

基于最小协方差行列式分布对齐的模型转移的研究

叶龙锋

温州大学电气与电子工程学院, 浙江 温州

收稿日期: 2026年3月9日; 录用日期: 2026年4月8日; 发布日期: 2026年4月17日

摘要

针对近红外光谱中无监督模型转移方法在目标域存在异常样本时预测性能不佳、鲁棒性不足的问题, 本文提出一种基于最小协方差行列式分布对齐的无监督模型转移方法(MCD-uDAR)。该方法通过MCD实现对源域与目标域预测值的稳健分布参数估计, 自动识别并抑制异常样本的干扰, 完成无监督条件下的跨域分布对齐与模型转移。仿真实验结果表明, 无异常样本时, 该方法可实现与源域理想模型相当的预测精度; 在10%~40%比例分布漂移异常样本干扰下, 其预测性能显著优于uDOP、di-PLS经典方法, 为复杂场景下的无监督模型转移提供了可靠方案。

关键词

近红外光谱, 最小协方差行列式, 无监督模型转移, 仪器标准化, 响应变量分布对齐

Research on Model Transfer Based on Minimum Covariance Determinant Distribution Alignment

Longfeng Ye

College of Electrical and Electronic Engineering, Wenzhou University, Wenzhou Zhejiang

Received: March 9, 2026; accepted: April 8, 2026; published: April 17, 2026

Abstract

To address the issues of unstable prediction performance and insufficient robustness in unsupervised model transfer methods for near-infrared spectroscopy when anomalous samples exist in the

target domain, this paper proposes an unsupervised model transfer method based on minimum covariance determinant alignment (MCD-uDAR). This method employs MCD to achieve robust distribution parameter estimation for predictions in both source and target domains, automatically identifying and suppressing interference from anomalous samples to accomplish cross-domain distribution alignment and model transfer under unsupervised conditions. Simulation results demonstrate that in the absence of anomalous samples, this method achieves prediction accuracy comparable to an ideal source domain model. Under interference from distribution-drifted anomalous samples at rates of 10% to 40%, its predictive performance significantly outperforms classical methods such as uDOP and di-PLS, providing a reliable solution for unsupervised model transfer in complex scenarios.

Keywords

Near-Infrared Ectroscopy, Minimum Covariance Determinant, Unsupervised Model Transfer, Instrumental Standardization, Response Variable Distribution Alignment

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近红外光谱(Near-infrared Spectrum, NIRS)分析技术凭借快速、无损的优势,已在工业在线监测、农产品品质检测、药物分析等领域实现广泛应用[1][2]。但在实际应用中,已建立的校准模型常因检测场景、仪器状态变化出现预测偏差,模型通用性不足。传统体验过更新数据集、重新建模的优化方式,存在耗时费力、检测成本高的局限。而模型转移方法可实现校准模型在不同仪器间的复用,有效避免重复建模所带来的时间与资源消耗[3]。目前, Dong [4]等人针对鸡蛋新鲜度 VIS-NIR 光谱模型跨品种预测能力下降问题,对比三种模型转移方法,在以矮脚鸡蛋为主品种时可实现蛋清 PH 值的高精度预测;冯佳豪[5]等人建立了淫羊藿与柔毛淫羊藿中 4 种核心黄酮成分的近红外光谱定量模型,通过 DS 与 PDS 算法实现了台式到便携式近红外仪器间的有效模型转移,为中药近红外分析的跨仪器推广提供了理论与技术支撑;Mishra [6]等人将域不变偏最小二乘(di-PLS)用于近红外光谱模型转移,无需标准样品即可将稻谷籽粒的蛋白质定量模型适配到米粉形态,实现了同一样品不同物理形态间的无标样模型自适应。

在无监督模型转移问题中,目标域样本通常仅包含光谱信息而缺乏对应的参考值,这使得异常样本难以通过传统残差分析或监督判别方式进行识别。在此情况下,若异常光谱在预测空间中占据一定比例,则会影响模型转移的预测效果。基于上述问题,本文通过引入鲁棒统计思想,提出一种基于最小协方差行列式的无监督模型转移方法(Minimum Covariance Determinant-assisted Unsupervised Distribution Alignment based on Response, MCD-uDAR),该方法通过对目标域预测响应变量分布进行稳健估计,有效降低异常样本在分布参数估计与对齐过程中的影响,从而提升了模型在存在异常样本情形下的可靠性。

2. 材料与方法

2.1. 仿真实验数据

本文构建了一组具有明确物理意义的仿真光谱数据[7][8]。首先,在给定的波长范围内共计 $p = 300$ 个

波长变量定义三个化学组分的纯物质光谱 $s_1, s_2, s_3 \in \mathbb{R}^{1 \times p}$ 。图 1 显示各组分光谱由高斯函数模拟其吸收峰：

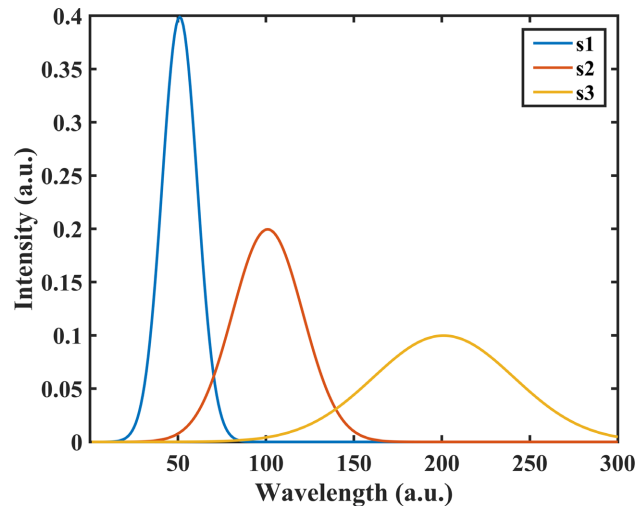


Figure 1. Spectra of three pure components

图 1. 三种纯组分光谱

$$s_k(\lambda) = \mathcal{N}(\lambda | \mu_k, \sigma_k^2) \quad k = 1, 2, 3 \quad (1)$$

