4.13)$$

(3) 更新 \mathbf{V}

当 \mathbf{W}^{k+1} 固定不变时，通过解决下述的优化方案^[29] 来对 \mathbf{V} 进行更新

$$\min_{\mathbf{V}} \quad \lambda_1 \sum_{i=1}^d \Phi_1(\|v^i\|_2) + \frac{\beta_2}{2} \|\mathbf{W}^{k+1} - \mathbf{V}\|_F^2 + \frac{\rho_3}{2} \|\mathbf{V} - \mathbf{V}^k\|_F^2. \quad (4.14)$$

可以得到以下算式

$$\min_{\mathbf{V}} \quad \lambda_1 \sum_{i=1}^d \Phi_1(\|v^i\|_2) + \frac{\beta_2 + \rho_3}{2} \|\mathbf{V} - \frac{\beta_2}{\beta_2 + \rho_3} \mathbf{W}^{k+1} - \frac{\rho_3}{\beta_2 + \rho_3} \mathbf{V}^k\|_F^2. \quad (4.15)$$

令 $\mathbf{Z} = \frac{\beta_2}{\beta_2 + \rho_3} \mathbf{W}^{k+1} + \frac{\rho_3}{\beta_2 + \rho_3} \mathbf{V}^k$ ，则可以更改为

$$\min_{\mathbf{V}} \quad \lambda_1 \sum_{i=1}^d \Phi_1(\|v^i\|_2) + \frac{\beta_2 + \rho_3}{2} \|\mathbf{V} - \mathbf{Z}\|_F^2. \quad (4.16)$$

而其中，式(4.16)可以分解为一系列向量优化问题

$$\min_{v^i} \lambda_1 \Phi_1(\|v^i\|_2) + \frac{\beta_2 + \rho_3}{2} \|v^i - z^i\|_2^2. \quad (4.17)$$

综上所述，得到完整优化流程如算法7所示。

Algorithm 7 求解 NAFS 的优化算法

输入：矩阵 $X \in \mathbb{R}^{m \times n}$ ，参数 $\lambda_1, \lambda_2, \beta_1, \beta_2, \varepsilon$

输出：投影矩阵 W

迭代：

- 1: 通过(4.8)更新 W^{k+1}
 - 2: 通过(4.13)更新 U^{k+1}
 - 3: 通过(4.17)更新 V^{k+1}
 - 4: 检查收敛性
-

4.3 数值实验

4.3.1 实验数据配置

本章引用的非凸变量以及特征选择参数与第三章数据配置一致，以下会利用 NAFS_L, NAFS_E, NAFS_H 作为新算法的简称。除此之外，以下新引入的数据进行了配置，令 β_1, β_2 为从 10^3 到 10^6 的等比数列，公差为 10^2 。令 λ_1, λ_2 为从 10^{-4} 到 10^{-1} 的等比数列，公差为 10^2 。

4.3.2 性能指标测试

表4.1和表4.2展示了 NOFS 及其对照方法和 NAFS 的准确率和归一化互信息结果。值得注意的是，随着所选特征数量的增加，部分方法在性能上都会呈现先上升后下降的趋势，甚至部分方法的性能产生了上下波动的情况。这是因为现实世界中的数据集往往包含大量冗余特征，过多的特征选择也可能导致性能下降。而 Average 列表示对应方法在不同数据集中的平均最佳性能，能够更好的避免了由于特征选择具有一定随机性而造成的波动。

在对不同特征选择算法的准确率和归一化互信息进行定量分析后，相比对照组算法，NAFS 方法在多个数据集上表现突出。尤其是归一化互信息性能的平均值要超对应的 NOFS 方法以及其他对照方法 1%-6% 上下。而其准确率平均值也是趋于

表 4.1 加入 NAFS 方法的 ACC 最佳性能

	COIL20	Isolet	USPS	Average
Lapscore	55.73%	54.48%	64.30%	58.17%
UFSOL	58.61%	50.45%	64.76%	57.94%
UDFS	57.80%	46.61%	56.63%	53.68%
SOGFS	52.92%	50.41%	-	51.67%
SRCFS	54.65%	59.87%	65.83%	60.11%
RUFS	57.32%	49.99%	65.25%	57.52%
SPEC	59.80%	52.38%	59.23%	57.13%
NOFS_L	58.95%	56.09%	66.18%	60.41%
NOFS_E	56.29%	53.87%	62.36%	57.51%
NOFS_H	58.98%	56.51%	66.20%	60.56%
NAFS_L	59.03%	54.45%	66.29%	59.92%
NAFS_E	58.11%	54.29%	64.41%	58.94%
NAFS_H	59.01%	56.67%	66.50%	60.73%

表 4.2 加入 NAFS 方法的 NMI 最佳性能

	COIL20	Isolet	USPS	Average
Lapscore	68.55%	71.45%	60.11%	66.70%
UFSOL	70.95%	68.95%	58.46%	66.12%
UDFS	71.71%	62.62%	50.06%	61.46%
SOGFS	65.24%	66.46%	-	65.85%
SRCFS	68.93%	73.98%	61.29%	68.07%
RUFS	71.34%	65.30%	59.80%	65.48%
SPEC	72.49%	67.73%	54.55%	64.92%
NOFS_L	74.71%	72.94%	61.22%	69.62%
NOFS_E	73.28%	72.01%	58.18%	67.83%
NOFS_H	74.83%	73.33%	61.51%	69.89%
NAFS_L	74.97%	73.40%	61.68%	70.01%
NAFS_E	74.31%	72.04%	60.04%	68.79%
NAFS_H	75.17%	73.99%	62.63%	70.60%

所用方法上部的位置。然而，对比 NOFS 方法常在较少特征数的情况下保持高性能，NAFS 方法通常与性能大于特征选择数相对较高或更高的情况下性能更优。

下述图片的表头为数据集，横坐标表示特征选择数量 (Number of Features)，纵坐标分别表示准确率 (ACC) 与归一化互信息 (NMI) 性能。从图片4.1可以看出，在 COIL20 数据集中，NAFS 的三组数据集相较于 NOFS 在特征选择数量变化时更趋于平稳，且最高数值都更为优越。同时，NAFS 对比 NOFS 在特征值 25-75 以及特征值大于 150 的时候，其性能指标数值高于 NOFS 方法的程度较大。

