

深度自编码器的近红外光谱转移研究

刘贞文¹, 徐玲杰^{2*}, 陈孝敬³

1. 集美大学港口与环境工程学院, 福建 厦门 361021
2. 温州大学机电工程学院, 浙江 温州 325035
3. 温州大学电气与电子工程学院, 浙江 温州 325035

摘要 近红外光谱分析中多变量校准模型的建立依赖于校准建模的光谱样本。然而, 近红外光谱测量环境的变化会导致同一被测物的光谱样本的偏移。为了削减光谱偏移后重新建立校准模型的成本, 提出一种基于深度自编码器(DAE)的非线性光谱转移方法, 以端到端的形式实现不同测量环境之间的光谱转移, 避免已有的线性光谱转移方法在非线形偏移光谱时效果不佳的情况。该方法在操作前不需要对光谱进行预处理和特征提取等操作, 可以实现原始光谱之间的转移, 是首个端到端的非线性光谱转移方法。为了实现光谱空间的有效转移, 设计了一种基于条件概率和参数最大似然法的误差函数惩罚项, 结合梯度反向传播算法优化深度自编码的网络参数。为了验证该方法的有效性, 引入两个公共的近红外光谱数据集, 分别是药片数据集和玉米数据集。利用本方法进行光谱转移的过程主要有: 根据 Kennard-Stone(KS)算法分别将两个数据集划分为校准集、验证集和测试集; 用校准集中的光谱样本输入深度自编码器, 根据设计的误差函数求出误差, 并用反向传播法迭代训练网络参数, 直至模型最优; 将预测集样本输入训练好的 DAE 转移模型, 可以发现转移后的光谱与相应的目标光谱谱线基本重合, 这说明该设计的转移模型的有效性。最后, 为了进一步验证本方法的优越性, 将该方法与经典的线性转移算法光谱空间变换(SST)和分段直接标准化(PDS)进行比较。这三种算法得到的转移光谱分别作为测试样本, 输入已建立的偏最小二乘(PLS)多变量校准模型, 通过比较预测均方根误差(RMSEP), 可以发现该方法在多变量校准模型中的预测结果的均方根误差均小于 SST 和 PDS, 分别提高了 5.7% 和 10.1%, 表明由非线性深度自编码器转移的光谱样本具有高效和实用的特点。

关键词 近红外光谱; 模型转移; 多变量校准; 深度自编码器

中图分类号: O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)07-2313-06

引言

近红外光谱已广泛应用于各个领域, 是一种快速、无损、环保、经济的实用技术^[1]。然而, 近红外光谱的微弱信号不仅包含化学信息, 还包含各种背景噪声。因此, 为了从光谱中得到准确的定量分析结果, 一个可靠的多元校正模型是必不可少的。在实际应用中, 可以使用一种以上的仪器来收集光谱。将一种仪器(主仪器)的校准模型直接应用于另一种仪器(次仪器)的光谱测量, 会使该模型失效。而针对新光谱的重新校准是费时的^[2-3]。校准转移的方法可以使用现有的模型来分析在新条件下(新仪器下)获得的新样品, 而不需要重新建立模型, 因此受到许多关注和研究^[4]。

模型校准转移主要可以通过三种策略实现: 模型预测值标准化、模型系数标准化以及光谱标准化(光谱转移)。简单的单变量斜率和偏差校正(SBC)^[5]在两个不同仪器上测量的光谱预测之间建立了线性方程。通过线性方程的偏置和斜率, 可以对新光谱的预测进行校正。但只有当仪器变化引起所有样本系统的光谱差异时, SBC 才是有效的^[6]。与 SBC 方法不同, 模型系数的标准化可以应用于不同样本间光谱差异的情况, 但通常这些方法需要大量的光谱样本才能得到满意的参数校正结果。通过标准化光谱进行模型转移是最常用的模型转移策略, 这类方法通过一个转移矩阵, 将次仪器上测量的样品的光谱转移为相应的主仪器上的光谱, 使其服从于主仪器的样本空间分布。

