

# 近红外光谱模型转移新算法\*

张 进<sup>1</sup> 蔡文生<sup>1</sup> 邵学广<sup>1,2,3,4,5\*\*</sup>

(1. 南开大学化学学院 分析科学研究中心 天津 300071; 2. 天津市生物传感与分子识别重点实验室 天津 300071; 3. 药物化学生物学国家重点实验室 天津 300071; 4. 天津化学化工协同创新中心 天津 300071; 5. 喀什大学化学与环境科学学院 喀什 844000)

**摘 要** 模型转移是解决分析仪器或分析方法通用性的关键技术。近红外光谱受测量仪器或测量条件的影响较大,模型转移对近红外光谱技术的实际应用尤为重要。本文综述了近年来近红外光谱分析中被广泛应用和新提出的模型转移算法,从计算原理角度梳理了有标样和无标样算法的联系和区别。有标样算法重点介绍了基于多元校正、因子分析、人工神经网络、多任务学习的模型转移方法,无标样算法重点介绍了基于光谱校正、模型参数校正和稳健建模的模型转移方法。从算法的角度分析了各种模型转移方法的特点和转移效果,并展望了模型转移算法的进一步发展。在综述的众多方法中分段直接标准化及其变体仍是模型转移的黄金标准,但是,基于因子分析的算法正变得受欢迎且基于神经网络和多任务学习的方法近年来也吸引了越来越多的注意。但是,在实际应用中,获得标准样品以在主机和子机上测得其光谱比较困难甚至是不可能的,无标样模型转移则更加实用。此外,随着仪器小型化、成像及超光谱成像的发展,模型转移在未来会变得愈加必不可少。

**关键词** 近红外光谱 化学计量学 模型转移 多任务学习 光谱空间转换

中图分类号:O657.33;O651 文献标识码:A 文章编号:1005-281X(2017)08-0902-09

## New Algorithms for Calibration Transfer in Near Infrared Spectroscopy\*

Jin Zhang<sup>1</sup>, Wensheng Cai<sup>1</sup>, Xueguang Shao<sup>1,2,3,4,5\*\*</sup>

(1. Research Center for Analytical Sciences, College of Chemistry, Nankai University, Tianjin 300071, China; 2. Tianjin Key Laboratory of Biosensing and Molecular Recognition, Tianjin 300071, China; 3. State Key Laboratory of Medicinal Chemical Biology, Tianjin 300071, China; 4. Collaborative Innovation Center of Chemical Science and Engineering (Tianjin), Tianjin 300071, China; 5. College of Chemistry and Environmental Science, Kashgar University, Kashgar 844000, China)

**Abstract** Calibration transfer is a key technique to ensure the consistency of instruments or analytical methods. Near infrared spectra is strongly influenced by the status of instrument or the environment of measurement. Therefore, calibration transfer is essential for practical applications of near infrared spectroscopy. This paper provides an overview of the state-of-the-art pertaining to calibration transfer methods, including those with and without standard samples, and the focus is on the methods based on multivariate calibration, factor analysis, neural network and multi-task learning for algorithms with standard samples and on the methods based on spectral correction, model coefficient correction and robust multivariate calibration for algorithms without standard

收稿:2017年4月13日,收修改稿:2017年6月28日,网络出版:2017年8月21日

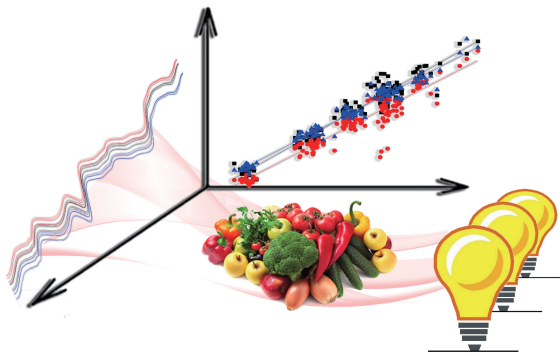
\* 国家自然科学基金项目(No. 21475068)资助

The work was supported by the National Natural Science Foundation of China (No. 21475068).

\*\* Corresponding authors e-mail: xshao@nankai.edu.cn

samples. The formulation and efficiency of different calibration transfer methods are analyzed from the viewpoint of the algorithms. Among the summarized methods, piecewise direct standardization (PDS) and its variants are still the gold standard for calibration transfer, but algorithms based on factor analysis is becoming popular and those based on neural network and multi-task learning has gradually attracted more attention in recent years. In practical applications, however, it is difficult or even impossible to obtain the standard samples for measuring their spectra on both master and slave instruments, calibration transfer without standard samples is more practical. Furthermore, with the development of the instrumentation in miniaturization, imaging and hyper-spectral imaging, calibration transfer will be more and more essential in the future.

**Key words** near infrared spectroscopy; chemometrics; calibration transfer; multi-task learning; spectral space transformation



Recent advances on calibration transfer with and without standard samples are summarized, including the algorithms based on factor analysis, neural network and multi-task learning.

Contents

- 1 Introduction
- 2 Algorithms with standard samples
  - 2.1 Multivariate calibration transfer
  - 2.2 Factor analysis methods
  - 2.3 Neural network
  - 2.4 Multi-task learning
  - 2.5 Others
- 3 Algorithms without standard samples
  - 3.1 Spectral correction methods
  - 3.2 Model coefficient correction methods
  - 3.3 Robust multivariate calibration
- 4 Conclusion

1 引言

近红外(near infrared, NIR)光谱是一种简单快速、绿色环保的新型分析技术,在许多领域已经得到广泛应用<sup>[1~3]</sup>。随着“互联网+”的推进,近红外光谱的网络化布局也得到迅速发展,近红外光谱的仪器类型更加多样,光谱的测量更加快捷,

在指导生产实践、制定统筹规划等方面逐渐发挥起重要作用<sup>[4, 5]</sup>。

随着近红外光谱应用场景的复杂化,经常会遇到检测条件或仪器自身发生变化的情况,如样品检查温/湿度变化<sup>[6~8]</sup>、样品形态改变<sup>[9]</sup>、仪器老化<sup>[10]</sup>和附件更换<sup>[8]</sup>等,此时采集的光谱往往会发生吸光度差异(difference in absorbance)和波长漂移(wavelength shifts)<sup>[11]</sup>,进而造成预测结果的偏差。解决这类问题的方法称为模型转移(calibration transfer)、模型传递或仪器/光谱标准化等<sup>[12~14]</sup>。通常将已经建立模型的仪器称为主机(master instrument),其他待转移的仪器称为子机(slave instrument)。模型转移所针对的情况复杂多样,本文仅以多台仪器间的模型转移为例对各类算法进行讨论。