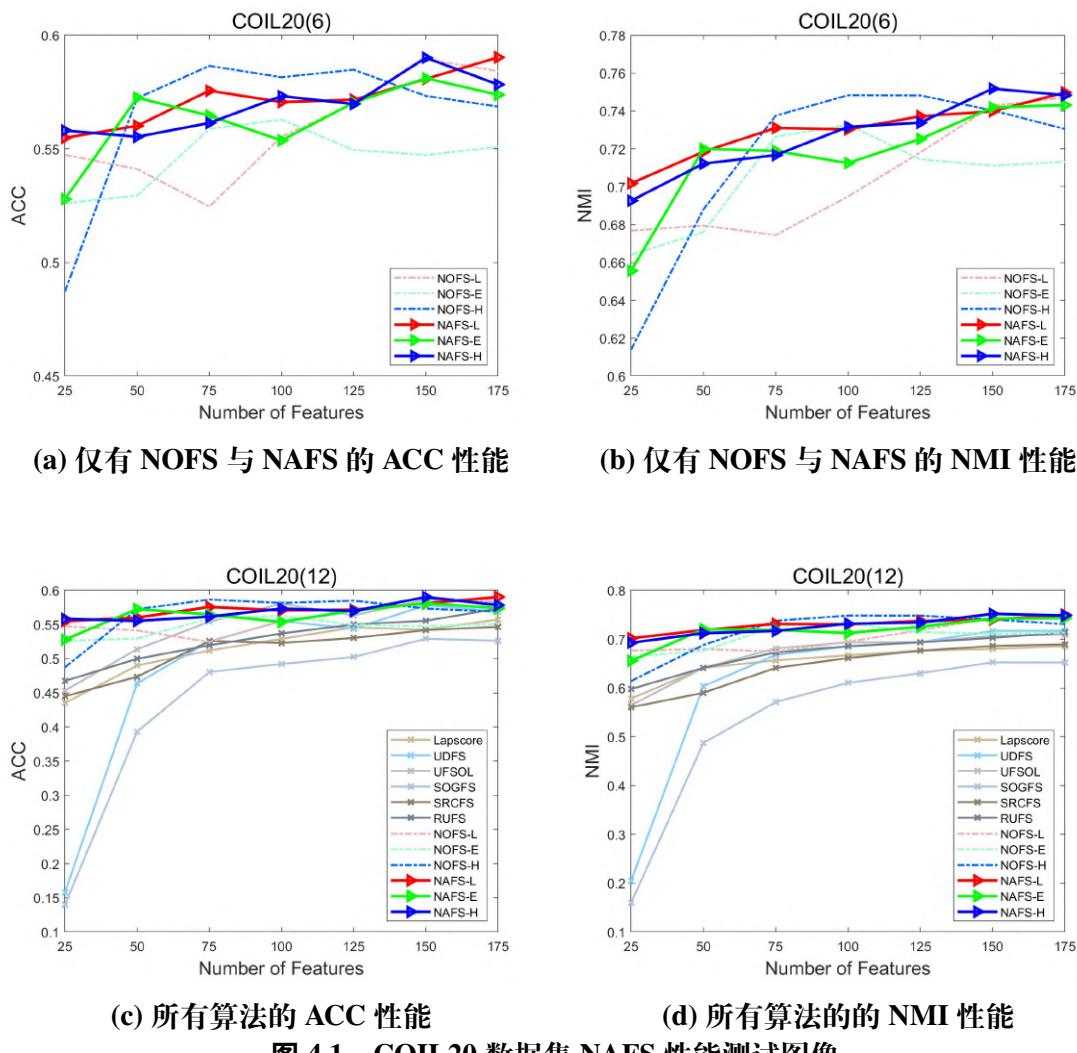


图 4.1 COIL20 数据集 NAFS 性能测试图像

从图片4.2可以看出，在 Isolet 数据集中，NAFS 方法在三组数据集的测试中与 NOFS 相比，其平均性能指标平均值相差不大。与此同时，其仍在特征选择数 25 或 175 时拥有更优的性能。而 NAFS 方法在特征选择数 100 或 125 时出现大幅度性能下滑，这在 NAFS_E 的曲线上表现尤为明显。

从图片4.3可以看出，在 USPS 数据集中，NAFS 系列方法的三组数据集与 NOFS 系列方法的准确率波动变化不明显，并在特征选择数量变化时更趋于平稳。尤其在

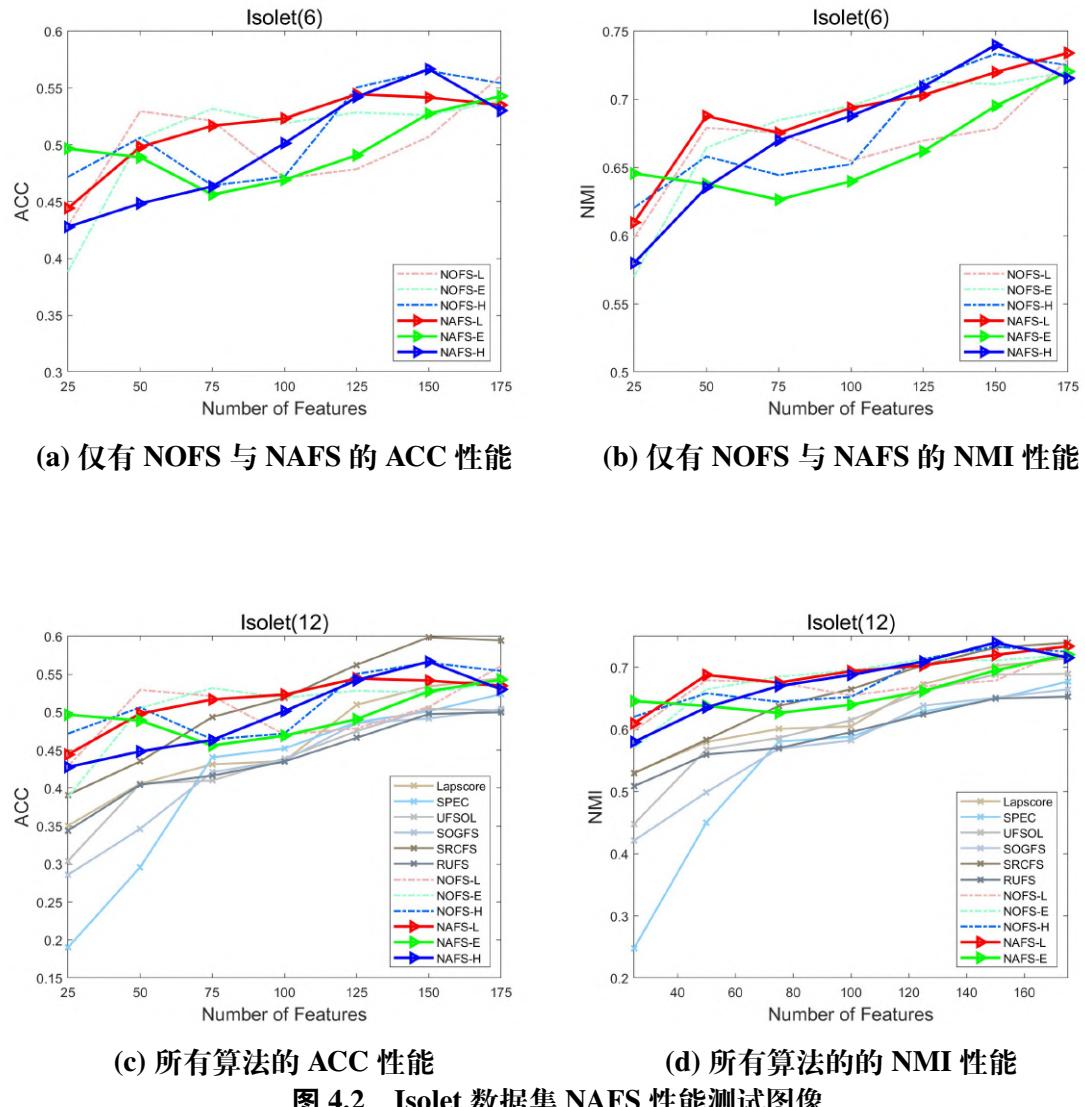
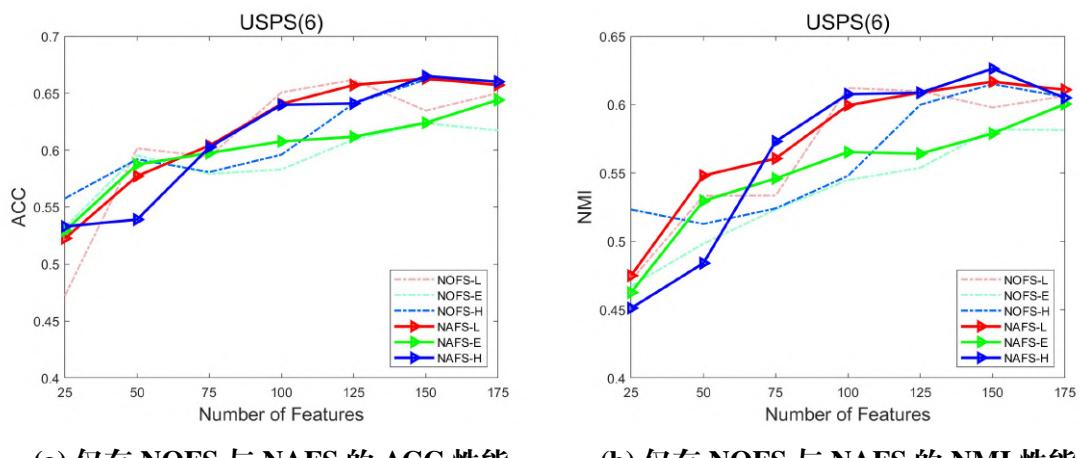