直接标准化(DS)^[7]以及分段直接标准化(PDS)^[8]是最具

收稿日期: 2019-10-08, 修订日期: 2020-01-30

基金项目: 集美大学科研启动资金项目(ZQ2019026)资助

作者简介: 刘贞文, 1979年生, 集美大学港口与环境工程学院副教授

e-mail: Liuzhenwen@jmu.edu.cn

* 通讯联系人 e-mail: grapefruitxlj@163.com

有代表性的标准化光谱方法。DS 直接将主仪器上测得的样品的光谱联系起来,并用变换矩阵来描述这种关系,主光谱的每个波长与次光谱的所有波长同时相关。然而,在 PDS 中,频谱相关被限制在一个更小的区域内。它对一个小窗口区域内的不同仪器测量的光谱之间建立了线性关系,以实现整体波段的非线性转换。典型相关分析(CCA)^[9]和光谱空间变换(SST)^[10]并没有直接对光谱进行校正,而是从不同光谱中提取出主成分之间的关系。SST 通过对不同仪器上测量的光谱进行对比,构建数据矩阵,并通过主成分分析(PCA)来估计光谱转换矩阵。CCA 是一种线性子空间学习方法,它先用子空间算法降低主光谱的维数,而后在低子空间中计算了光谱间变换矩阵。

以上提到的算法,都是基于线性变换实现预测值、模型系数和光谱的转移。然而,在不同仪器上测量的近红外光谱不仅包含遵循朗伯-比尔定律的吸收信号,还包含与仪器状态,测量条件,粒径,甚至温度和湿度相关的信号^[11],这导致光谱之间非线性的差异。因此,仅使用简单的线性模型很难描述在不同仪器上测量的近红外光谱的差别^[12]。本工作使用深度自编码(deep autoencoder, DAE)的方法,建立了不同仪器之间的非线性映射,在两个公共数据集上表现优于其他传统线性光谱标准化方法。

1 原理与方法

1.1 深度自编码器神经网络

深度自编码(DAE)是深度神经网络(DNN)的一种特殊类型。DAE 是隐藏层不少于两层的自编码器,它由编码器和解码器两个部分组成。作为一种非线性特征提取方法,编码器建立了自适应多层编码网络,将高维数据转换为低维码。解码器使用了与编码器相同网络结构,可以从代码中重构原始数据。每一隐藏层的权值代表输入变量的不同重构。为了恢复输入信号,DAE 必须捕获输入最重要的特性。在编码过程中,使用反向传播算法最小化原始数据与重构数据之间的差异来训练隐含层的权值。Ting 等^[13]利用 DAE 提取近红外光谱的主要特征,并将此方法用于近红外光谱的分类。

图 1 为隐藏层数三层的 DAE。从该图中可以看出,DAE 的输入与输出维度相同,且编码器与解码器具有对称的网络结构。万能近似定理^[14]保证至少有一层隐藏层且隐藏单元足够多的前馈神经网络能以任意精度近似任意函数(在很大范围内),这是非平凡深度(至少有一层隐藏层)的一个主要优点^[15]。所以,深度自编码器在给定足够多的隐藏单元的情况下,能以任意精度近似任何从输入到编码的映射。这也是深度自编码器优于浅层自编码器之处。

1.2 深度自编码器进行光谱标准化的过程

光谱样本对 (x_1, x_2) 代表仪器一和仪器二对同一个样本测量后得到的谱线。设 L 是自编码器的均方误差函数,惩罚自编码器输出 $\hat{x}_2 = g(f(x_1))$ 与 x_2 之间的差异,则对单个样本,DAE 的误差方程可以表示为

$$L = \frac{1}{2} \|\hat{x}_2 - x_2\|^2 \quad (1)$$

假设仪器一样本数据集为 $X_1 = \{x_1^1, x_1^2, \dots, x_1^m\}$, 样本分布空间为 \hat{p}_1 ; 仪器二样本数据集为 $X_2 = \{x_2^1, x_2^2, \dots, x_2^m\}$; 样本分布空间为 \hat{p}_2 。