模型转移主要通过主机和子机检测信号之间的联系,拟合出对应的转移函数,以保证多台仪器预测结果的一致性和准确性<sup>[15]</sup>。用来建立转移函数的光谱称为标准光谱,采集标准光谱所用的样品为标准样品,目前普遍使用的标准样品为主机校正集覆盖范围内的样品。标准样品的选择、制备、运输、存储和采集均需按照一定的规范进行,例如标准样品的数量、代表性、灭活、密封、运输和储存条件等。

标准样品的光谱是建立模型转移校正函数的重要依据。按照是否需要在所有仪器上采集一一对应的标准光谱,模型转移可以划分为有标样和无标样模型转移<sup>[15]</sup>。大部分模型转移算法为有标样算法,如斜率截距法(slope/bias, S/B)<sup>[16]</sup>、Shenks 算法(Shenks algorithm)<sup>[17]</sup>、直接标准化(direct standardization, DS)<sup>[18]</sup>和分段直接标准化(piecewise direct standardization, PDS)<sup>[18]</sup>等算法。在实际应用中,当无法获取标准样品或难以在所有仪器上采集标准光谱时,无标样模型转移算法就起

到关键性作用。常见的方法包括信号处理方法,如小波变换(wavelet transform, WT)<sup>[19]</sup>、有限脉冲响应(finite impulse response, FIR)方法<sup>[20]</sup>和正交信号校正(orthogonal signal correction, OSC)<sup>[21]</sup>算法等。

按照函数校正的对象,也可以将模型转移算法分为三类<sup>[11, 22]</sup>,一是对预测结果校正的算法,如S/B等;二是对光谱校正的算法,如Shenks、DS、PDS、FIR、OSC等;三是对模型参数校正的算法,如两步偏最小二乘(two step partial least squares)方法<sup>[23]</sup>等。

已有文献对模型转移的方法进行了总结。Fearn<sup>[11]</sup>总结了2001年以前的模型转移算法,并指出制约近红外光谱分析广泛应用的最大障碍可能是仪器间的一致性,而模型转移从实验角度为提高仪器一致性提供了一种方案。同年,褚小立等<sup>[15]</sup>介绍了模型转移常用算法的原理和实际应用,并指出各类算法的局限性和发展前景。2002年,Brown等<sup>[22]</sup>从光谱的差异类型、产生原因,到仪器选择、模型维护和标准样品规范等方面介绍了与模型转移相关的工作。2009年,胡昌勤等<sup>[24]</sup>总结了各种模型转移方法及其适用范围。本文总结了2000年以后提出或广泛应用的模型转移算法,从算法角度将各类算法进行分类梳理,并展望了红外光谱模型转移方法的进一步发展方向。

## 2 有标样模型转移算法

有标样模型转移算法是指在计算过程中,通过标准样品在主机和子机上采集的对应光谱建立光谱、预测值或模型参数之间的函数关系,进而实现预测结果校正。大部分利用光谱校正的模型转移方法可以表示成子机光谱 $\mathbf{X}_s$ 和转换矩阵 $\mathbf{F}$ 相乘得到主机光谱 $\mathbf{X}_m$ 的形式<sup>[18]</sup>。

$$\mathbf{X}_m = \mathbf{X}_s \mathbf{F} \quad (1)$$

此类方法具有结构统一、计算效率高的特点,比较适用于光谱线性差异的校正。

模型转移中大部分方法为有标样方法。从算法角度看,有标样方法主要包括基于多元校正、因子分析、神经网络和多任务学习的模型转移算法。

### 2.1 基于多元校正的模型转移

基于多元校正的模型转移是指通过多元回归、求逆等运算对近红外光谱进行校正的一类方法。该类方法一直是模型转移研究的热点之一,最典型的是DS和PDS算法<sup>[18]</sup>。在随后的二十多年中又发展了利用光谱预处理、变量选择、小波变换等方法改

进的DS和PDS方法<sup>[25]</sup>。

DS是一种通过主机和子机标准光谱( $\mathbf{X}_m$ ,  $\mathbf{X}_s$ ),利用最小二乘原理直接求解转换矩阵 $\mathbf{F}$ 的方法。

$$\mathbf{F} = (\mathbf{X}_s^T \mathbf{X}_s)^{-1} \mathbf{X}_s^T \mathbf{X}_m \quad (2)$$

然而由于光谱性质随波长的区域性变化和不等性噪声等原因,实际应用中往往难以达到理想的效果。为了减小光谱差异区域性变化带来的误差,PDS利用窗口内的局部光谱进行模型转移。该方法在一定程度上提升了模型转移效果,但是由于在不同波长处的窗口宽度无法自适应调整,该算法仍有改进的空间。

小波变换有效解决了近红外光谱多尺度分析问题<sup>[26]</sup>,为多尺度自适应模型转移提供了一种思路。2001年Brown等<sup>[25]</sup>提出一种通过WT结合DS和PDS的模型转移方法——小波变换联合直接校正(wavelet hybrid direct standardization, WHDS)。该方法利用WT对光谱信号进行降噪和压缩,再通过小波逆变换重构出不同尺度的光谱信号,然后对不同尺度的光谱信号分别进行模型转移。由于不同尺度的光谱信号具有特定的频率特征,改进的方法针对不同尺度的光谱采用独立的模型转移参数,提高了DS或PDS对区域性变化的稳健性。Han等<sup>[27]</sup>和田高友等<sup>[28]</sup>分别提出了基于小波系数进行模型转移的方法。该方法先利用WT对主机和子机的标准光谱提取小波系数,再通过DS或PDS对不同尺度的小波系数进行转换,最后直接利用校正的小波系数进行建模分析。与传统DS和PDS相比,该方法省掉了小波变换逆运算,同时对光谱进行压缩,其模型转移计算效率和效果均有提升。