参数分别设为 $(\mu_1, \sigma_1) = (0, 1)$ 、 $(\mu_2, \sigma_2) = (5, 2)$ 和 $(\mu_3, \sigma_3) = (15, 4)$ ，以表征窄、中、宽三种典型峰形。所有光谱构成矩阵 $S \in \mathbb{R}^{3 \times p}$ 。

随后，生成 $m = 80$ 个样本的浓度矩阵 $C \in \mathbb{R}^{80 \times 3}$ ，其服从分布 $\mathcal{N}(5, 1)$ ，基于线性混合模型生成理想光谱数据 $X = C \cdot S$ 。

为模拟不同仪器条件引入的系统性响应差异，对光谱数据通过响应函数进行滤波处理。源域与目标域均采用长度为 $L = 9$ 的有限冲击响应(FIR)滤波器，其中源域响应函数设为均值滤波器：

$$h_s = \frac{1}{9}[1, 1, \dots, 1]^T \quad (2)$$

目标域响应函数则由随机权重构成：

$$h_t = \frac{1}{9}[\gamma_1, \gamma_2, \dots, \gamma_9]^T \quad \gamma_i \sim \mathcal{N}(0, 1) \quad (3)$$

对每条光谱信号进行滤波并采用边界复制策略处理起始波长，得到源域和目标域光谱。

为进一步增强仿真数据的真实性，在滤波后的光谱中引入加性背景干扰。首先，引入加性背景干扰以模拟基线漂移、低频环境噪声。该背景信号由 $N_b = 20$ 个高斯型谱峰随机叠加构成，其形式为：

$$b(\lambda) = \frac{1}{N_b} \sum_{i=1}^{N_b} \mathcal{N}(\lambda | \mu_i, \sigma_i^2) \quad (4)$$

其中高斯峰的中心位置 $\mu_i \in [0, 25]$ 和标准差 $\sigma_i \in [1, 10]$ 随机生成。源域和目标域分别独立生成背景项 β_s 和 β_t ，并按照信号幅度的 1% 比例叠加至光谱中：

$$X_d = X_d + 0.01 \cdot \beta_d \quad d \in \{s, t\} \quad (5)$$

在此基础上，为模拟实际测量过程中的随机误差，进一步向源域和目标域光谱中加入零均值高斯白噪声 $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ ，其中噪声标准差 σ^2 设定为信号标准差的 0.1%，最终得到的光谱表示为：

$$X_d = X_d + 0.01 \cdot \varepsilon_d \quad d \in \{s, t\} \quad (6)$$

2.2. 异常值

在真实数据集中，常常会出现与大多数样本分布显著不同的数据点，这类样本在机器学习与化学计量学领域中通常被称为异常值[9]。异常样本其产生原因复杂多样，既可能来源于测量环境波动、仪器状态变化，也可能由基础数据出错或样本本身引起[10]。在无监督模型转移问题中，目标域样本仅包含光谱信息而不具备对应的标签信息，异常样本的构造与识别必须严格限定在光谱特征及其统计分布层面，以避免引入任何不符合无监督设定的先验信息。基于这一原则，本文在目标域光谱数据中引入分布漂移异常样本，用以系统评估模型转移方法在复杂目标域场景下的鲁棒性。

分布漂移异常值模拟量因仪器长期漂移或样品来源系统性变化导致的数据整体分布参数改变。对于 $m \times n$ 的目标域原始光谱矩阵 X_t 。首先，基于全部目标域光谱样本估计其均值向量和协方差矩阵：

$$\mu_x = \frac{1}{n} \sum_{i=1}^n x_i \quad \sigma_x = Cov(X_t) \quad (7)$$

在此基础上，构造一个新的光谱分布，其均值和协方差相对于原始分布发生系统性偏移：

$$\mu_x^* = \mu_x + \alpha \cdot \sigma_x \quad \sigma_x^* = \beta \cdot \sigma_x \quad (8)$$

其中， α 为均值偏移因子， β 为协方差缩放因子。

随后，从该新的正态分布中随机生成比例为 ρ 的光谱样本，并用其替换目标域中对应数量的原始样本，从而形成分布漂移的异常样本。

2.3. 大豆数据集

大豆数据集：该数据集包含 60 个大豆样本，来自 2 台不同的 NIR 光谱仪器上测量，分别记为 R1 和 R2，每台仪器测得的光谱数据波长范围为 1100~2500 nm，波长间隔为 4 nm，共 300 个预测变量，每个大豆样本包含 3 个属性信息，即：水分、油脂、蛋白质，本文仅对油脂进行研究。

2.4. 最小协方差行列式分布对齐

本文提出基于最小协方差行列式(Minimum Covariance Determinant, MCD)分布对齐的无监督模型转移方法，旨在解决目标域无标签下的异常识别问题。MCD [11]是一种具有高鲁棒性的多元位置与协方差估计方法。其基本思想是通过迭代与马氏距离，寻找数据中协方差行列式中最小的一个子集，从而实现位置参数和协方差矩阵的估计。异常样本的马氏距离通常大于正常样本，可作为异常识别的核心判别依据，低维场景下 MCD 可通过马氏距离与截止值 $\sqrt{\chi_{p,0.975}^2}$ 对比完成异常判别。

基于源域中的光谱矩阵 X_s 和响应变量 y_s ，通过偏最小二乘回归建立预测模型，其形式为：

$$y_s = X_s \beta \quad (9)$$

其中 β 为回归系数。在无监督的情境下，目标域的真实响应变量 y_t 不可获得，但可通过源域训练所得的回归系数 β 对目标域样本 X_t 进行预测：

$$\hat{y}_t = X_t \beta \quad (10)$$

在获得源域响应值 y_s 和目标域预测值 \hat{y}_t ，MCD-uDAR 分别对两者进行最小协方差行列式估计，以获得各自对应的稳健均值和标准差：

$$\left[\hat{\mu}_s^{robust}, \hat{\delta}_s^{robust} \right] = MCD(y_s) \quad (11)$$