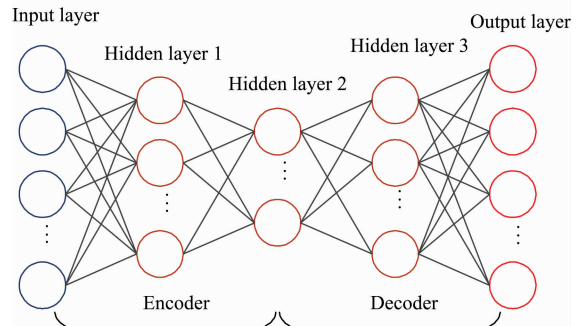


图 1 具有三层隐藏层的深度自编码器

Fig. 1 DAE with three encoder layers

所以整体的损失函数为

$$L = \frac{1}{m} \sum_{i=1}^m \frac{1}{2} \|\hat{x}_2^i - x_2^i\|^2 \quad (2)$$

为了实现两个仪器光谱样本之间空间的映射,我们根据以下数理统计的原理引入针对 DAE 参数的误差惩罚项。我们希望从 \hat{p}_1 中采样的样本,经过深度自编码器(DAE)的重构之后,得到服从 \hat{p}_2 分布的相应光谱样本,设这个重构过程为 $p_{\text{reconstruct}}(X_2 | X_1; \theta)$ (在给定输入 X_1 的条件下得到 X_2 的概率),其中 θ 代表 DAE 的网络参数。根据条件概率最大似然估计定理,对 θ 的条件最大似然估计是

$$\theta_{\text{ML}} = \operatorname{argmax}_{\theta} p_{\text{reconstruct}}(X_2 | X_1; \theta) \quad (3)$$

因为样本之间都是独立同分布的,所以式(3)可以表示为

$$\theta_{\text{ML}} = \operatorname{argmax}_{\theta} \prod_{i=1}^m p_{\text{reconstruct}}(x_2^i | x_1^i; \theta) \quad (4)$$

用 \log 函数将式(4)中的乘积转化为求和的形式,即对数似然

$$\begin{aligned} \theta_{\text{ML}} &= \operatorname{argmax}_{\theta} \sum_{i=1}^m \log p_{\text{reconstruct}}(x_2^i | x_1^i; \theta) \\ &= \operatorname{argmin}_{\theta} \left[- \sum_{i=1}^m \log p_{\text{reconstruct}}(x_2^i | x_1^i; \theta) \right] \end{aligned} \quad (5)$$

故可以对 DAE 的误差函数 L 加入针对参数 θ 的惩罚项

$$L' = - \sum_{i=1}^m \log p_{\text{reconstruct}}(x_2^i | x_1^i; \theta) \quad (6)$$

得到以下惩罚项函数

$$L' = - \sum_{i=1}^m x_2^i \log(x_2^i) + (1 - x_2^i) \log(1 - x_2^i) \quad (7)$$

其中 \hat{x}_2^i 是 DAE 的输出。

根据式(2)和式(7)整体损失函数可以定义为

$$\begin{aligned} L &= \frac{1}{m} \sum_{i=1}^m \frac{1}{2} \|\hat{x}_2^i - x_2^i\|^2 + \lambda L' \\ L &= \frac{1}{m} \sum_{i=1}^m \frac{1}{2} \|\hat{x}_2^i - x_2^i\|^2 - \lambda \sum_{i=1}^m x_2^i \log \hat{x}_2^i + \\ &\quad (1 - x_2^i) \log(1 - \hat{x}_2^i) \end{aligned} \quad (8)$$

其中 λ 为惩罚项权重参数。得到损失函数后,即可用梯度反向传播方法对网络中的参数 θ 进行迭代优化,其中 α 是学习率。

$$\theta = \theta - \alpha \frac{\partial L}{\partial \theta} \quad (9)$$

2 数据集介绍

利用两个近红外光谱数据集对该方法的有效性进行了验证。第一个是国际漫反射会议网站发布的药品片剂数据集 (<http://www.idrc-chambersburg.org/shootout2002.html>)。该数据集包含了 655 个药物片剂样品的近红外光谱以及每个样本的重量、硬度和活性药物成分(API)的测定值。利用美国 Foss 公司的两台光谱仪,测定的光谱波长范围为 600~1 898 nm,分辨率为 2 nm(纳米)。该数据集的 655 个样本被分为校准、验证和预测集三个部分,分别包含 155, 40 和 460 个样本。两台仪器得到的光谱样本被用来互相转移,使用 API 验证光谱转移后定量校准模型的预测效果。在计算中,由于 1 794~1 898 nm 波段的光谱噪声太大,无法获得合理的校准结果^[16],因此使用了建议的 600~1 792 nm 波段的 597 个变量。数据集中的可能的不合理谱线:校准集中的第 19, 122, 126 和 127 条,预测集中的第 11, 145, 267, 295, 342, 313, 341 和 343 条谱线被剔除^[6],剩余的数据集中包含 642 个样本。