王菊香等<sup>[29]</sup>提出了一种小波多尺度分段直接校正(wavelet multi-scale piecewise direct standardization, WM-PDS)的方法,该方法针对不同尺度的小波系数应用不同窗口宽度的PDS。与小波系数直接进行模型转移相比,该方法虽然在一定程度上增加了计算复杂度,但是转移后的光谱预测准确度明显提升。Wang等<sup>[30]</sup>提出了一种遗传算法直接校正(genetic algorithm-direct standardization, GA-DS)的模型转移方法。该方法首先通过GA选择特征变量,然后用DS进行模型转移。变量选择一定程度上降低了光谱的维数,减小了模型转移中随机误差和共线性引起的过拟合。Yuan等<sup>[31]</sup>提出了一种改进的PDS(Rank-Kennard-Stone-PDS, Rank-KS-PDS)算法,该方法通过化学值辅助KS算法挑选标



准样品,再通过 PDS 进行模型转移。化学值辅助的选择方法能够选择在目标值空间中尽可能均匀分布的标准样品。在同样的模型转移效果下,该方法可以减少使用标准样品的数量。

DS 和 PDS 作为一类基于多元校正的模型转移方法,已经经过了广泛的应用和检验,成为一种衡量模型转移效果的标准方法。结合光谱预处理、变量选择和标准样品选择等方法,PDS 能够提供更加稳健和准确的模型转移结果。

2.2 基于因子分析的模型转移

因子分析是一种通过对数据矩阵进行特征分析、旋转变换等操作,以获得有关信息的数学方法。

$$\mathbf{X}_{ij} = \sum_{n=1}^N \mathbf{T}_{in} \mathbf{P}_{jn}, (i = 1, 2, \dots, I; j = 1, 2, \dots, J) \tag{3}$$

其中  $\mathbf{X}_{ij}$  为第  $i$  条光谱的第  $j$  个元素,  $\mathbf{T}_{in}$  为第  $n$  个因子的得分向量的第  $i$  个元素,  $\mathbf{P}_{jn}$  为第  $n$  个因子载荷向量的第  $j$  个元素。基于因子分析的模型转移方法主要利用因子分析方法提取光谱特征信息,并利用光谱特征构造具有低秩特征的投影矩阵,从而达到模型转移的目的。此类方法能够降低数据维数、提高计算效率、减小噪声干扰和过拟合现象,经过数十年的发展已经成为一类比较成熟的模型转移方法。

2005 年,王艳斌等<sup>[32]</sup>提出了一种基于目标因子分析的模型转移方法 (calibration transfer based on target factor analysis, TTFA),为基于因子分析的模型转移提供了一种新思路,TTFA 算法包括四个基本步骤:

- (1) 将主机标准光谱  $\mathbf{X}_m$  进行主成分分析 (principal component analysis, PCA),得到得分  $\mathbf{T}_m$  和载荷  $\mathbf{P}_m$  矩阵。
- (2) 同样将子机标准光谱  $\mathbf{X}_s$  进行 PCA,得到得分  $\mathbf{T}_s$  和载荷  $\mathbf{P}_s$  矩阵。
- (3) 通过广义逆运算得到子机得分矩阵  $\mathbf{T}_s$  向主机得分矩阵  $\mathbf{T}_m$  转换的最小二乘解  $\mathbf{T}$ 。
- (4) 则任意子机光谱  $\mathbf{X}_{s, \text{test}}$  可以转换为  $\mathbf{X}_{m, \text{test}}$ ,

$$\mathbf{X}_{m, \text{test}} = \mathbf{X}_{s, \text{test}} \mathbf{P}_s \mathbf{T} \mathbf{P}_m^T \tag{4}$$

分离出转换矩阵  $\mathbf{F}$ ,如公式(5)。

$$\mathbf{F} = \mathbf{P}_s (\mathbf{T}_s^T)^+ \mathbf{T}_m \mathbf{P}_m \tag{5}$$

其中“+”代表广义逆。利用 TTFA 和 PDS 分别对模拟光谱和真实光谱进行模型转移,结果表明 TTFA 与 PDS 均能得到模型转移效果,而 TTFA 在光谱的非线性差异校正上具有一定优势。

光谱空间转换 (spectral space transformation, SST)<sup>[33]</sup>是一种利用标准光谱构造差异光谱投影空间的模型转移的算法。该算法的特点在于将不同仪器采集的标准光谱连接为增广矩阵,对增广矩阵  $\mathbf{X}_{\text{comb}}$  进行 PCA (忽略残差矩阵),如公式(6),得到不同仪器光谱的载荷 (loading),利用载荷构造转换矩阵  $\mathbf{F}$ ,并进行模型转移。

$$\mathbf{X}_{\text{comb}} = [\mathbf{X}_m, \mathbf{X}_s] = \mathbf{T} \mathbf{P} = \mathbf{T} [\mathbf{P}_m^T, \mathbf{P}_s^T] \tag{6}$$

其中  $\mathbf{T}$  和  $\mathbf{P}$  分别为增广矩阵  $\mathbf{X}_{\text{comb}}$  的 PCA 的得分和载荷,  $\mathbf{P}_m$  和  $\mathbf{P}_s$  分别为主机和子机的载荷。子机待测光谱  $\mathbf{X}_{s, \text{test}}$  可以通过公式(7)转换为  $\mathbf{X}_{m, \text{test}}$ 。

$$\mathbf{X}_{m, \text{test}} = \mathbf{X}_{s, \text{test}} + \mathbf{X}_{s, \text{test}} (\mathbf{P}_s^T)^+ (\mathbf{P}_m^T - \mathbf{P}_s^T) \tag{7}$$

为了结构统一,可将上式分离出转换矩阵  $\mathbf{F}$ ,如公式(8)。

$$\mathbf{F} = \mathbf{I} + (\mathbf{P}_s^T)^+ (\mathbf{P}_m^T - \mathbf{P}_s^T) \tag{8}$$

其中  $\mathbf{I}$  为单位矩阵。SST 具有结构简单、对标准样品个数不敏感等优点,是一种比较理想的基于因子分析的模型转移参考方法。

基于典型相关性分析 (canonical correlation analysis, CCA)<sup>[34]</sup>的模型转移是一种经典的基于因子分析的模型转移方案。该算法最大的特点在于提取了标准光谱间具有最大相关的“典型变量”,通过典型变量间的转换达到模型转移的目的。然而 CCA 仅考虑提取的典型变量的最大相关,可能引入与目标无关的冗余信息,从而使模型转移函数变得复杂。因此, Du 等<sup>[35]</sup>提出了一种利用偏最小二乘 (partial least squares, PLS) 提取与目标值相关且方差最大的因子,再利用 CCA 进行模型转移的方法。该方法一定程度上避免了 CCA 将某些与目标无关的特征作为典型成分的可能,因而能够对仅与目标成分直接相关的光谱差异进行校正,提高了模型转移的针对性。