图 4.2 Isolet 数据集 NAFS 性能测试图像



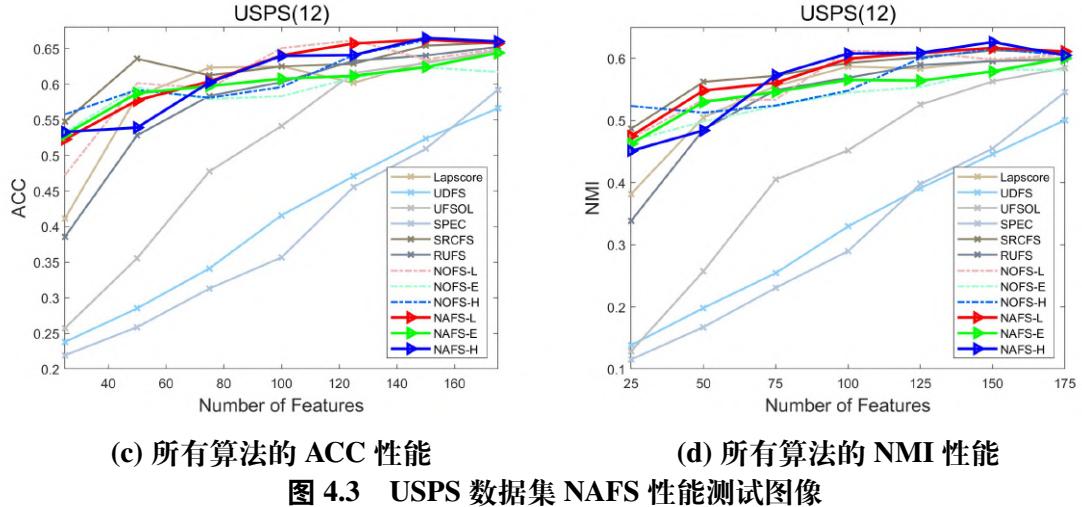


图 4.3 USPS 数据集 NAFS 性能测试图像

特征选择 100-125 之间有更优的性能指标。其中，NAFS_L 相较于 NOFS_L 在大部分特征选择数当中性能更优。但该方法在特征选择数为 175 时出现性能下滑。

综上可以看出，相较于其他方法而言，NAFS 方法受特征选择数量的影响更小。与此同时，其性能指标大部分也都处于平均水平之上，并对比 NOFS 方法更为优异，尤其在特征选择数量 100 上下以及 150-175 上下性能表现相对具有优势。

4.3.3 噪声干扰测试

不同类型的噪声会对特征选择性能产生不同程度的影响。在设计和选择特征选择算法时，需要考虑噪声类型的影响，以提高算法的性能和稳定性。本章引用的噪声函数以及其参数设置与第三章数据配置一致。

在不加入噪声和不同类型噪声对比下，对前文方法（NAFS）的表现如下：

(1) 在 COIL20 数据集上的性能表现

由图可见，在噪声的干扰下，绝大部分方法所测试的性能结果都有所下滑。其中，新方法（NAFS）在抗噪声方面表现出色，其整体性能受噪声影响较小，尤其在特征选择数 25-125 之间，相较于其他方法，NAFS 系列方法的性能指标表现更为优越。

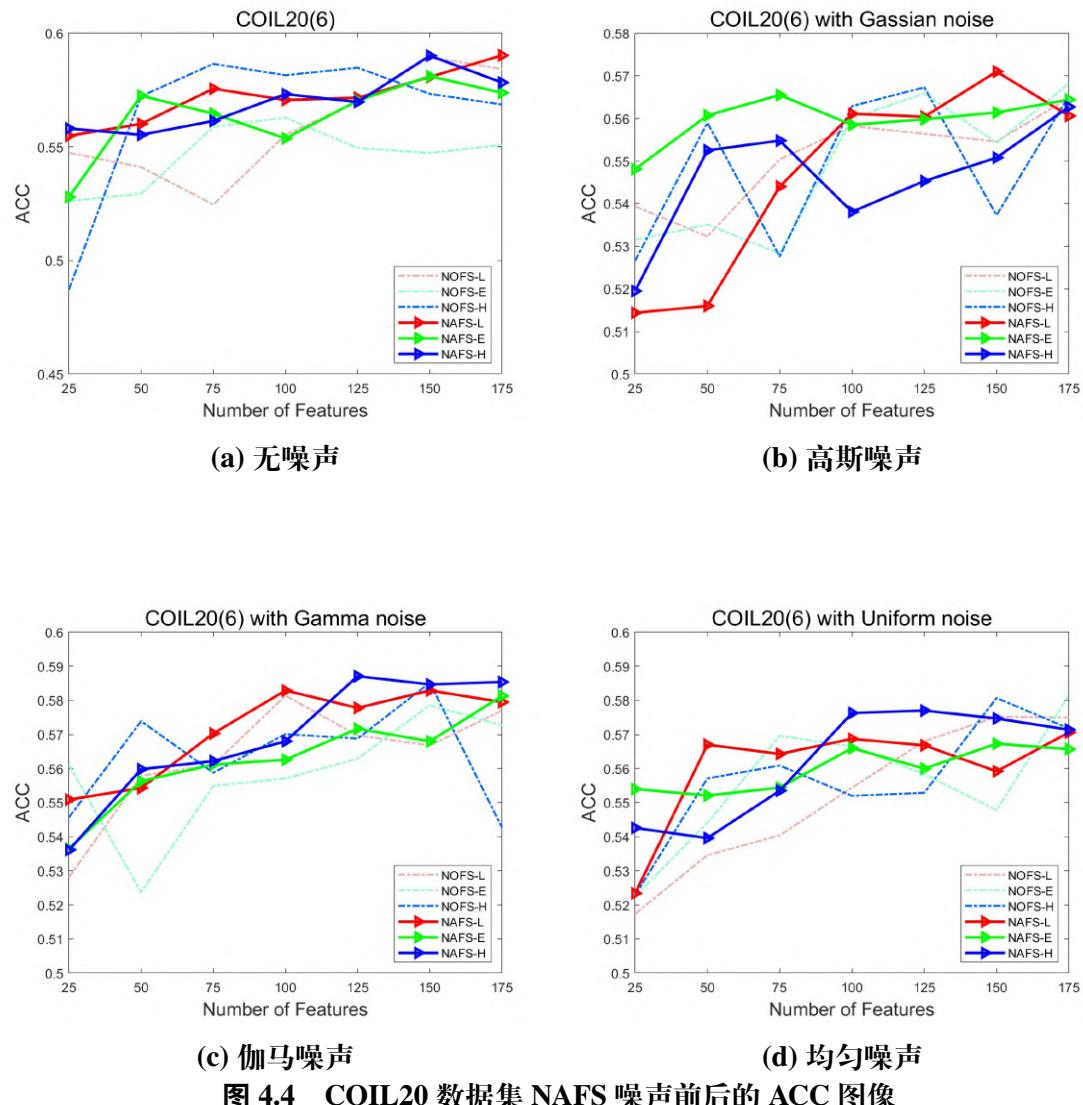
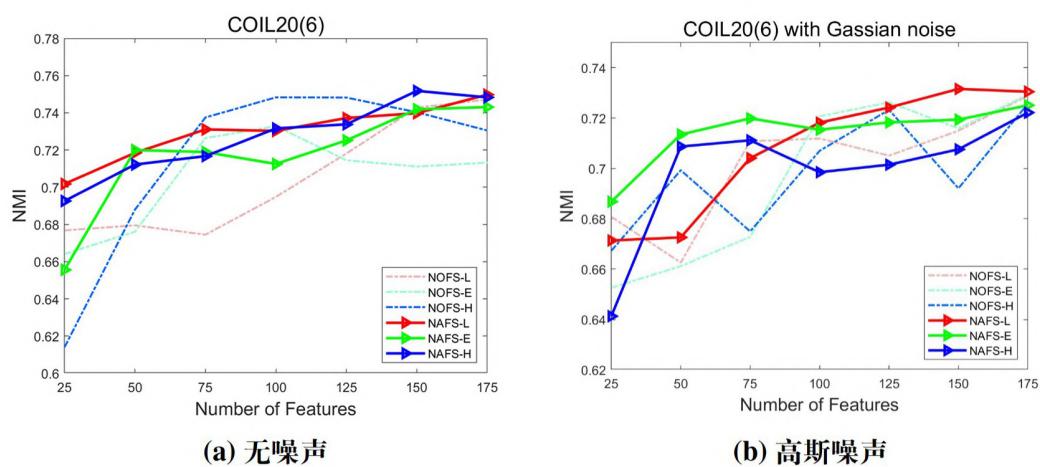