这一步骤在方法中实现异常样本识别和分布参数稳健估计两大功能。对于源域响应值 y_s 通常源于实验，但仍可能包含因操作或样本异常引入的离群点，MCD 估计能够在不依赖强分布假设的前提下提供稳健估计。

$$\left[\hat{\mu}_t^{robust}, \hat{\delta}_t^{robust} \right] = \text{MCD}(\hat{y}_t) \quad (12)$$

对于目标域预测值 \hat{y}_t ，若对应的光谱数据存在异常样本，则预测结果将发生偏离；此时若采用传统估计，异常值会严重扭曲分布参数。MCD 方法通过识别并排除离群点，可以提取主题分布的稳健位置和分布估计，从而为后续分布匹配校正提供可靠的参数基础。

基于正态分布的线性变换性质，MCD-uDAR 通过以下公式将目标域预测值的分布与预测响应值的分布进行对齐：

$$\hat{y}_t^* = \frac{\hat{y}_t - \hat{\mu}_t^{robust}}{\hat{\delta}_t^{robust}} \cdot \hat{\delta}_s^{robust} + \hat{\mu}_s^{robust} \quad (13)$$

2.5. 对比方法

无监督动态正交投影(Unsupervised Dynamic Orthogonal Projection, uDOP) [12]模型转移方法是动态正交投影(DOP)结合光谱空间变换(SST)的无监督版本，主要用于没有标签数据的场景。该方法依赖源域数据和响应值以及目标域数据中单独的光谱测量值，通过对源域和目标域的奇异值分解(Singular Value Decomposition, SVD)得到负载和投影，投影并结合源域负载重新构建目标域中的目标样本。

域不变最小二乘法(Domain-invariant Partial Least Squares, di-PLS) [13]方法是在普通 PLS 的基础上加入了域正则化，在最大化潜在变量投影域目标属性的协方差以保证模型的预测能力的前提下，通过惩罚源域和目标域数据在潜在空间中投影的方差差异，实现两域数据在潜在空间的分布对齐，从而提取对域间系统差异不敏感的域不变特征。

2.6. 评价指标

为了全面评价所建立模型的性能，采用预测均方根误差(Root Mean Square Error of Prediction, RMSEP)和残差预测偏差(Residual Predictive Deviation, RPD) [14]作为评价指标。预测均方根误差是通过计算预测值与真实值之间的误差来衡量模型的预测误差，数值越小，表示模型的预测能力越强；其计算公式为：

$$\text{RMSEP} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (14)$$

其中， y_i 是第 i 个观测值， \hat{y}_i 是模型预测的第 i 个值， n 是观察值的总数。

残差预测偏差是一种用于评估校准模型在预测数据集上性能的统计指标，其定义为总体样本标准差(δ)与预测均方根误差。RPD 值越大，表示实际值与预测值之间的一致性越好，反映出模型具有更强的泛化能力，能够在未知样本上进行准确的预测。其计算公式为：

$$\text{RPD} = \frac{\delta}{\text{RMSEP}} \quad (15)$$

3. 结果与分析

3.1. 无异常情况下的结果讨论

图 2 是生成的仿真光谱，表 1 汇总了 MCD-uDAR 及其它无监督模型转移方法的预测结果，以评估在引入鲁棒机制后模型校正性能的变化情况。实验结果表明，以 Y_1 作为响应变量的转移任务中，MCD-

uDAR 整体预测性能与评估来自源域测试集的性能表现高度一致，其 RMSEP 与 RPD 指标基本相同。图 3 则说了未经校准的时候，其分布有明显的差异，但经过 MCD-uDAR 方法校准后能够有效改善此类情况；在以 Y_2 和 Y_3 为响应变量的转移任务中，MCD-uDAR 利用基于 MCD 估计构建的稳健马氏距离，图 4 在预测空间中分别识别出 4 个和 1 个异常样本。对于被判定为异常的预测点，其对应的样本在后续分布校正与性能评估阶段被视为不可靠观测，因此在评价指标计算时予以剔除。该设计旨在避免少量异常样本对分布参数估计及模型转移效果评估产生不成比例的干扰，从而更真实地反映模型在主体样本上的预测能力。此外，对比方法中，di-PLS 在该仿真场景下表现出良好的预测性能，而 uDOP 方法可能未能有效捕捉源域和目标域之间的真实系统系统差异，导致其校准效果相对较差。

3.2. 分布漂移异常样本的结果讨论

本文通过在目标域中构造具有分布偏移特征的异常样本来模拟现实中的分布漂移现象。具体方法是在目标域样本中引入均值偏移量为 4、协方差缩放因子为 3 的合成异常样本。该设置不仅使目标域样本在特征空间中出现显著的整体平移，同时扩大了样本内部的离散程度，从而较为全面地模拟了实际应用中可能出现的复杂分布变化。通过向目标域样本中分别添加 10% 至 40% 不同比例的异常样本，模拟了不同程度的光谱整体分布偏移情形。不同异常样本比例下的目标域光谱如图 5 所示，可见随着异常样本