光谱回归 (spectral regression, SR)<sup>[36, 37]</sup>是一种基于流形学习 (manifold learning, ML) 的数据降维框架。与 PCA、PLS 等方法相比,该方法避免了频繁做稠密矩阵的特征值分解过程,大大提高了因子分析的计算效率。Peng 等<sup>[38]</sup>提出了一种基于 SR 的模型转移方法对玉米和药片的近红外光谱进行了模型转移,结果表明 SR 的转移效果明显优于 PDS。

交替三线性分解 (alternating tri-linear decomposition, ATLD)<sup>[39]</sup>是一种基于交替最小二乘的迭代分解算法,该方法对三维数据的分解具有较清晰的“物理意义”和明显的“二阶优势”<sup>[40]</sup>。本课题组提出利用 ATLD 对近红外光谱进行模型转移的

方法<sup>[12]</sup>,通过对三台(或三台以上)光谱仪器的标准光谱构成的三维矩阵进行分解,得到相对样品浓度 **A**、光谱 **B** 和仪器 **C** 解矩阵。通过矩阵 **C** 建立不同仪器的相对关系进行模型转移,其具体步骤如下:

(1)通过 ATLD 对三台(或三台以上)仪器采集的标准光谱进行分解,得到沿不同方向的解 **A**( $I * N$ )、**B**( $J * N$ )和 **C**( $K * N$ )。

(2)根据 ATLD 可知:

$$\mathbf{X}_{:,1} - \mathbf{X}_{:,2} = \mathbf{X}_{:,1} \mathbf{F}_{k1}^+ \mathbf{F}_{k1} - \mathbf{X}_{:,1} \mathbf{F}_{k1}^+ \mathbf{F}_{k2} \quad (9)$$

其中  $\mathbf{F}_{ki}$  为  $\text{diag}(\mathbf{c}_i) \mathbf{B}^T$ ,  $\mathbf{c}_i$  为矩阵 **C** 的第  $i$  行。

(3)从  $k1$  仪器采集的光谱可通过公式(10)转移到  $k2$  仪器的光谱:

$$\mathbf{X}_{:,2} = \mathbf{X}_{:,1} + \mathbf{X}_{:,1} \mathbf{F}_{k1}^+ \mathbf{F}_{k2} - \mathbf{X}_{:,1} \mathbf{F}_{k1}^+ \mathbf{F}_{k1} \quad (10)$$

其中转换矩阵 **F** 可通过公式(11)求得:

$$\mathbf{F} = \mathbf{I} + \mathbf{F}_{k1}^+ (\mathbf{F}_{k2} - \mathbf{F}_{k1}) \quad (11)$$

由公式(11)和公式(8)可以看出 ATLD 的转换矩阵 **F** 与 SST 形式一致,其实质则都是基于特征向量的转换,然而 ATLD 在特征向量分解上则更具“物理意义”,具有明显的“二阶优势”。通过对玉米和烟叶的近红外光谱进行模型转移,分别采用 PDS、SST 和 ATLD 进行模型转移,结果表明 SST 和 ATLD 的效果优于 PDS。

### 2.3 基于人工神经网络的模型转移

人工神经网络(artificial neural network, ANN)是一种仿生物神经网络的结构和功能的计算模型,用于对数据之间的函数关系进行估计和近似。20 世纪末, Walczak 等<sup>[41]</sup>和 Duponchel 等<sup>[42]</sup>已经提出了基于 ANN 进行模型转移的思路。然而由于 ANN 模型训练效率低,容易对局部光谱过拟合,因此转移效果没有显著提升,并未引起足够重视。

极限学习机(extreme learning machine, ELM)是一种单隐层的 ANN,与传统前馈型 ANN 不同的是 ELM 不需要训练输入层到隐层的权重,能够极大地提高神经网络的训练速度<sup>[43]</sup>。2016 年, Liang 等<sup>[44]</sup>提出了一种基于极限学习机自编码(transfer via extreme learning machine auto-encoder, TEAM)的近红外光谱模型转移方法,其基本思路是对子机标准光谱和主机标准光谱建立多个 ELM 的组合模型,进而通过组合模型进行模型转移。分别对玉米、烟叶和药品的近红外光谱进行模型转移,结果表明经过 TEAM 转移后的光谱预测结果明显优于 PDS。由于极限学习机训练速度快,因此该方法比较适用于多模型组合,从而避免了单个 ANN 模型的局部过拟合现象。随着 ANN 在深度学习和人工智能领域的蓬

勃发展,利用各种 ANN 模型进行模型转移成为一类比较实用的方法。

### 2.4 基于多任务学习的模型转移

多任务学习(multi-task learning, MTL)是一类基于归纳迁移学习的机器学习算法,已经被广泛应用于邮件过滤、文本分类、WIFI 定位等<sup>[45]</sup>。公式(12)为一种 MTL 的基本框架<sup>[46]</sup>,MTL 的学习过程通常包括多个任务的共同学习( $\Sigma(\cdot)$ )和任务间联系的惩罚( $\Omega(\cdot)$ )。

$$\min_{\mathbf{w}_i} \left( \sum_{i=1}^I \left( \|\mathbf{y}_i - \mathbf{X}_i^T \mathbf{w}_i\|^2 + \lambda_i \mathbf{w}_i \mathbf{C}^{-1} \mathbf{w}_i^T \right) + \Omega(\mathbf{C}) \right) \quad (12)$$

其中  $\mathbf{y}_i$  和  $\mathbf{X}_i$  分别为第  $i$  个任务的目标值和校正样本,  $\mathbf{w}_i$  为对应的模型参数,  $\lambda$  为权重调整参数, **C** 为任务间的联系,  $\Omega(\mathbf{C})$  为任务间联系的惩罚项。