第二个数据集是玉米数据集 (<http://software.eigenvector.com/Data/Corn/index>),包含了 m5, mp5, mp6 三种仪器上测量的近红外光谱,以及 80 个玉米样品的水分、油脂、蛋白质和淀粉含量。每条光谱测量范围为 1 100~2 498 nm,分辨率为 2 nm。三种仪器测量的光谱被用来互相校正,用蛋白质含量来验证光谱转移后定量校准模型的预测结果。

在计算之前,每个数据集都会根据经典的 Kennard-Stone(KS)^[17]算法被划分为校准集、验证集和预测集。药片集的 642 个样本被分为 400 个样本的校准建模集,以及各包含 121 个样本的验证集和预测集。玉米集的 80 样本被分为 40, 20, 20 的校准建模集、验证集和预测集。校准建模集和验证集的光谱样本会被用于建立定量分析的多变量回归模型以及训练用于光谱转移的 DAE 模型。预测集的光谱样本用来检验 DAE 的效果,同时检验转移后的光谱在原多变量回归模型中的预测结果。

3 结果与讨论

3.1 基于 DAE 的光谱校正转移

针对药片数据集和玉米数据集,我们分别设计了两个编码器层数为四层的深度自编码器网络,结构分别为[597 400 200 100 200 400 597]和[700 350 175 80 175 350 700]。因此,输入的仪器一的光谱维度,经过编码器从 597 逐步变为 400, 200, 100, 再经过对称的解码器,从维度 100 逐步变为 200, 400, 597 的输出光谱,输出的维度为 597 的光谱即为我们所需的符合仪器二样本分布空间的光谱。自编码器的损失函数和参数优化方式根据式(8)和式(9)获得。进行药片数据集光谱转移的自编码器网络结构如图 2 所示。玉米数据集的

样本也同上进行光谱的转移。

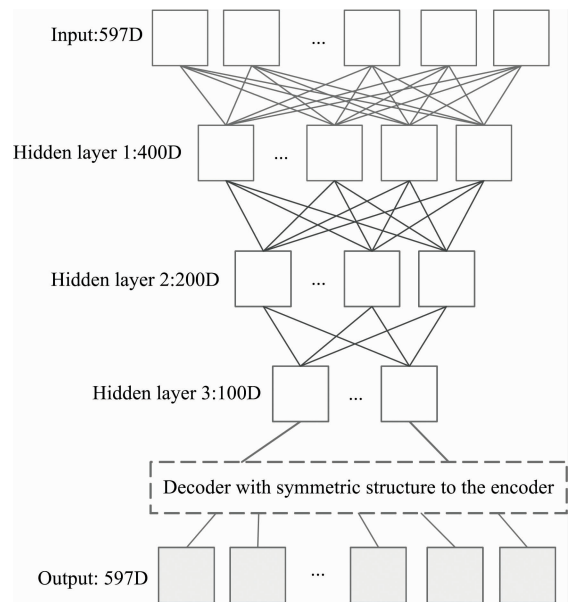


图 2 针对药片数据集光谱转移的深度自编码器网络结构
Fig. 2 Network structure for spectral transfer of Tablet dataset

我们使用 sigmoid 激活函数来限制每层神经元的输出接近于 0。经过多次代码的运行,确定将式(8)中的惩罚系数 λ 设置为 0.01,将式(9)中的学习率 α 设置为 0.001,迭代训练 5 000 个回合后,模型收敛,并达到最佳的转移光谱的效果。

3.2 转移光谱的验证

图 3 展示了从两个数据集的预测集中随机选取的两台仪器上对应的光谱样本,对它们进行相互的光谱转移后的可视化结果。图 3(a)是药片数据集的两台仪器上对应的光谱相互转移的结果,实线是真实的一组具有相同属性值的,由不同仪器测量出的光谱样本,虚线是经过 3.1 中设计的深度自编码器模型转移后的光谱。从图 3(a)中可得,两条谱线的差异主要集中在波段 0~90 和 500~597 nm,而在其余波段内,两条光谱线几乎重合,说明这两条光谱之间存在着很明显的非线性偏移。在这主要差异的两个波段中,虚线与实线重合,这表明光谱得到了成功的转移。图 3(b)~(d)是玉米样本集的三台光谱仪 m5, mp5 和 mp6 上的光谱相互转移的结果。从三张图可以看出, m5 的光谱与 mp5 和 mp6 之间偏移较大,而 mp5 与 mp6 光谱偏移小。图中虚实线吻合,说明深度自编码器能够有效地对三台仪器得出玉米光谱样本进行两两之间的转移。