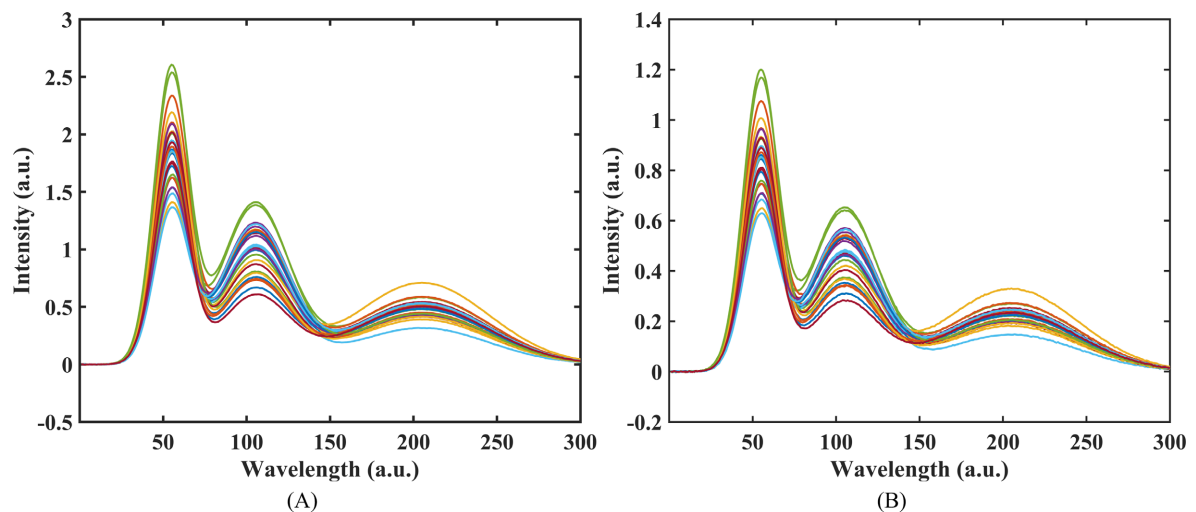


Figure 2. Spectra of simulated datasets: (A) Source domain; (B) Target domain

图 2. 模拟数据集光谱: (A) 源域; (B) 目标域

Table 1. Evaluation results of different model transfer method on Simulation Data

表 1. 不同模型转移算法在仿真数据上的评估结果

| 转移方法 | RMSEP | | | RPD | | |
|-------------------|--------|--------|--------|---------|---------|---------|
| | Y_1 | Y_2 | Y_3 | Y_1 | Y_2 | Y_3 |
| None ^a | 0.0121 | 0.0115 | 0.012 | 61.9851 | 71.0945 | 56.4196 |
| None | 2.6108 | 2.7509 | 2.74 | 0.2887 | 0.2979 | 0.2477 |
| uDOP | 2.6063 | 2.7647 | 2.7411 | 0.2831 | 0.2902 | 0.2424 |
| di-PLS | 0.0771 | 0.1997 | 0.1911 | 9.5765 | 4.0189 | 3.4769 |
| MCD-uDAR | 0.0123 | 0.0118 | 0.0111 | 60.9567 | 45.2814 | 55.3318 |

注: a 为评估了来自源域的测试集; b 为评估了来自目标域的测试集。

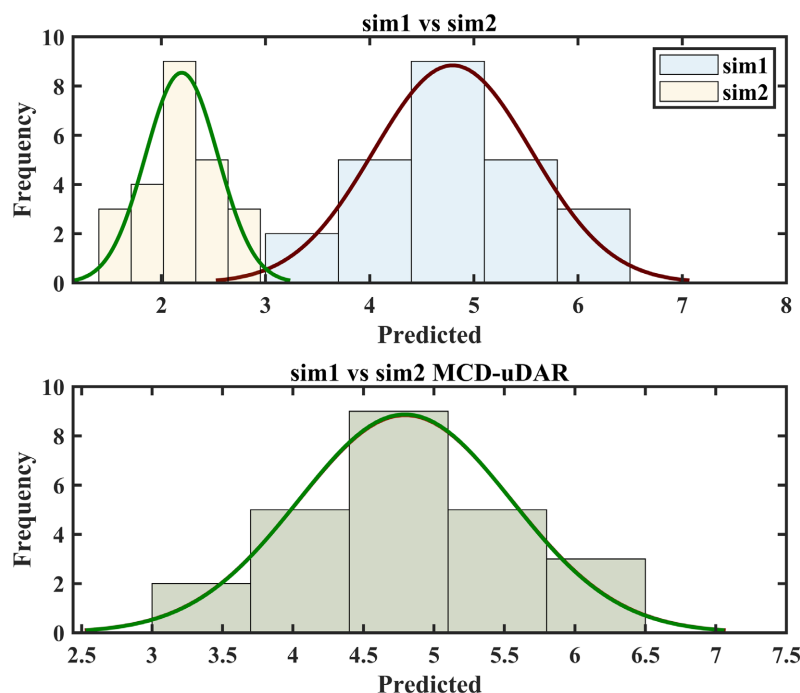


Figure 3. Prediction histograms of the source domain model for the target domain in the simulation data, along with the fitted normal distribution curves

图 3. 仿真数据中源域模型对目标域的预测直方图与正态分布拟合曲线

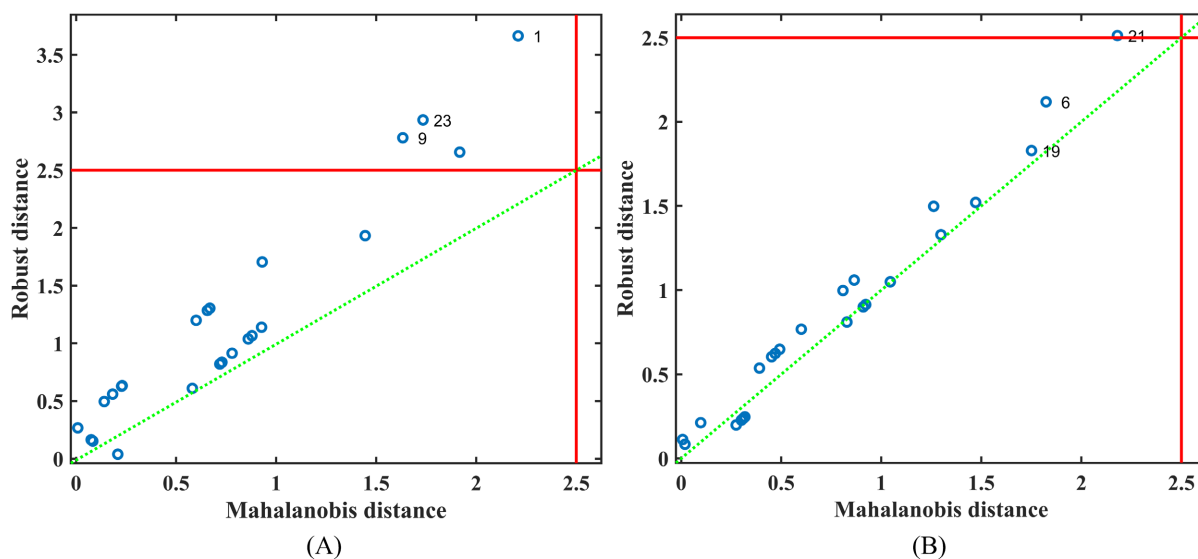


Figure 4. Martens-Robust distance for simulated data; (A) Y_2 as response variable; (B) Y_3 as response variable