近两年,已经有多篇文献报道通过 MTL 对分析化学仪器差异进行校正。例如电子鼻响应的模型转移<sup>[47]</sup>,该问题与近红外光谱模型转移具有极大的相似性,甚至可以互相借鉴。2015 年, Yu 等<sup>[48]</sup>将支持向量机(support vector machine, SVM)和二范数惩罚( $\Omega$  为  $L_2$ )的多任务学习框架结合,提出了一种近红外光谱的模型转移方法(regularized multi-task learning, RMTL)。分别对玉米、药片和醇溶液的近红外光谱进行模型转移, RMTL 和 PDS 均表现出良好的模型转移效果。2016 年, Yu 等<sup>[13]</sup>在多任务学习框架的基础上提出了一种基于迹惩罚( $\Omega$  为  $\text{tr}(\cdot)$ )的近红外光谱模型转移算法。该方法能够有效防止模型转移过程中的过拟合现象,在引入非线性核函数后能足够校正大部分非线性差异,从而带来更稳健的模型转移效果。

### 2.5 其他方法

还有部分算法基于单变量校正、峰位置对齐和光谱拓展等方法在特定条件下也能达到模型转移的效果。2014 年, 樊永显等<sup>[49]</sup>提出一种基于光谱缩放平移的一元线性回归(simple linear regression direct standardization, SLRDS)模型转移方法,该方法适用于光谱具有较小的线性差异的情况。本课题组提出了一种基于拓展光谱(extended spectra)的模型转移方法<sup>[50]</sup>,该方法利用一组标准样品估计子机和主机之间的差异光谱,利用差异光谱将主机校正集转移为子机校正集,从而扩充子机建模光谱数量,建立子机稳健的模型。2016 年,本课题组提出了一种基于双模型策略(dual model strategy)的模型转移方法<sup>[51]</sup>,该方法通过建立两个模型(主模型和校正

模型)达到模型转移的目的,主模型为主机光谱建立的模型,校正模型实质是描述不同仪器预测值的比值,则子机采集的任意光谱可以通过主模型预测结果与校正模型预测的比值得到准确的预测值。彭黔荣等<sup>[52]</sup>提出一种改进的 Shenks 方法,该方法通过特征峰位置匹配代替 Shenks 中的波数校正,使光谱波数校正过程更加稳健。

### 3 无标样模型转移算法

在实际 NIR 模型转移的应用中,往往因为难以得到稳定的标准样品或标准样品无法同时所有仪器上采集光谱,而导致了有标样模型转移难以实施<sup>[20]</sup>。无标样模型转移是可以在一定程度上解决这种问题,即在建立校正函数的过程中无需使用一一对应的标准光谱。

按照校正对象分类,无标样模型转移可以分为无标样光谱校正、无标样模型参数校正和建立稳健模型。

#### 3.1 无标样光谱校正

无标样光谱校正是指利用基线校正、光谱求导、数字滤波、信号平滑等方法减小仪器间光谱差异的方法。最早的无标样模型转移方法为基于 FIR<sup>[20]</sup>的光谱校正。

Morris 等<sup>[53]</sup>提出了通过优化窗口宽度结合 FIR 和窗口多元散射校正(windowed multiplicative scatter correction, W-MSC)消除仪器间光谱差异的方法。对飞机燃料的近红外光谱进行模型转移,结果表明窗口大小影响模型转移效果。窗口过小时容易导致光谱被过分滤波,样品自身携带的信息丢失;窗口过大时光谱差异消除效果不明显。选择合适大小的窗口后用 MSC 和 FIR 能够有效消除仪器间光谱差异,校正后的子机光谱预测误差能够满足分析需求。

利用小波变换对近红外光谱进行背景扣除能够有效消除变化背景带来的预测误差<sup>[19]</sup>。本课题组提出了一种基于小波变换的数字滤波器,该方法通过遗传算法优化小波函数的组合,建立了一种直接定量分析的方法<sup>[54]</sup>。该方法建立的模型能够对三组具有差异的玉米近红外光谱的蛋白质含量同时进行准确的预测,表明该方法能够有效过滤变化的背景对光谱分析带来的干扰,可以达到理想的模型转移效果。

另外,外部参数正交化(external parameter orthogonalisation, EPO)<sup>[55,56]</sup>、正交空间回归(orthogonal space regression, OSR)<sup>[7,57]</sup>等方法也是

基于滤波达到无标样模型转移的目的。

#### 3.2 无标样模型参数校正

无标样模型参数校正是指以纯物质光谱、无信息光谱、主机模型参数等已知条件为基础,利用空间转换或优化算法求出子机模型的过程。近年来,无标样的模型参数校正逐渐引起人们的重视。Kalivas 和 Kunz 等<sup>[58-60]</sup>基于吉洪诺夫正则化(Tikhonov regularization, TR)提出了针对光谱随时间连续变化和离散变化的模型维护方案。该方案在预测残差最小化的基础上加入了包含纯光谱、无关光谱等的正则化约束,从而达到提高模型适用性、降低预测误差的目的。

$$\min_{\mathbf{w}_m} \left( \|\mathbf{y} - \mathbf{X}\mathbf{w}_m\|^a + \lambda \|\mathbf{L}\mathbf{w}_s\|^b + \eta\Omega(\mathbf{w}_s) + \dots \right) \quad (13)$$

其中, $\lambda$  和  $\eta$  为权重调整参数, $\mathbf{w}_m$  和  $\mathbf{w}_s$  分别为主机和子机模型参数;第一项为预测残差最小化,第二项为对子机模型的约束,第三项为包括纯光谱、无关光谱等的约束; $a$  和  $b$  为一种惩罚长度的概念(范数),当其为 0 和 1 时,所求的解具有稀疏特性,而为 2 时,所求的解相对平滑。该方案是一类对主机光谱、无信息光谱、纯组分光谱信息的加权融合方法,在模型转移、模型维护甚至无模型预测等方面都有较为深入的研究。

具有稀疏约束的 TR 模型转移方法能够减少参与计算的变量数,降低模型复杂度和过拟合风险。然而当光谱分辨率过高时,该方法选择的变量过于稀疏,会导致不稳健和解释性差的问题。为了解决连续性差的问题,She 等<sup>[61]</sup>提出了一种增加连续性约束的吉洪诺夫模型维护方法。结果表明加入了连续化约束的吉洪诺夫正则化模型维护方案保证模型的连续性的同时对不同仪器采集的光谱均能够准确预测。因此充分利用无信息光谱、纯组分光谱能够改善模型稳健性和实用性,为无标样模型转移提供了一种新思路。