图 4.4 COIL20 数据集 NAFS 噪声前后的 ACC 图像



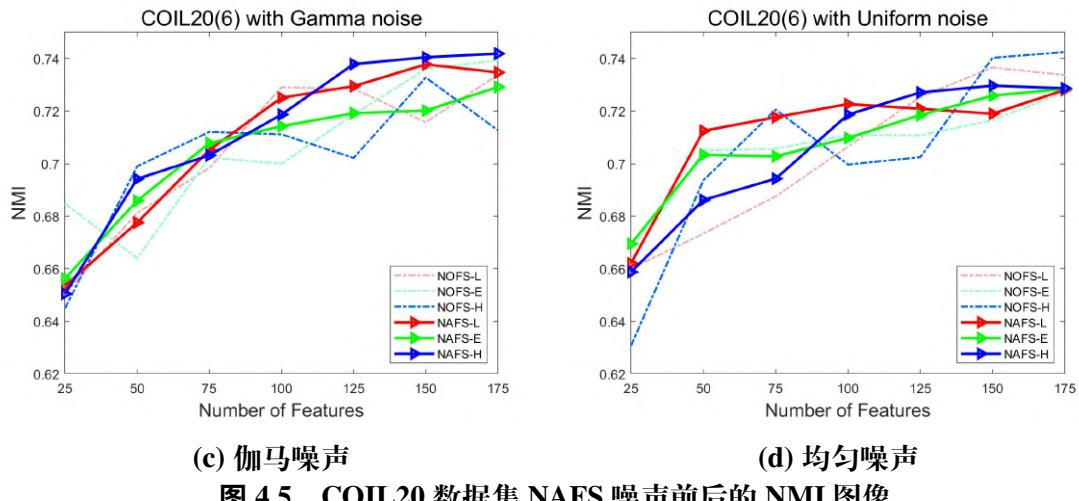
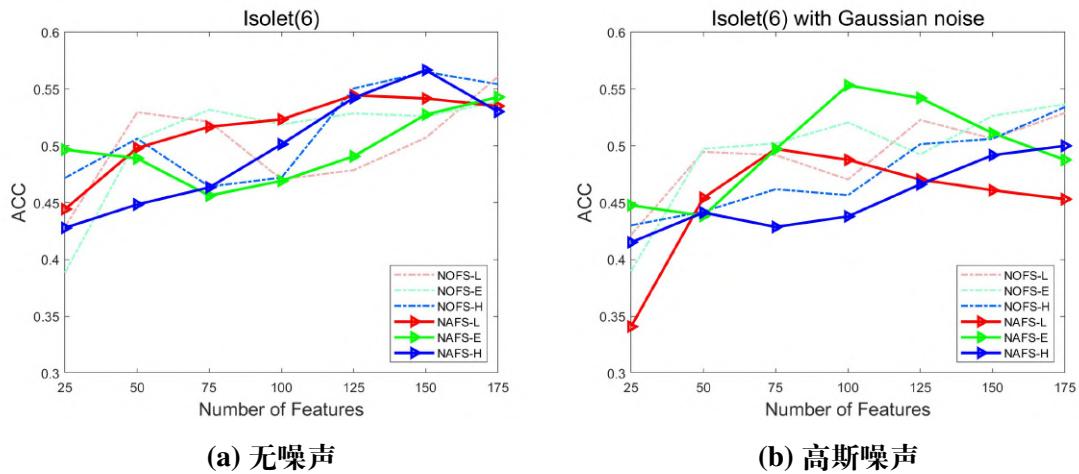


图 4.5 COIL20 数据集 NAFS 噪声前后的 NMI 图像

(2) Isolet 数据集的性能表现

由于噪声的存在具有不确定性，在高斯噪声的干扰下，NAFS_E 的准确率有所提升，而 NAFS_H 则是在均匀噪声的影响下数据也有向上浮动的趋势。这种现象可能是由于该方法在噪声干扰下更加突出地识别了数据的内在结构，使得特征选择更加精准。由此可见，NAFS 方法在 Isolet 数据集上的性能数值变化各异，但其强大的抗噪声干扰能力使其在噪声环境下表现出更为优越的性能。该方法的抗噪性能在特征数量选择为 100 时尤为显著。

综上所述，NAFS 方法在没有噪声干扰的情况下，性能差异与 NOFS 方法不大，但其在噪声环境下的强大鲁棒性使其整体性能更为优越。这表明，NAFS 方法在实际应用中，尤其是在噪声数据处理和特征选择任务中，具有显著的优势和潜力。