图 4. 仿真数据的马氏 - 鲁棒距离; (A) Y_2 为响应变量; (B) Y_3 为响应变量

比例的增加, 各特征波峰的形状与强度差异将会更加明显。表 2 的实验结果表明, 在该类分类漂移异常样本下, MCD-uDAR 在该场景下表现出更稳定的预测性能。即使在异常样本比例达到 40% 的情况下, 其 RMSEP 仍为 0.9473, 表现出较强的抗分布漂移能力。由于引入鲁棒统计的思想, MCD-uDAR 通过 MCD 方法对预测响应变量的分布进行稳健估计, 从而获得对异常样本不敏感的均值和标准差。在分布对齐过程中, 明显偏离主体分布的异常样本不会主导统计结果, 从而有效减弱了异常样本对模型转移过程的干

扰, 使得目标域预测响应变量的主体分布特征能够被更加准确地刻画。图 6 给出了在不同异常样本比例条件下仿真数据在 MCD 模型中的鲁棒马氏距离分布情况。可以观察到, 部分样本的鲁棒距离明显超过预设阈值, 表明这些样本在特征空间中显著偏离主要数据分布, 具有典型的异常样本特征。MCD 能够对这些样本进行自动识别与剔除, 从而在统计估计阶段降低其影响, 进一步提高模型的泛化能力与稳定性。

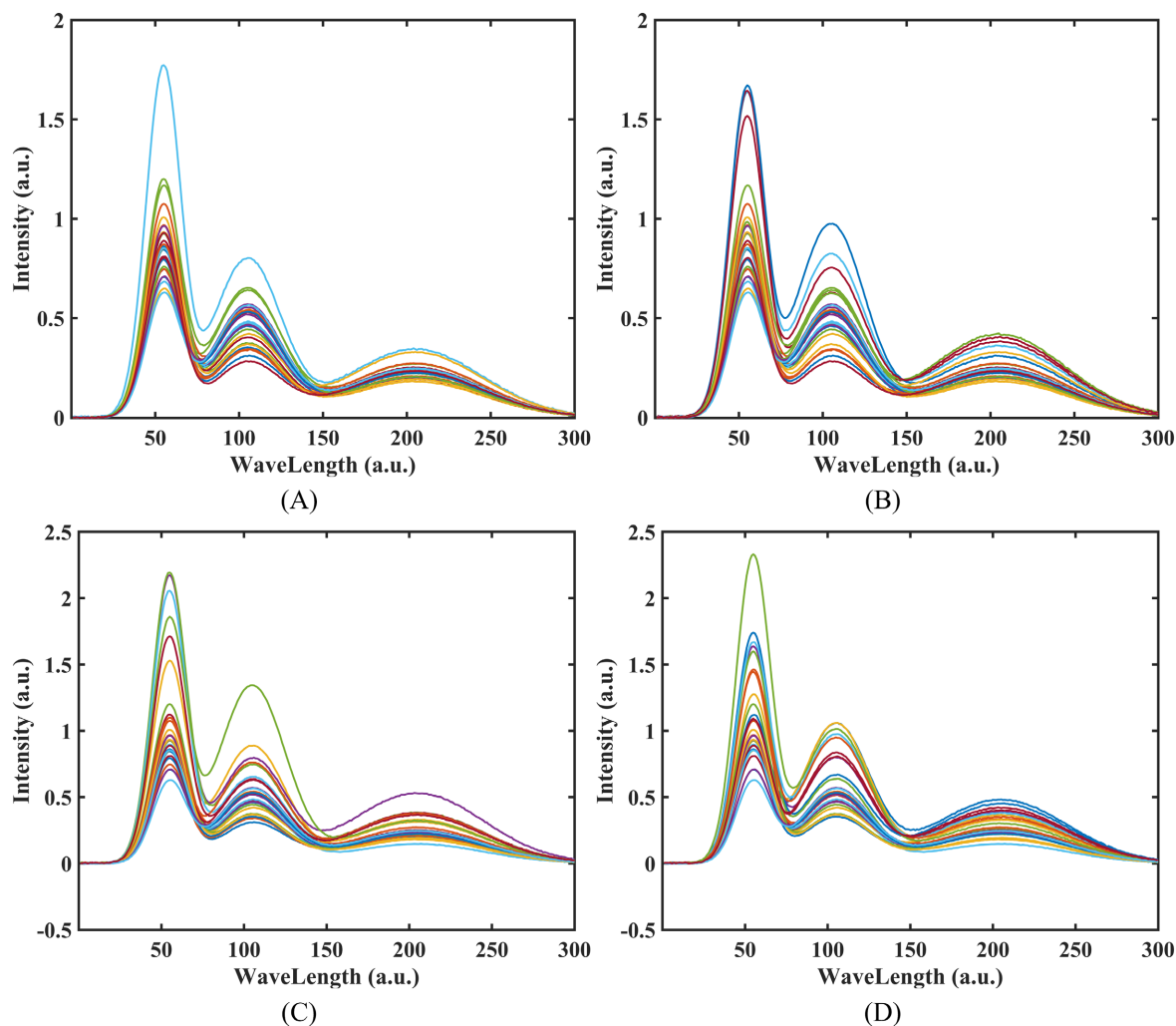


Figure 5. Target domain spectra with varying proportions of outliers incorporated: (A) 10%; (B) 20%; (C) 30%; (D) 40%
图 5. 掺入不同比例异常值的目标域光谱; (A) 10%; (B) 20%; (C) 30%; (D) 40%