本课题组提出了一种双模型的无标样模型转移方法<sup>[62]</sup>,该方法假设光谱差异与目标值差异呈线性相关,通过建立与光谱差异相关的预测结果校正模型达到模型转移的目的。通过该方法、SST 和 PDS 对玉米和烟叶的近红外光谱进行模型转移,结果表明该方法能够达到与 PDS 和 SST 相似的转移结果,但该方法无需标准样品,在实际应用中更加实用。近期,本课题组提出了另一种基于模型线性相关的线性模型校正(linear model correction, LMC)无标样模型转移方法<sup>[63]</sup>,该方法在双模型策略的<sup>[49, 51]</sup>基



基础上提出了模型参数线性相关的推论,通过对主机和子机模型参数的相关系数  $\text{corr}(\mathbf{w}_s, \mathbf{w}_m)$  进行约束,并对子机模型预测残差最小化,从而求得转移后的子机模型。

$$\begin{aligned} \min_{\mathbf{w}_s} & \left( \|\mathbf{y}_s - \mathbf{X}_s \mathbf{w}_s\|^2 \right) \\ \text{s. t. } & \text{corr}(\mathbf{w}_s, \mathbf{w}_m) \geq \lambda \end{aligned} \quad (14)$$

由于向量的相关系数仅与两个向量的夹角大小相关,而与其长度无关。因此,  $\lambda$  可以统一设定为一个固定值,使得算法在参数选择上更加智能化。

### 3.3 建立稳健模型

建立稳健模型,提高模型适应性又被称为“广义的模型转移”<sup>[24]</sup>。此类方法通常利用光谱预处理、变量选择和全局建模等方法建立起对所有仪器和条件都适用的稳健模型。

通过合理的光谱预处理结合稳健校正、多模型共识识别等方法也可以达到稳健建模、减小仪器间预测误差的目的。例如多元散射校正(multiplicative scatter correction, MSC)<sup>[64]</sup>、正交信号校正(orthogonal signal correction, OSC)<sup>[65]</sup>和广义最小二乘(general least squares, GLS)<sup>[66]</sup>等。基于光谱预处理的方法操作简单,容易实现,在光谱差异不大时效果比较明显。然而当光谱差异显著,尤其是光谱峰位置漂移时,此类方法较难达到理想的效果。

基于稳健变量选择的稳健模型建立方法比较适用于光谱峰位置漂移的情况。而光谱预处理过程往往与变量选择、滤波、降噪等过程相关,因此为了建立稳健的公共模型,通常要与其他无模型转移技巧共同使用<sup>[67]</sup>。例如稳定自适应重加权采样(stability competitive adaptive reweighted sampling, SCARS)<sup>[68]</sup>、连续投影组合模型算法(successive projection algorithm-subagging, SPA-Subagging)<sup>[69]</sup>等。

全局建模指将所有包含差异因素的光谱都加入校正集中,建立一个统一的模型<sup>[24]</sup>。建立的全局模型能够有效改善不同条件下模型预测结果,但是准确性往往不如单独建立的模型<sup>[22, 70]</sup>。

## 4 结论

随着“互联网+”的到来,近红外光谱的快速分析优势将被进一步放大,近红外光谱分析技术已经进入人们生产、生活的方方面面。近红外数据的采集、存储、管理与分析将会面临巨大的变革。近红外光谱的数据终端将会朝着轻量化、便携式的方向发展

展,其重要的特点在于快捷丰富的通讯功能,能够在复杂和极端环境轻松上传光谱信息并下载分析结果。近红外终端多元化对模型转移算法提出了新要求:稳健、低复杂度和智能化。稳健性是指模型转移算法能够应对可能出现的各种复杂检测环境,保证相对稳定的校正结果。低复杂度是指模型转移参数在保证转移效果的前提下尽可能建立因子数少、参数空间小的转换函数,进而提高模型转移的泛化能力。智能化是指系统能够针对光谱采集环境自动调整校正参数,以优化方法代替传统的人为干预参数确定,结合专家经验的启发式算法代替人为干预转移算法,向模型自动更新、转移和推荐的方向发展。

20 世纪 80 年代,傅里叶变换近红外逐渐成为主导产品,仪器间一致性得到大幅提升,对模型转移算法的需求得到了缓解。然而,随着仪器小型化和网络化需求的提升,仪器间一致性要求重新引起人们的关注。光电二极管阵列检测器(photo-diode array, PDA)/电荷耦合阵列检测器(charge-coupled device array detector, CCD)在小型近红外光谱仪发展中备受青睐。CCD 近红外光谱仪无可移动的光学部件,结构简单、成本低、极易实现小型化,然而其分辨率低,阵列内部检测器一致性难以保持,阵列检测器间的响应也不如傅里叶变换型近红外一致性好。目前对 CCD 类型的近红外模型转移研究还处在实验室研究阶段,并未考虑检测器内部响应变化的不一致<sup>[70~73]</sup>。根据阵列型检测器的近红外的特点,常规模型转移关于光谱线性相关、模型参数线性相关或检测结果线性相关的假设均难以满足或彻底失效。研究局部校正、自适应局部矫正或基于已知信息的启发式校正的模型转移算法将对便携式近红外模型转移非常关键。

近红外光谱模型转移发展至今已经取得了一定的成果,可以预见在便携式近红外蓬勃发展,快速分析从实验室走向千家万户,模型转移算法发展和应用的潜力依然非常巨大。PDS 作为一种经典的模型转移方案,无论是在理论参照和实际应用中都仍然是首选对象。然而 PDS 在稳健性、抗干扰能力、转移后光谱的异常噪声消除等方面依然还有提升空间。光谱在高维空间中的变化可能是非线性的,非线性模型转移方法在进一步提高模型转移准确性方面还有很大发展空间。近几年,卷积神经网络在机器视觉、嗅觉、语音文字处理等领域发展迅速,多隐层和局部感知的特性为模型转移算法研究发展提供了新思路。模型转移算法发展中依然有很多问题需

要注意:标准样品需求量大、转移参数过多、不稳健、计算复杂度高等。无标样模型转移过程简单,避免了标准样品制备、运输、采集严格的条件限制,为模型转移智能化提供了理论依据。