3.3 多变量校准模型验证

进行光谱转移是为了将偏移的光谱进行转移后,直接应用于已建立的多变量校准模型中,来获得有效的定性定量分析结果。因此,为了进一步研究本方法的有效性,将本方法与经典的线性转移方法 PDS 和 SST 算法中得到的转移光谱分别作为测试样本输入已建立的多变量校准模型,通过比较多变量模型输出的均方根误差(RMSEP)来验证转移的效果。多变量校准模型通过经典的偏最小二乘法(PLS)算法建立^[18]。通过 7 次交叉验证,确定了 PLS 模型中潜在变量

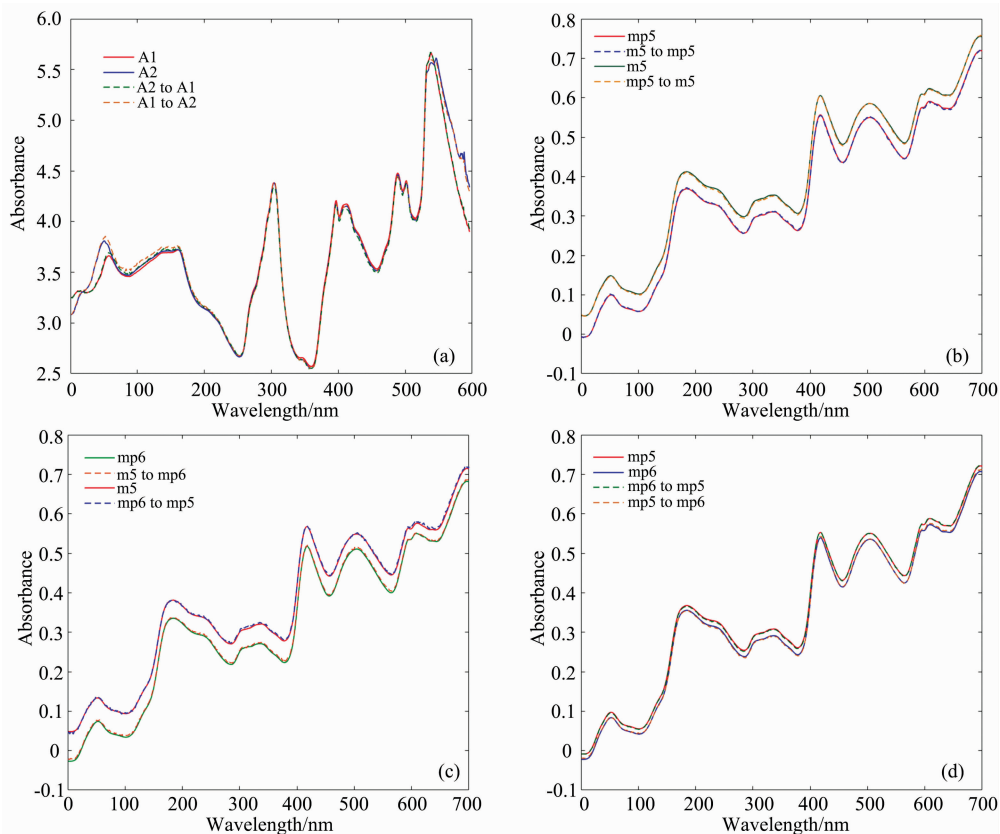


图 3 两个数据集的不同仪器上的光谱相互转移的结果, (a) 中的 A1 和 A2 代表药片数据集两台光谱仪上测出的光谱

Fig. 3 Transfer results of DAE for two data sets, A1 and A2 in figure (a) represent the spectra of the tablets measured by two instruments

表 1 药片数据集进行光谱转移后, 转移光谱在多变量校准模型中的 RMSEP

Table 1 RMSEP of transferred tablet spectrum in a multivariate calibration model

建模光谱	验证光谱	参数	RMSEP
A1	A1	nLV=8	2.893 3
	A2	nLV=8	8.875
	A2 to A1(PDS)	w=15, nLV=6	3.712 2
	A2 to A1(SST)	nPC=4	3.654 4
	A2 to A1(DAE)	$\lambda=0.01, \alpha=0.001$	3.472 9
A2	A2	nLV=9	2.971 4
	A1	nLV=9	6.068 7
	A1 to A2(PDS)	w=15, nLV=6	3.764 9
	A1 to A2(SST)	nPC=4	3.861 7
	A1 to A2(DAE)	$\lambda=0.01, \alpha=0.001$	3.578 1