Table 2. Evaluating the performance of different model transfer algorithms on simulation data using Y_1 as the response variable
表 2. 以 Y_1 为响应变量评估不同模型转移算法在仿真数据上的评估结果

| 转移方法 | 10% | | 20% | | 30% | | 40% | |
|----------|--------|---------|--------|--------|--------|--------|--------|--------|
| | RMSEP | RPD | RMSEP | RPD | RMSEP | RPD | RMSEP | RPD |
| None | 2.5489 | 0.2957 | 2.38 | 0.3167 | 2.4781 | 0.3042 | 2.2879 | 0.3295 |
| uDOP | 2.751 | 0.2683 | 2.3933 | 0.3083 | 2.669 | 0.2765 | 2.3533 | 0.3136 |
| di-PLS | 0.5905 | 1.2497 | 0.6287 | 1.1738 | 0.9225 | 0.7999 | 0.8655 | 0.8526 |
| MCD-uDAR | 0.0359 | 21.7658 | 0.3369 | 2.1706 | 0.2919 | 2.4716 | 0.9473 | 0.831 |

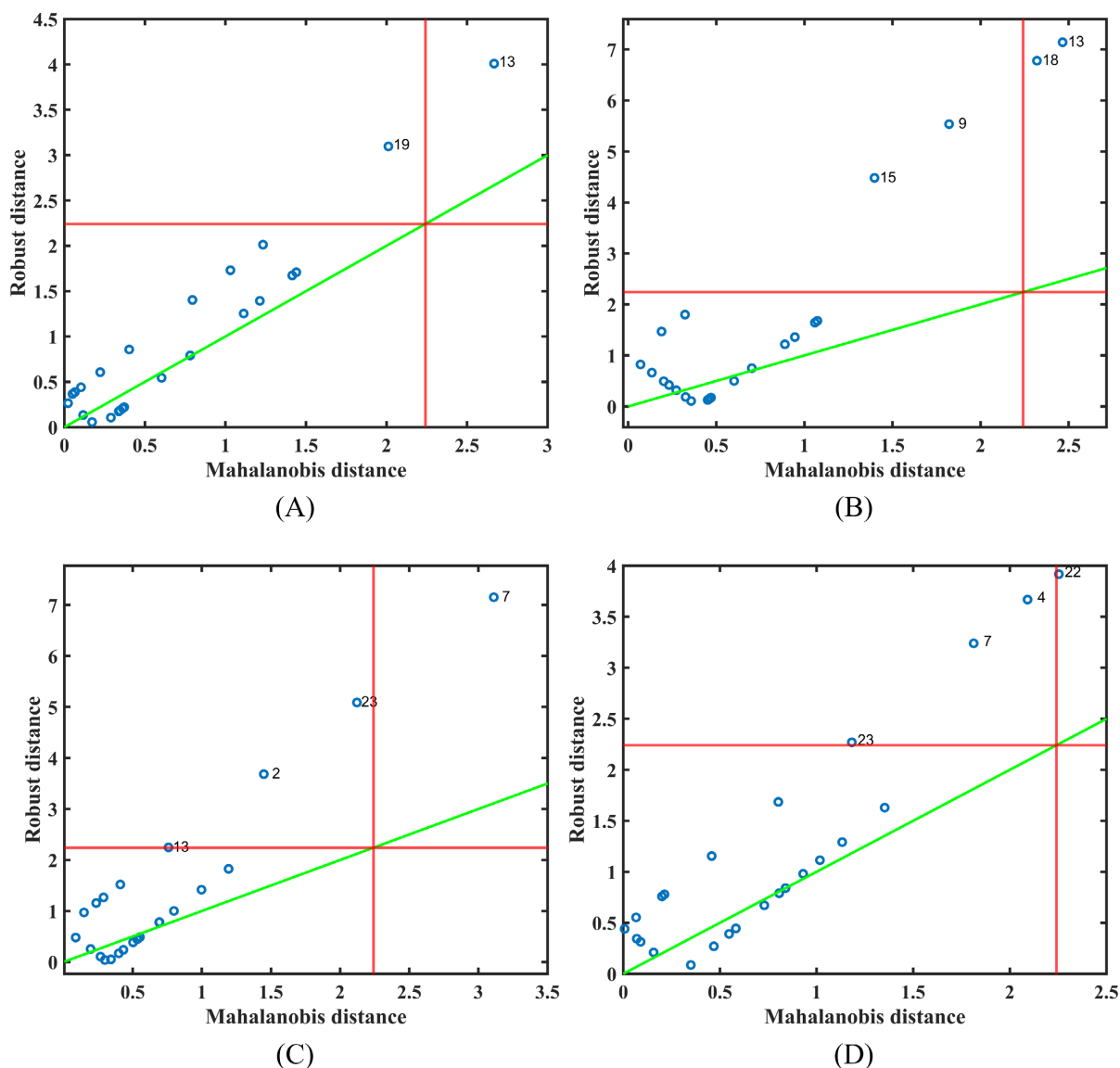


Figure 6. Martens-Robust distance for simulated data; (A) 10%; (B) 20%; (C) 30%; (D) 40%