模型转移是对光谱仪器间一致性的保障和补充,为近红外便携化、网络化发展提供了软件基础。但开发模型转移新算法的同时应注重算法的实用性,应从解决实际问题,推进近红外光谱在大数据时代更广泛的应用为出发点。

- 缩写列表
- F**:转换矩阵
  - I**:单位矩阵
  - X**:光谱矩阵
  - y**:目标值向量
  - X<sub>m</sub>**、**X<sub>s</sub>**:主机/子机标准光谱
  - X<sub>s,test</sub>**、**X<sub>m,test</sub>**:主机/子机待测光谱
  - X<sub>comb</sub>**:多组光谱数据沿某一方向构成的增广矩阵
  - X<sub>,:,k</sub>**:三维矩阵 **X<sub>ijk</sub>** 沿 *K* 方向的第 *k* 个切片矩阵
  - T<sub>m</sub>**、**T<sub>s</sub>**:主机/子机因子分析的得分矩阵或潜变量
  - P<sub>m</sub>**、**P<sub>s</sub>**:主机/子机因子分析的载荷矩阵或潜变量轮廓
  - A**、**B**、**C**:三维矩阵分解算法沿不同方向的解(例如:目标浓度、波长、仪器等)
  - λ**、**η**:权重调整参数
  - w**:模型参数
  - w<sub>m</sub>**、**w<sub>s</sub>**:主机/子机模型参数
  - diag(c<sub>i</sub>)**:以向量 **c<sub>i</sub>** 为对角元素的对角矩阵
  - Ω(·)**:对目标施加的惩罚函数

参 考 文 献

[ 1 ] Huang H B, Yu H Y, Xu H R, Ying Y B. J. Food Eng. , 2008, 87(3) : 303.

[ 2 ] Huppert T J, Diamond S G, Franceschini M A, Boas D A. Appl. Opt. , 2009, 48(10) : D280.

[ 3 ] Prieto N, Roche R, Lavín P, Batten G, Andrés S. Meat Science, 2009, 83(2) : 175.

[ 4 ] 刘言(Liu Y), 蔡文生(Cai W S), 邵学广(Shao X G). 科学通报(Chinese Science Bulletin), 2015, 60(08) : 704.

[ 5 ] 刘言(Liu Y), 蔡文生(Cai W S), 邵学广(Shao X G). 科学通报(Chinese Science Bulletin), 2015, 60(08) : 694.

[ 6 ] Wang Y D, Kowalski B R. Anal. Chem. , 1993, 65(9) : 1301.

[ 7 ] Barboza F D, Poppi R J. Anal. Bioanal. Chem. , 2003, 377

(4) : 695.

[ 8 ] Liu X, Huang C J, Han L J. Energy & Fuels, 2015, 29(10) : 6450.

[ 9 ] Pereira L S A, Carneiro M F, Botelho B G, Sena M M. Talanta, 2016, 147: 351.

[10] Leion H, Folestad S, Josefson M, Sparen A. J. Pharm. Biomed. Anal. , 2005, 37(1) : 47.

[11] Fearn T. J. Near Infrared Spectrosc. , 2001, 9(4) : 229.

[12] Liu Y, Cai W S, Shao X G. Anal. Chim. Acta, 2014, 836: 18.

[13] Yu B F, Ji H B, Kang Y. Spectrosc. Lett. , 2016, 49(1) : 23.

[14] Bouveresse E, Hartmann C, Massart D L, Last I R, Prebble K A. Anal. Chem. , 1996, 68(6) : 982.

[15] 褚小立(Chu X L), 袁洪福(Yuan H F), 陆婉珍(Lu W Z). 光谱学与光谱分析(Spectroscopy and Spectral Analysis), 2001, 21(06) : 881.

[16] Osborne B G, Fearn T. Int. J. Food Sci. Technol. , 1983, 18(4) : 453.

[17] Shenk J S, Westerhaus M O. US 4866644, 1989.

[18] Wang Y D, Veltkamp D J, Kowalski B R. Anal. Chem. , 1991, 63(23) : 2750.

[19] Ma C X, Shao X G. J. Chem. Inf. Model. , 2004, 44(3) : 907.

[20] Blank T B, Sum S T, Brown S D, Monfre S L. Anal. Chem. , 1996, 68(17) : 2987.

[21] Sjöblom J, Svensson O, Josefson M, Kullberg H, Wold S. Chemom. Intell. Lab. Syst. , 1998, 44(1/2) : 229.

[22] Feudale R N, Woody N A, Tan H W, Myles A J, Brown S D, Ferré J. Chemom. Intell. Lab. Syst. , 2002, 64(2) : 181.

[23] Forina M, Drava G, Armanino C, Boggia R, Lanteri S, Leardi R, Corti P, Conti P, Giangiacomo R, Galliena C, Bigoni R, Quartari I, Serra C, Ferri D, Leoni O, Lazzeri L. Chemom. Intell. Lab. Syst. , 1995, 27(2) : 189.

[24] 张学博(Zhang X B), 冯艳春(Feng Y C), 胡昌勤(Hu C Q). 药物分析杂志(Chinese Journal of Pharmaceutical Analysis), 2009, 29(08) : 1390.

[25] Tan H W, Brown S D. J. Chemom. , 2001, 15(8) : 647.

[26] 田高友(Tian G Y), 袁洪福(Yuan H F), 刘慧颖(Liu H Y), 陆婉珍(Lu W Z). 光谱学与光谱分析(Spectroscopy and Spectral Analysis), 2003, 23(6) : 1111.

[27] Yoon J G, Lee B W, Han C H. Chemom. Intell. Lab. Syst. , 2002, 64(1) : 1.

[28] 田高友(Tian G Y), 褚小立(Chu X L), 袁洪福(Yuan H F), 陆婉珍(Lu W Z). 分析化学(Chinese Journal of Analytical Chemistry), 2006, 34(07) : 927.

[29] 王菊香(Wang J X), 李华(Li H), 邢志娜(Xing Z N), 郭恒光(Guo H G). 分析化学(Chinese Journal of Analytical Chemistry), 2011, 39(06) : 846.

[30] Wang J X, Qu J, Li H, Han X, Xu G. Pet. Sci. Technol. , 2012, 30(19) : 1975.