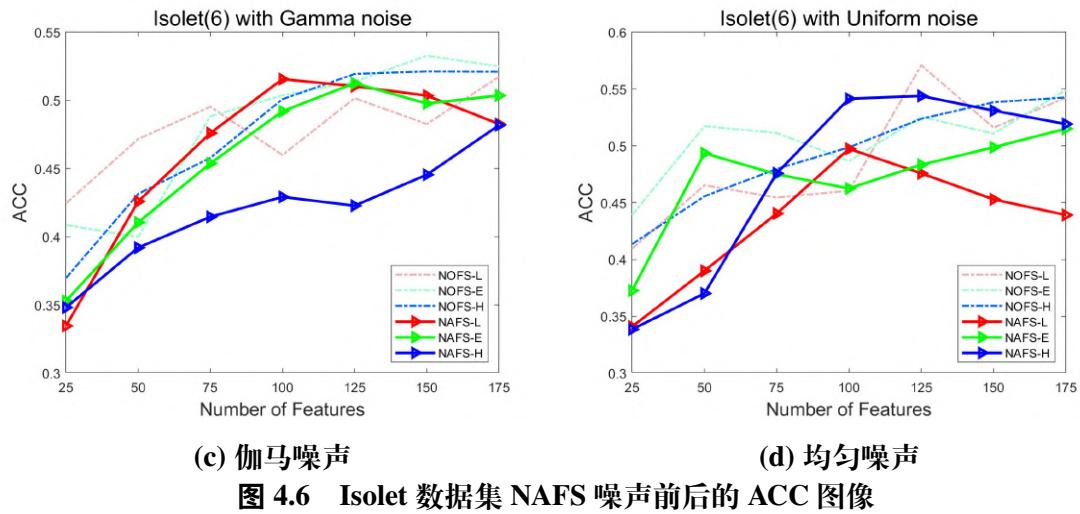


图 4.6 Isolet 数据集 NAFS 噪声前后的 ACC 图像

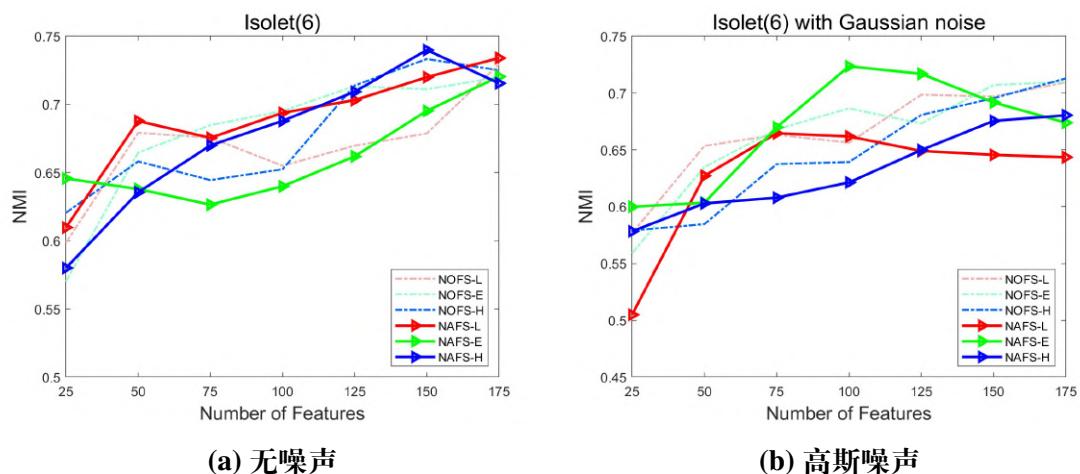


图 4.7 Isolet 数据集 NAFS 噪声前后的 NMI 图像

(3) USPS 数据集的性能表现

由图可见，噪声的干扰下，NAFS 与 NOFS 波动差异不明显。其中，NAFS 方法在噪声影响下特征选择数对其准确率性能差异化影响较大，其最高最低数值之间的差距有所增加。而在均匀噪声影响下，NAFS 方法在低频段获取了极佳的性能优势。同时，两组方法在噪声干扰下的归一化互信息性能指标数据趋于一致，其中 NAFS 在特征选择数为 175 时通常表现出更为优越的性能。