图 6. 仿真数据的马氏 - 鲁棒距离图; (A) 10%; (B) 20%; (C) 30%; (D) 40%

3.3. 大豆数据集的结果讨论

图 7 展示了两台仪器测得的大豆数据的光谱, 在该数据中, 以 R1 为源域, R2 为目标域进行跨仪器场景下的讨论, 表 3 则总结了不同算法对该数据集的预测结果, 结果表明, 未经校准的情况下, 以 R1 仪器建立的 PLS 模型在对 R2 进行预测时, 其 RMSEP 为 20.5873, RPD 为 0.1755, 其结果正如图 8 中上子图未经校准的结果一致, 其分布存在显著差异, 而经过 MCD-uDAR 方法校准后的 RMSEP 仅为 0.9917, RPD 为 2.0919, 经过校正后的两台仪器分布曲线如图 8 中的下子图所示, 能够实现完美的重合, 表明两者之间以无明显差异, 优于用于对比的 di-PLS 和 uDOP 算法, 值得意外的, uDOP 未能实现有效校正, 预测性能进一步劣化, RMSEP 升至 32.5399, RPD 为 0.1079. 图 9 给出了在大豆数据集在 MCD 模型中的鲁棒马氏距离分布情况, 可以观察到基于 MCD 的稳健距离可更灵敏地识别偏离主体分布的异常样本, 有效避免异常值对响应变量分布均值、标准差估计的干扰, 为后续分布对齐提供可靠的参数支撑, 显著

Table 3. Evaluation results of different calibration transfer methods on soy
表 3. 不同模型转移算法在大豆上的评估结果

| 转移方法 | RMSEP | RPD |
|----------|---------|--------|
| None | 20.5873 | 0.1755 |
| uDOP | 32.5399 | 0.1079 |
| di-PLS | 1.3308 | 2.6392 |
| MCD-uDAR | 0.9917 | 2.0919 |

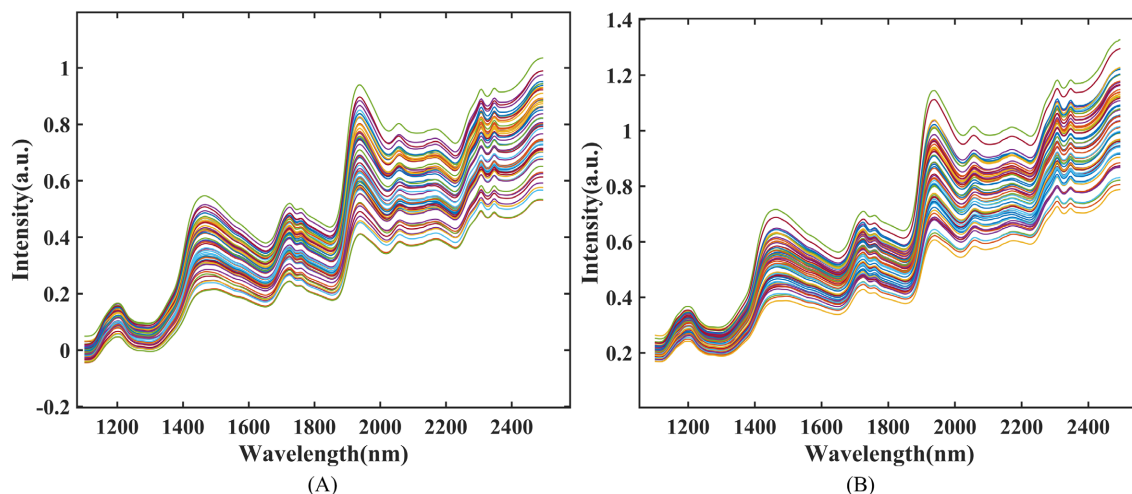


Figure 7. Spectra of simulated datasets: (A) R1; (B) R2
图 7. 大豆数据集; (A) R1; (B) R2

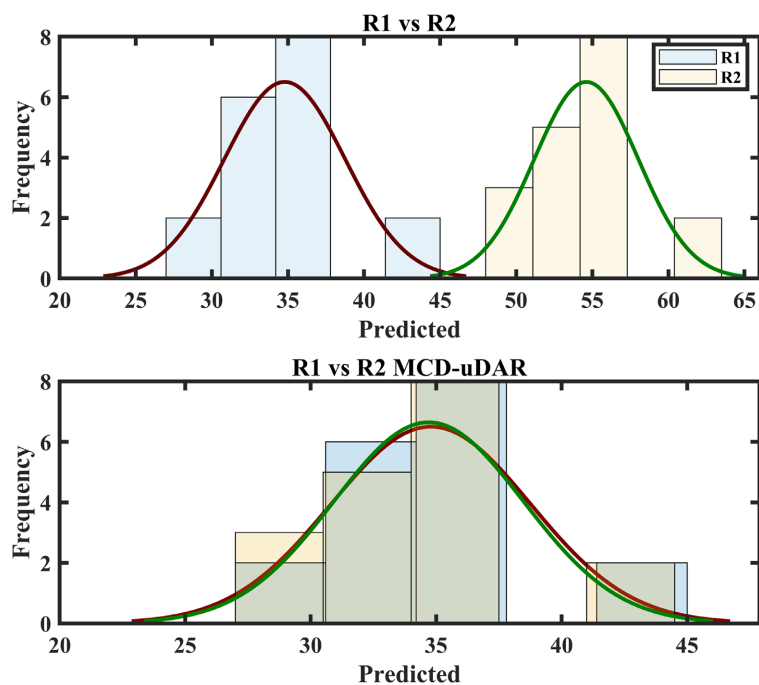


Figure 8. Prediction histograms of the source domain model for the target domain in the soy data, along with the fitted normal distribution curves