注: nLV: 偏最小二乘法(PLS)的潜在变量个数; w: PDS 中选用的窗口尺寸; nPC: SST 算法中选取的主成分个数

Note: nLV: Number of latent variables in PLS; w: Window size used in PDS; nPC: Number of principal components in SST

(nLV)的最优数量。如表 1 和表 2 所示, 显然, 当用一台仪器的校准集光谱建立的模型直接预测另一台仪器的光谱, 效果很差。但当三种方法得到的转移光谱被预测时, RMSEP 显著降低。通过对三种方法的比较, 本方法得到的转移光谱

在原校准模型中的表现略优于常用的 SST 和 PDS 方法。不管是哪两台光谱仪之间的相互转移, 均得到了类似的结果。

表 2 玉米数据集进行光谱转移后, 转移光谱在多变量校准模型中的 RMSEP

Table 2 RMSEP of transferred corn spectrum in a multivariate calibration model

建模光谱	验证光谱	参数	RMSEP
m5	m5	nLV=19	0.098 5
	mp5	nLV=19	1.013
	mp6	nLV=19	1.085 4
	mp5 转 m5(PDS)	w=17, nLV=8	0.144 8
	mp5 转 m5(SST)	nPC=3	0.150 3
	mp5 转 m5(DAE)	$\lambda=0.01, \alpha=0.001$	0.133 9
	mp6 转 m5(PDS)	w=17, nLV=8	0.169 5
	mp6 转 m5(SST)	nPC=3	0.178 1
	mp6 转 m5(DAE)	$\lambda=0.01, \alpha=0.001$	0.162 9
mp5	mp5	nLV=6	0.148 3
	m5	nLV=6	0.553
	mp6	nLV=6	0.248 6
	m5 转 mp5(PDS)	w=17, nLV=8	0.155 7
	m5 转 mp5(SST)	nPC=3	0.161 1
	m5 转 mp5(DAE)	$\lambda=0.01, \alpha=0.001$	0.141

续表 2

	mp6 转 mp5(PSD)	w=17, nLV=8	0.15
	mp6 转 mp5(SST)	nPC=3	0.159 8
	mp6 转 mp5(DAE)	$\lambda=0.01, \alpha=0.001$	0.146 6
mp6	mp6	nLV=9	0.155 2
	m5	nLV=9	0.677 9
	mp6	nLV=9	0.192 2
	m5 转 mp6(PSD)	w=17, nLV=8	0.225 1
	m5 转 mp6(SST)	nPC=3	0.231 9
	m5 转 mp6(DAE)	$\lambda=0.01, \alpha=0.001$	0.201 3
	mp5 转 mp6(PSD)	w=17, nLV=8	0.169 7
	mp5 转 mp6(SST)	nPC=3	0.179 5
	mp5 转 mp6(DAE)	$\lambda=0.01, \alpha=0.001$	0.162 3

注: nLV: 偏最小二乘法(PLS)的潜变量个数; w: PDS 中选用的窗口尺寸; nPC: SST 算法中选取的主成分个数

Note: nLV: Number of latent variables in PLS; w: Window size used in PDS; nPC: Number of principal components in SST

4 结 论

研究了深度自编码在近红外光谱转移中的应用, 探索了一种非线性的光谱转移方法。通过数理统计的相关知识, 设计了符合光谱转移需求的误差函数。同时, 针对药片数据集和玉米数据集, 给出了详细网络结构和模型参数。将预测集样本输入训练好的 DAE 转移模型, 可以发现转移后的光谱与相应的目标光谱谱线基本重合, 说明本转移模型的有效性。最后, 为了进一步验证本方法的优越性, 将本方法与 SST 和 PDS 方法进行比较。将这三种算法得到的转移光谱分别作为测试样本, 输入已建立的偏最小二乘(PLS)多变量校准模型, 通过比较 RMSEP, 可以发现本方法在多变量校准模型中的预测结果的均方根误差均小于 SST 和 PDS, 平均提高了 5.7% 和 10.1%, 进一步证明了本方法高效和实用的优点。