[31] Liang C, Yuan H F, Zhao Z, Song C F, Wang J J. Chemom. Intell. Lab. Syst. , 2016, 153:51.



- [32] 王艳斌(Wang Y B), 袁洪福(Yuan H F), 陆婉珍(Lu W Z). 光谱学与光谱分析(Spectroscopy and Spectral Analysis), 2005, 25(03): 398.
- [33] Du W, Chen Z P, Zhong L J, Wang S X, Yu R Q, Nordon A, Littlejohn D, Holden M. Anal. Chim. Acta, 2011, 690(1): 64.
- [34] Fan W, Liang Y Z, Yuan D L, Wang J J. Anal. Chim. Acta, 2008, 623(1): 22.
- [35] Zheng K Y, Zhang X, Iqbal J, Fan W, Wu T, Du Y P, Liang Y Z. J. Chemom., 2014, 28(10): 773.
- [36] Cai D, He X F, Han J W. IEEE 11th International Conference on Computer Vision. IEEE, 2007. 1.
- [37] Yan S, Xu D, Zhang B, Zhang H J, Yang Q, Lin S. IEEE Trans. Pattern Anal. Machine Intelligence, 2007, 29(1): 40.
- [38] Peng J T, Peng S L, Jiang A, Tan J. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2011, 78(4): 1315.
- [39] Wu H L, Shibukawa M, Oguma K. J. Chemom., 1998, 12(1): 1.
- [40] 吴海龙(Wu H L), 俞汝勤(Yu R Q). 化学通报(Chinese Science Bulletin), 2011, 74(9): 771.
- [41] Despagne F, Walczak B, Massart D L. Appl. Spectrosc., 1998, 52(5): 732.
- [42] Duponchel L, Ruckebusch C, Huvenne J, Legrand P. J. Near Infrared Spectrosc., 1999, 7(3): 155.
- [43] Huang G B, Zhu Q Y, Siew C K. Neurocomputing, 2006, 70(1/3): 489.
- [44] Chen W R, Bin J, Lu H M, Zhang Z M, Liang Y Z. Analyst, 2016, 141(6): 1973.
- [45] Pan S J, Yang Q. IEEE Trans. Knowl. Data Eng., 2010, 22(10): 1345.
- [46] Evgeniou T, Pontil M. Regularized Multi-Task Learning. Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Seattle: ACM, 2004. 109.
- [47] Yan K, Zhang D. Sens. Actuators, B: Chem., 2016, 225: 288.
- [48] Yu B F, Ji H B. Anal. Methods, 2015, 7(6): 2714.
- [49] 杨辉华(Yang H H), 张晓凤(Zhang X F), 樊永显(Fan Y X), 谢谱模(Xie P M), 褚小立(Chu X L). 分析化学(Chinese Journal of Analytical Chemistry), 2014, 42(09): 1229.
- [50] 王家俊(Wang J J), 者为(Zhe W), 刘言(Liu Y), 蔡文生(Cai W S), 邵学广(Shao X G). 中国烟草学报(Acta Tabacaria Sinica), 2014, 20(06): 1.
- [51] Wang J J, Li Z F, Wang Y, Liu Y, Cai W S, Shao X G. Spectrosc. Lett., 2016, 48(5): 348.
- [52] 彭黔荣(Peng Q R), 张进(Zhang J), 刘娜(Liu N), 胡芸(Hu Y), 刘剑(Liu J), 姬厚伟(Ji H W), 王芳(Wang F). CN 105628645A, 2016.
- [53] Kramer K E, Morris R E, Rose-Pehrsson S L. Chemom. Intell. Lab. Syst., 2008, 92(1): 33.
- [54] Li X Y, Xu Z H, Cai W S, Shao X G. Anal. Chim. Acta, 2015, 880: 26.
- [55] Amat-Tosello S, Dupuy N, Kister J. Anal. Chim. Acta, 2009, 642(1/2): 6.
- [56] Roger J M, Chauchard F, Bellon M V. Chemom. Intell. Lab. Syst., 2003, 66(2): 191.
- [57] Lin Z Z, Xu B, Li Y, Shi X Y, Qiao Y J. J. Chemom., 2013, 27(11): 406.
- [58] Kalivas J H, Siano G G, Andries E, Goicoechea H C. Appl. Spectrosc., 2009, 63(7): 800.
- [59] Shahbazikhah P, Kalivas J H. Chemom. Intell. Lab. Syst., 2013, 120: 142.
- [60] Kunz M R, Kalivas J H, Andries E. Anal. Chem., 2010, 82(9): 3642.
- [61] Kunz M R, She Y Y. J. Chemom., 2013, 27(9): 233.
- [62] Li X Y, Cai W S, Shao X G. J. Near Infrared Spectrosc., 2015, 23(5): 285.
- [63] Liu Y, Cai W S, Shao X G. Spectrochim Acta A, 2016, 169: 197.
- [64] Geladi P, Macdougall D, Martens H. Appl. Spectrosc., 1985, 39(3): 491.
- [65] Wold S, Antti H, Lindgren F, Öhman J. Chemom. Intell. Lab. Syst., 1998, 44(1/2): 175.
- [66] Martens H, Høy M, Wise B M, Bro R, Brockhoff P B. J. Chemom., 2003, 17(3): 153.
- [67] Malli B, Birlutiu A, Natschlager T. Chemom. Intell. Lab. Syst., 2017, 161: 49.
- [68] 张晓羽(Zhang X Y), 李庆波(Li Q B), 张广军(Zhang G J). 光谱学与光谱分析(Spectroscopy and Spectral Analysis), 2014, 34(05): 1429.
- [69] Martins M N, Galvao R K H, Pimentel M F. J. Braz. Chem. Soc., 2010, 21(1): 127.
- [70] Greensill C V, Wolfs P J, Spiegelman C H, Walsh K B. Appl. Spectrosc., 2001, 55(5): 647.
- [71] Greensill C V, Walsh K. J. Near Infrared Spectrosc., 2002, 10(1): 27.
- [72] Alamar M C, Bobelyn E, Lammertyn J, Nicolaï B M, Moltó E. Postharvest Biol. Technol., 2007, 45(1): 38.
- [73] Hayes C J, Walsh K B, Greensill C V. J. Near Infrared Spectrosc., 2016, 24(1): 59.