图 8. 大豆数据中源域模型对目标域的预测直方图与正态分布拟合曲线

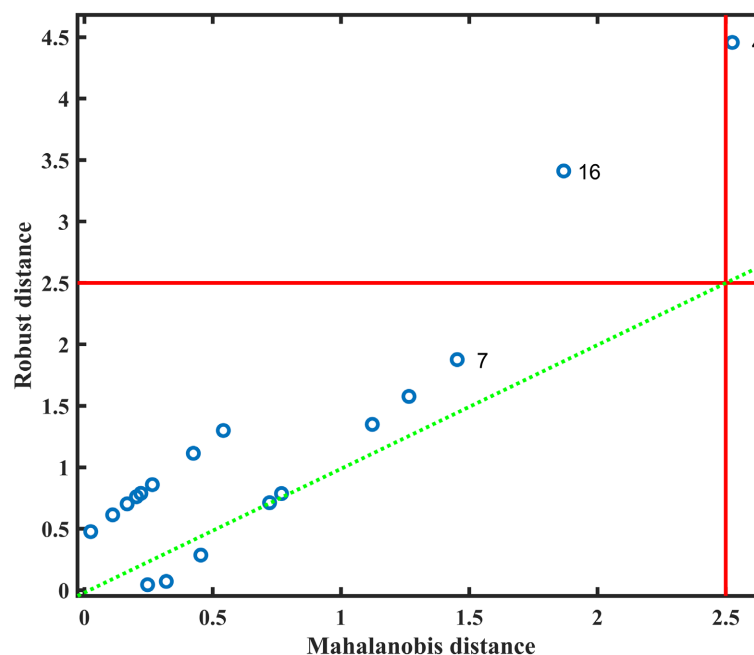


Figure 9. Martens-Robust distance for Soys data
图9. 大豆数据的马氏 - 鲁棒距离图

提升了模型转移过程的稳定性与抗干扰能力。

4. 结论

本文针对传统无监督模型转移方法在异常样本时表现不稳定的问题，提出了一种结合最小协方差行列式的分布对齐模型转移方法(MCD-uDAR)。该方法通过对预测响应变量进行稳健均值和标准差的估计，有效抑制了异常值对分布对齐过程中的干扰。在仿真数据集和大豆数据集上的验证表明，在存在异常样本的情况下，MCD-uDAR 仍能保持较好的预测性能，且显著优于 uDOP 和 di-PLS 方法。实验结果表明，引入鲁棒机制不仅能有效识别并减弱异常样本的影响，还可在无监督条件下实现目标域数据分布的准确对齐，从而为复杂实际场景提供了一种更为可靠、稳健的模型转移方案。

参考文献

- [1] Liu, J., Zhang, J., Tan, Z., Hou, Q. and Liu, R. (2022) Detecting the Content of the Bright Blue Pigment in Cream Based on Deep Learning and Near-Infrared Spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, **270**, Article ID: 120757. <https://doi.org/10.1016/j.saa.2021.120757>
- [2] Shi, C. and Sang, J. (2025) Qualitative Recognition of Waste Textiles Based on Near-Infrared Spectroscopy and ModernTCN. *Microchemical Journal*, **214**, Article ID: 113902. <https://doi.org/10.1016/j.microc.2025.113902>
- [3] Fearn, T. (2001) Standardisation and Calibration Transfer for near Infrared Instruments: A Review. *Journal of Near Infrared Spectroscopy*, **9**, 229-244. <https://doi.org/10.1255/jnirs.309>
- [4] Dong, X., Dong, J., Li, Y., Xu, H. and Tang, X. (2019) Maintaining the Predictive Abilities of Egg Freshness Models on New Variety Based on VIS-NIR Spectroscopy Technique. *Computers and Electronics in Agriculture*, **156**, 669-676. <https://doi.org/10.1016/j.compag.2018.12.012>
- [5] 冯佳豪, 程泉翔, 张智勇, 等. 淫羊藿与柔毛淫羊藿中朝藿定 A、B、C 和淫羊藿苷的近红外光谱定量分析及模型转移研究[J]. *分析测试学报*, 2025, 44(6): 995-1004.
- [6] Mishra, P. and Nikzad-Langerodi, R. (2021) A Brief Note on Application of Domain-Invariant PLS for Adapting Near-Infrared Spectroscopy Calibrations between Different Physical Forms of Samples. *Talanta*, **232**, Article ID: 122461. <https://doi.org/10.1016/j.talanta.2021.122461>

-
- [7] Rabatel, G., Marini, F., Walczak, B. and Roger, J. (2019) VSN: Variable Sorting for Normalization. *Journal of Chemometrics*, **34**, e3164. <https://doi.org/10.1002/cem.3164>
- [8] Xie, Z., Chen, X., Roger, J., Ali, S., Huang, G. and Shi, W. (2024) Calibration Transfer via Filter Learning. *Analytica Chimica Acta*, **1298**, Article ID: 342404. <https://doi.org/10.1016/j.aca.2024.342404>
- [9] Verboven, S., Hubert, M. and Goos, P. (2012) Robust Preprocessing and Model Selection for Spectral Data. *Journal of Chemometrics*, **26**, 282-289. <https://doi.org/10.1002/cem.2446>
- [10] 何佳艳, 李亭, 郭长凯, 等. 近红外光谱法快速无损测定奶粉的脂肪含量[J]. 食品与发酵工业, 2017, 43(10): 228-233.
- [11] Rousseeuw, P.J. (1984) Least Median of Squares Regression. *Journal of the American Statistical Association*, **79**, 871-880. <https://doi.org/10.1080/01621459.1984.10477105>
- [12] Fonseca Diaz, V., Roger, J. and Saeyns, W. (2022) Unsupervised Dynamic Orthogonal Projection. An Efficient Approach to Calibration Transfer without Standard Samples. *Analytica Chimica Acta*, **1225**, Article ID: 340154. <https://doi.org/10.1016/j.aca.2022.340154>
- [13] Nikzad-Langerodi, R., Zellinger, W., Lughofer, E. and Saminger-Platz, S. (2018) Domain-Invariant Partial-Least-Squares Regression. *Analytical Chemistry*, **90**, 6693-6701. <https://doi.org/10.1021/acs.analchem.8b00498>
- [14] Chang, C., Laird, D.A., Mausbach, M.J. and Hurburgh, C.R. (2001) Near-Infrared Reflectance Spectroscopy-Principal Components Regression Analyses of Soil Properties. *Soil Science Society of America Journal*, **65**, 480-490. <https://doi.org/10.2136/sssaj2001.652480x>