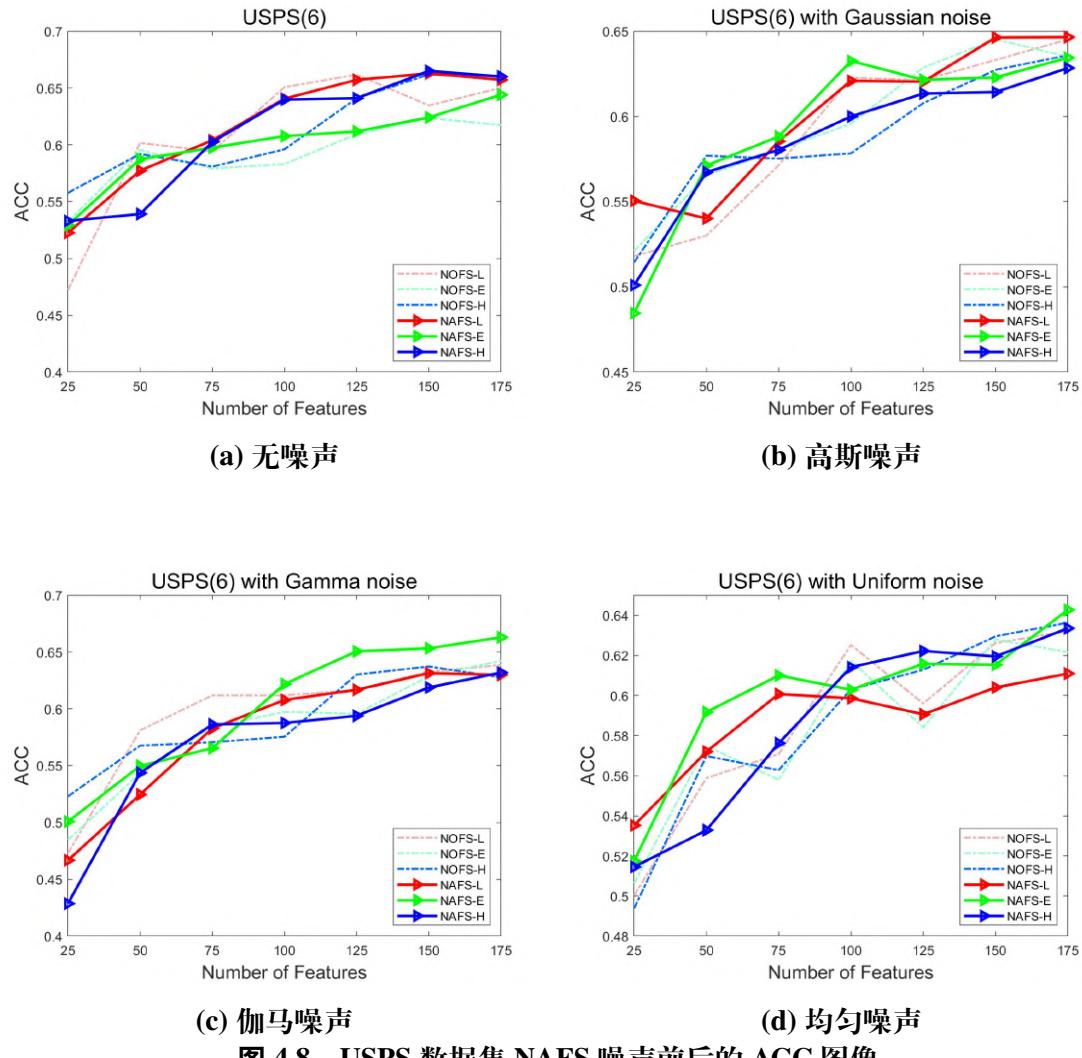


图 4.8 USPS 数据集 NAFS 噪声前后的 ACC 图像

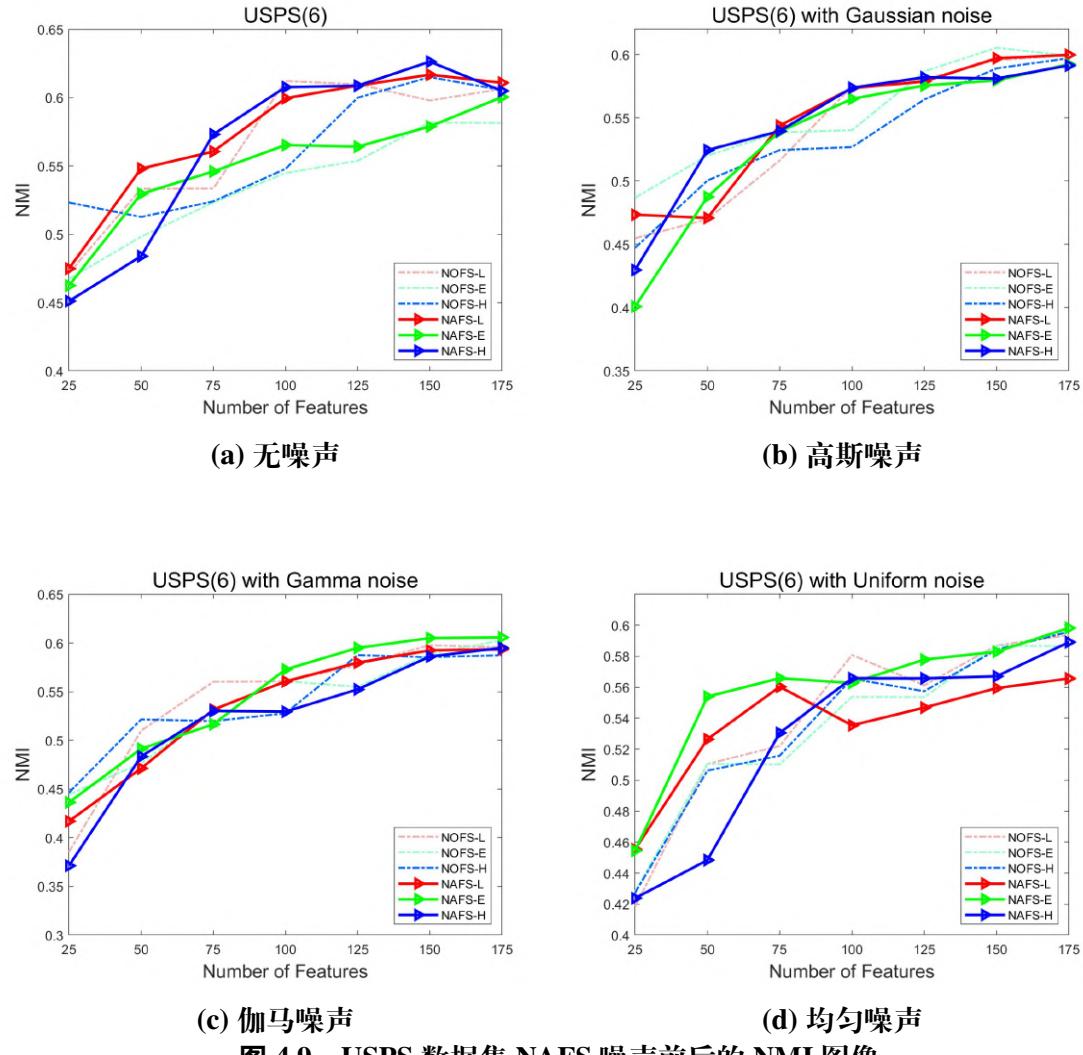


图 4.9 USPS 数据集 NAFS 噪声前后的 NMI 图像

4.4 本章小结

本章提出了一种新型无监督特征选择算法框架，不仅结合了非凸正则化和结构化稀疏主成分分析技术，还引入了元素稀疏项，从而进一步增强了特征选择的性能。通过应用近端交替极小化法，开发了一种迭代算法来解决这个问题。实验结果证明了新方法在抗噪声方面表现出色，其性能指标受噪声影响较小。这也就是说新方法在实际应用中，尤其是在噪声数据处理和特征选择任务中，具有显著的优势和开发潜力。

第五章 总结与展望

5.1 总结

本文主要讨论了无监督特征选择领域中的各类方法与模型的性能优势，提出了基于结构稀疏和混合稀疏的两种无监督特征选择新方法。通过对两种方法在多个常见、开放数据集上进行性能指标测试，验证了其在不同应用下的性能优势，并将其与现有方法进行了对比。

首先，针对无监督特征选择中易出现的过度拟合与缺乏结构性等问题，文章提出了一种在主成分分析中引入正则化非凸模型项的结构稀疏非凸主成分分析方法，该方法通过在目标函数中添加正则化项，实现了在减少误差的同时控制模型复杂性的目标。为了验证该方法的有效性，利用常见数据集 COIL20、Isolet 和 USPS 对该算法进行了全面评估，并与现有的方法进行了详细比较。测试结果表明在引入非凸函数后，该方法在测试数据集上的绝大部分性能优于其他参照方法。值得注意的是，此方法在特征选择数较小的情况下，表现出了极高的准确率和归一化互信息性能，这能够证明其在处理高维数据时可以选用更少的特征选择数，从而提高了模型的运行效率。

此外，为解决上述方法在面对不同数据分布时可能产生错误特征选择的问题，文章接着提出了一种在优化过程中引入元素稀疏的混合稀疏无监督特征选择新方法。该模型的设计目标是通过减少过拟合现象来提高模型的泛化能力，增强模型的效率和鲁棒性。为了优化该模型，应用近端交替极小化算法进行逐步优化，这使得其中的计算过程变得更为简便和高效。测试结果表明该方法相较前述方法，其准确率与归一化互信息更为优越。然而，其在特征选择数量出现变化时，呈现性能数值较为稳定的上下波动表现，说明其受到特征选择数的影响较小。同时，在噪声影响的测试数据下，展示了优越的稳定性和鲁棒性。说明其即使在较为复杂和不确定的噪声环境中，依然能够有效地识别和选择重要特征，保持较高的模型性能。该算法为促进处理噪声影响提供了更为有效的解决方案，未来可以进一步探索其在更大规模数据集和更多实际应用场景中的潜力。

综上所述，本文提出了两种新颖且有效的无监督特征选择方法，在一定程度上成功的解决了特征选择中的部分过度拟合和模型复杂性问题，使得模型的性能得到了显著的提高。与此同时，通过测试验证了所提出方法的有效性和实用性，证实了研究方法具备良好的泛化能力与鲁棒性。这能为未来特征选择领域的进一步研究和应用提供有力支持。

5.2 展望

在未来的研究中，可以进一步拓展和改进提出的无监督特征选择方法，以适应更广泛的应用场景和挑战。通过不断改进和完善无监督特征选择方法，为机器学习领域的发展和应用提供更加有效和可靠的工具和技术。

- (1) **探索更多的非凸优化方法：**进一步探索和研究更多种类的非凸优化方法，进一步改进特征选择模型的性能和鲁棒性，提高其在更广泛数据集和应用场景中的适用性和泛化能力。通过尝试其他不同的非凸函数和优化算法，发现更有效的特征选择策略。
- (2) **考虑多模态数据：**在实际应用中，许多场景下的数据是多模态的，包含多种类型的信息。未来的研究可以探索如何将无监督特征选择方法扩展到多模态数据上，以更好地挖掘不同数据源之间的关联和信息。
- (3) **应用于实际场景：**提出的无监督特征选择方法应用于更广泛的实际场景中，例如医疗诊断、金融预测、图像识别等领域。通过与实际应用场景进行模拟，综合深度学习等前沿技术，验证方法的有效性和实用性，进一步改进和优化算法。