References

- [1] Abdali S, Daneshfar A, Ghaedi M, et al. *Analytical Methods*, 2016, 8: 6074.
- [2] Huang G Z, Chen X J, Li L M, et al. *Chemometrics and Intelligent Laboratory Systems*, 2020, 201: 103986.
- [3] Feudale R N, Woody N A, Tan H, et al. *Chemometrics and Intelligent Laboratory Systems*, 2002, 64: 181.
- [4] Sulub Y, LoBrutto R, Vivilecchia R, et al. *Analytica Chimica Acta*, 2008, 611: 143.
- [5] Bouveresse E, Hartmann C, Massart D L, et al. *Analytical Chemistry*, 1996, 68: 982.
- [6] Zhang J, Guo C, Cui X, et al. *Analytica Chimica Acta*, 2019, 1050: 25.
- [7] Wang Y, Veltkamp D J, Kowalski B R. *Analytical Chemistry*, 1991, 63: 2750.
- [8] Tan H W, Brown S D. *Journal of Chemometrics*, 2001, 15: 647.
- [9] Fan W, Liang Y, Yuan D, et al. *Analytica Chimica Acta*, 2008, 623: 22.
- [10] Du W, Chen Z P, Zhong L J, et al. *Analytica Chimica Acta*, 2011, 690: 64.
- [11] Kalivas J H. *Journal of Chemometrics*, 2012, 26: 218.
- [12] Jansen J J, Hoefsloot H C J, van der Greef J, et al. *Analytica Chimica Acta*, 2005, 530: 173.
- [13] Liu T, Li Z, Yu C, et al. *Infrared Physics & Technology*, 2017, 87: 124.
- [14] Cybenko G. *Mathematics of Control, Signals and Systems*, 1989, 2: 303.
- [15] Goodfellow I, Bengio Y, Courville A. *Deep Learning*, MIT Press, 2016.
- [16] Zhang J, Cui X, Cai W, et al. *Journal of Chemometrics*, 2018, 32: e2971.
- [17] Kennard R W, Stone L A. *Technometrics*, 1969, 11(1): 137.
- [18] Chen X J, Xu Y L, Meng L W, et al. *Sensors and Actuators B: Chemical*, 2020, 311: 127924.

Near Infrared Spectroscopy Transfer Based on Deep Autoencoder

LIU Zhen-wen¹, XU Ling-jie^{2*}, CHEN Xiao-jing³

1. College of Harbour and Environmental Engineering, Jimei University, Xiamen 361021, China

2. College of Mechanical and Electrical Engineering, Wenzhou University, Wenzhou 325035, China

3. College of Electrical and Electronic Engineering, Wenzhou University, Wenzhou 325035, China

Abstract The establishment of a near-infrared spectroscopy multivariate calibration model relies on calibration samples. However, changes in the near-infrared spectroscopy measurement environment can cause differences between the spectra easily. In order to reduce the consumption of rebuilding calibration model on offset spectrum, this paper proposes a nonlinear spectral transfer method based on deep autoencoder (DAE), which can realize the spectral transfer in an end-to-end way. This method compensates for the poor performance of existing linear spectral transfer methods in the face of nonlinear offset spectra. Moreover, this method can realize the transfer between raw spectra without data process and feature extraction operations. In the paper, we propose an error function penalty term based on the conditional probability distribution and parameter maximum likelihood method, and combine it to the gradient back-propagation algorithm to optimize the parameters of DAE. In order to verify the effectiveness of the method, we perform the proposed method on tablet dataset and corn dataset, which are both public near-infrared spectral datasets. First, we divide the two datasets into the calibration set, validation set, and prediction set using Kennard-Stone (KS) method respectively. Then, we design a network structure that conforms to the spectral sample dimension of the dataset. Finally, samples in calibration set are input to the DAE, and the network parameters are iteratively optimized by the proposed error function and back-propagation algorithm. After the transfer model is established, we compare it with the spatial transformation (SST) and piecewise direct standardization (PDS), both of them are classical linear transfer algorithm. The transferred spectrum obtained by these three algorithms are respectively inputted into the established multivariate calibration model, and we can find that the root means square error of prediction (RMSEP) of the proposed method averagely improves 5.7% and 10.1% than SST and PDS respectively, which can be demonstrated that the spectral samples transferred by the nonlinear deep autoencoder are highly efficient and useful.

Keywords Near infrared spectroscopy; Model transfer; Multivariate calibration; Deep auto-encoder

(Received Oct. 8, 2019; accepted Jan. 30, 2020)

* Corresponding author