本文的研究成果为无监督特征选择领域的发展提供了新的思路和方法，同时也为未来相关研究和应用提供了有益的参考和启示。与此同时，无监督特征选择模型的进一步优化也是当前在各领域备受瞩目的课题之一。

插图索引

图 1.1 特征选择示例	1
图 1.2 特征选择方法分类	2
图 1.3 无监督特征选择流程	3
图 1.4 封装迭代算法	3
图 1.5 应用场景示意图 ^[25]	4
图 3.1 COIL20 数据集图像	17
图 3.2 USPS 数据集图像	17
图 3.3 非凸函数图像	18
图 3.4 COIL20 数据集性能测试图像	20
图 3.5 Isolet 数据集性能测试图像	21
图 3.6 USPS 数据集性能测试图像	21
图 3.7 高斯噪声干扰图像	22
图 3.8 伽马噪声干扰图形	23
图 3.9 均匀噪声干扰图像	23
图 3.10 COIL20 数据集噪声前后的 ACC 图像	24
图 3.11 COIL20 数据集噪声前后的 NMI 图像	25
图 3.12 Isolet 数据集噪声前后的 ACC 图像	26
图 3.13 Isolet 数据集噪声前后的 NMI 图像	26
图 3.14 USPS 数据集噪声前后的 ACC 图像	27
图 3.15 USPS 数据集噪声前后的 NMI 图像	28
图 4.1 COIL20 数据集 NAFS 性能测试图像	34
图 4.2 Isolet 数据集 NAFS 性能测试图像	35
图 4.3 USPS 数据集 NAFS 性能测试图像	36
图 4.4 COIL20 数据集 NAFS 噪声前后的 ACC 图像	37
图 4.5 COIL20 数据集 NAFS 噪声前后的 NMI 图像	38
图 4.6 Isolet 数据集 NAFS 噪声前后的 ACC 图像	39
图 4.7 Isolet 数据集 NAFS 噪声前后的 NMI 图像	39
图 4.8 USPS 数据集 NAFS 噪声前后的 ACC 图像	40
图 4.9 USPS 数据集 NAFS 噪声前后的 NMI 图像	41

表格索引

表 3.1	数据详情	16
表 3.2	ACC 最佳性能	19
表 3.3	NMI 最佳性能	20
表 4.1	加入 NAFS 方法的 ACC 最佳性能	33
表 4.2	加入 NAFS 方法的 NMI 最佳性能	33

参考文献

- [1] Liu H, Motoda H. Feature selection for knowledge discovery and data mining [M]. Springer, 1998.
- [2] Solorio-Fernández S, Carrasco-Ochoa J A, Martínez-Trinidad J F. A review of unsupervised feature selection methods [J]. Artificial Intelligence Review, 2020, 53 (2): 907–948.
- [3] Kotsiantis S B. Feature selection for machine learning classification problems: A recent overview [J]. Artificial Intelligence Review, 2011, 42: 157–176.
- [4] Sheikhpour R, Sarram M, Gharaghani S, et al. A survey on semi-supervised feature selection methods [J]. Pattern Recognition, 2017, 64 (2016): 141–158.
- [5] Alelyani S, Tang J, Liu H. Feature selection for clustering: A review [M] // Alelyani S, Tang J, Liu H. Chapman and Hall/CRC, 2018: 2018: 29–60.
- [6] Guyon I, Elisseeff A, De AM A. An introduction to variable and feature selection [J]. Journal of Machine Learning Research, 2003, 3: 1157–1182.
- [7] Dy J G, Brodley C E. Feature selection for unsupervised learning [J]. Journal of Machine Learning Research, 2004, 5: 845–889.
- [8] Curts R J, Campbell D E. Pattern recognition algorithms for data mining [M]. Chapman & Hall/CRC Press, 2004.
- [9] Liu H. Feature Engineering for Machine Learning and Data Analytics [M]. CRC Press, 2018.
- [10] He X, Cai D, Niyogi P. Laplacian score for feature selection [J]. Advances in Neural Information Processing Systems, 2005, 18 (5): 504–514.
- [11] Du S, Ma Y, Li S, et al. Robust unsupervised feature selection via matrix factorization [J]. Neurocomputing, 2017, 241: 115–127.
- [12] Law M H C, Figueiredo M A T, Jain A K. Simultaneous feature selection and clustering using mixture models [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26 (9): 1154–1166.
- [13] Guo J, Zhu W. Dependence guided unsupervised feature selection [C]. In Proceedings of the AAAI Conference on Artificial Intelligence, Palo Alto, California, 2018: 2232–2239.
- [14] Li Z, Yang Y, Liu J, et al. Unsupervised feature selection using nonnegative spectral analysis [C]. In Proceedings of the AAAI Conference on Artificial Intelligence, Palo Alto, California, 2012: 1026–1032.
- [15] Yang Y, Shen H, Ma Z, et al. $\ell_{2,1}$ -norm regularized discriminative feature selection for unsupervised learning [C]. In Proceedings of the Twenty-second International Joint Conference on Artificial Intelligence, Freiburg, Germany, 2011: 1589–1594.
- [16] Mairal J, Elad M, Sapiro G. Sparse representation for color image restoration [J]. IEEE Transactions on Image Processing, 2008, 17 (1): 53–69.
- [17] Geng T, Sun G, Xu Y, et al. Image compressed sensing recovery based on multi-scale group sparse representation [C]. In 2018 International Conference on Systems, Signals and Image Processing (IWSSIP), 2018: 1–5.

- [18] Chen M, Renna F, Rodrigues M R D. Compressive sensing with side information: How to optimally capture this extra information for GMM signals? [J]. IEEE Transactions on Signal Processing, 2018, 66 (9): 2314–2329.
- [19] Candes E J, Romberg J. Quantitative robust uncertainty principles and optimally sparse decompositions [J]. Foundations of Computational Mathematics, 2006, 6 (2): 227–254.
- [20] Candes E J, Romberg J K, Tao T. Stable signal recovery from incomplete and inaccurate measurements [J]. Communications on Pure and Applied Mathematics, 2005, 59 (8): 1207–1223.
- [21] Wright J, Yang A Y, Ganesh A, et al. Robust face recognition via sparse representation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31 (2): 210–227.
- [22] Harandi M, Sanderson C, Shen C, et al. Dictionary learning and sparse coding on Grassmann manifolds: An extrinsic solution [C]. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2013: 3120–3127.
- [23] Ritter G. Robust cluster analysis and variable selection [M]. London: CRC Press, 2015.
- [24] 段立娟, 郭亚静, 解晨瑶. 基于特征聚类和等距映射的无监督特征选择算法 [J]. 北京工业大学学报, 2024, 50 (3): 325–332.
- [25] Spruyt V. The curse of dimensionality in classification [J]. Feature Extraction, 2014.
- [26] Pearson K. Lii. on lines and planes of closest fit to systems of points in space [J]. Philosophical Magazine Journal of Science, 1901, 2 (11): 559–572.
- [27] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning; Data mining, inference and prediction [M]. 2nd ed. New York, NY, USA: Springer-Verlag, 2009.
- [28] Hancock P, Burton A, Bruce V. Face processing: Human perception and principal components analysis [J]. Memory Cognition, 1996, 24 (1): 26–40.
- [29] Alter O, Brown P O, Botstein D. Singular value decomposition for genome-wide expression data processing and modeling [J]. Proceedings of the National Academy of Sciences of the United States of America, 2000, 97 (18): 10101–10106.
- [30] Zou H, Hastie T, Tibshirani R. Sparse principal component analysis [J]. Journal of computational and graphical statistics, 2006, 15 (2): 265–286.
- [31] 伍国华, 张雯菲, 毛成辉. 一种基于变量约简的稀疏优化算法 [J]. 控制与决策, 2022, 37 (6): 1550–1558.
- [32] Li Z, Nie F, Bian J, et al. Sparse PCA via $\ell_{2,p}$ -norm regularization for unsupervised feature selection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45 (4): 1392–1404.
- [33] Wang Z, Lu H, Liu H. Tighten after relax: Minimax-optimal sparse PCA in polynomial time [C]. In Proceedings of the International Conference on Machine Learning (ICML), 2014: 2–3.
- [34] Jolliffe I T, Trendafilov N T, Uddin M. A modified principal component technique based on the LASSO [J]. Journal of Computational and Graphical Statistics, 2003, 12 (3): 531–547.
- [35] Tan M, Hu Z, Yan Y, et al. Learning sparse PCA with stabilized ADMM method on stiefel manifold [J]. IEEE Transactions on Knowledge and Data Engineering, 2021, 33 (3): 1078–1088.
- [36] 李本正. 基于协同变换域的稀疏和低秩张量填充的建模与算法研究 [J], 2023, 1: 11–12.
- [37] Attouch H, Bolte J, Svaiter B. Convergence of descent methods for semi-algebraic and tame problems: Proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods [J]. Mathematical Programming, 2013, 137: 91–129.

- [38] Bolte J, Sabach S, Teboulle M. Proximal alternating linearized minimization for nonconvex and nonsmooth problems [J]. Mathematical Programming, 2014, 146: 459–494.
- [39] Zhang X, Wang D, Zhou Z, et al. Robust low-rank tensor recovery with rectification and alignment [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33 (1): 238–255.
- [40] 汪廷华, 胡振威, 占宏祥. 一种新颖的无监督特征选择方法 [J]. 山东大学学报, 2024, 59 (3).
- [41] Li Z, Nie F, Bian J, et al. Sparse PCA via $\ell_{2,p}$ -norm regularization for unsupervised feature selection [J]. Pattern Analysis and Machine Intelligence, 2023, 45 (4): 5322–5328.
- [42] He X, Cai D, Niyogi P. Laplacian score for feature selection [C]. In Advances in Neural Information Processing Systems, 2005, 18.
- [43] Guo J, Quo Y, Kong X, et al. Unsupervised feature selection with ordinal locality [C]. In IEEE International Conference on Multimedia and Expo (ICME), 2017: 1213–1218.
- [44] Yang Y, Shen H T, Ma Z, et al. $\ell_{2,1}$ -norm regularized discriminative feature selection for unsupervised learning [C]. In International Joint Conference on Artificial Intelligence (IJCAI), 2011: 1589–1594.
- [45] Nie F, Zhu W, Li X. Unsupervised feature selection with structured graph optimization [C]. In Proceedings of the Association for the Advancement of Artificial Intelligence Conference on Artificial Intelligence (AAAI), 2016, 30: 1302–1308.
- [46] Huang D, Cai X, Wang C D. Unsupervised feature selection with multi-subspace randomization and collaboration [J]. Knowledge-Based Systems, 2019, 182: 104856.
- [47] Qian M, Zhai C. Robust unsupervised feature selection [C]. In Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence (IJCAI), 2013: 1621–1627.
- [48] Zhao Z, Liu H. Spectral feature selection for supervised and unsupervised learning [C]. In Proceedings of the 24th International Conference on Machine Learning, Corvallis, 2007: 1151–1157.
- [49] Lu C, Tang J, Yan S, et al. Nonconvex nonsmooth low-rank minimization via iteratively reweighted nuclear norm [J]. IEEE Transactions on Image Processing, 2016, 25 (2): 829.
- [50] Friedman J H. Fast sparse regression and classification [J]. International Journal of Forecasting, 2012, 28 (3): 722–738.
- [51] Gao C, Wang N, Yu Q, et al. A feasible nonconvex relaxation approach to feature selection [C]. In Proceedings of the AAAI Conference on Artificial Intelligence, 2011, 25: 356–361.
- [52] Donoho D L, Johnstone I M. Ideal spatial adaptation by wavelet shrinkage [J]. Biometrika, 1994, 81 (3): 425–455.
- [53] Abedinzadeh Torghabeh F, Modaresnia Y, Hosseini S A. Auto-UFSTool: An Automatic Unsupervised Feature Selection Toolbox for MATLAB [J]. Journal of AI and Data Mining, 2023, 11 (4): 517–524.

致 谢

在我即将完成本科学业、步入全新的人生阶段之际，向所有在我的学术道路上给予过支持和帮助的人们表示最诚挚的感谢。

首先，衷心感谢我的父母。在我成长的每一阶段，都给予了我悉心的照料和无微不至的关怀，这些无条件的支持都成为了我的坚强后盾和支柱。无论是面对学业上的挑战还是生活中的挫折，您们始终站在我身边，鼓励我、支持我。您们的爱和付出让我在追求梦想的道路上充满了力量和勇气。不仅为我提供了良好的生活和学习环境，更在精神上给予我无尽的支持。这些教会了我坚韧、勤奋和感恩，使我在未来的道路上勇往直前。

此外，感谢在学术上给予我帮助和指导的学长学姐们，是你们的帮助令我更思进取、渡过难关。在我迷茫时提出宝贵的建议，在我困难时即时伸出了援手。无论是在课程选择还是在课题研究中，总是不遗余力地分享自己的经验和见解，这让我少走了许多弯路。正是因为有你们的帮助，我才能在学术道路上不断进步、不断超越自己。你们是我在大学生活中的榜样，激励我不断追求卓越和进步。

同时，感谢所有在大学生涯中相遇过的朋友们，是你们的陪伴和支持让我的大学时光变得更加丰富多彩。我们一起经历了许多难忘的时刻，每一个快乐的欢声笑语间都充满了轻松与真挚的情感。是与你们的友谊让我在异乡的大学生活中感受到了家的温暖，我们一同分享喜悦，共同面对挑战，在你们的理解和支持下，我度过了许多艰难的时刻，这些美好的回忆将永远珍藏在我的心中。愿我们在未来的日子里都能继续保持联络，友谊长存，共同迎接人生的新篇章。

最后，特别感谢我的老师们，尤其是指导我完成本篇论文的修贤超老师。感谢您对我的悉心指导和耐心解答。您博学和严谨的态度不仅帮助我在学术上取得了进步，更是让我在学术思维方式上得到了极大的提升。在论文写作过程中，您的每一次细致批阅和耐心修改，都让我受益匪浅。您不仅传授了我丰富的知识，更教会了我如何严谨治学、如何独立思考。感谢您对我的悉心栽培和无私奉献，您是我学术道路上的引路人，您的教诲将铭刻在我的心中，指引我走向更加辉煌的未来。

我会倍加珍惜这段宝贵的大学经历，不忘旧情，为实现美好未来而不断奋斗。

武乃琼
上海大学
2024年5月24日

