

文章编号 1004-924X(2014)02-0296-08

# 近红外光谱稀疏分量分析检测柴油品质参数

周 扬<sup>1,2\*</sup>, 戴曙光<sup>2</sup>, 葛丁飞<sup>1</sup>

(1. 浙江科技学院 信息与电子工程学院, 浙江 杭州 310023;  
2. 上海理工大学 光电信息与计算机工程学院, 上海 200093)

**摘要:** 由于光谱盲源分离中的独立分量分析方法(ICA)在柴油品控参数近红外光谱定量分析时预测效果不理想, 稳定性不高, 本文提出了一种在稀疏特性下的盲源分离近红外光谱分析思路——近红外光谱稀疏分量分析法, 并用该方法预测了柴油沸点、密度、芳烃总量等品控参数。首先利用柴油校正集光谱样本训练冗余字典并完成光谱在该字典下的稀疏变换, 接着完成混合矩阵估计, 最后用混合矩阵与柴油品控参数建立回归预测模型。针对混合矩阵估计中光谱稀疏度不为一时聚状特征模糊导致无法确定聚类数的问题, 提出将 AP 聚类算法应用于聚类过程。实验表明, 近红外光谱稀疏分量分析法对柴油沸点、密度、芳烃总量预测的相关系数( $R$ )、预测均方根误差(RMSEP)分别达到了 98.91%, 99.68%, 99.43% 和 2.84,  $0.88 \times 10^{-3}$ , 0.59, 性能优于 ICA 及全谱偏最小二乘(PLS)等传统方法。该方法可作为一种柴油品控参数检测的有效盲源分离定量分析方法, 并可推广于其它光谱检测领域。

**关键词:** 近红外光谱; 柴油检测; 独立分量分析法; 稀疏分量分析法; 盲源分离

**中图分类号:** O657.33; U473.12 **文献标识码:** A **doi:** 10.3788/OPE.20142202.0296

## Detection of diesel quality parameters by near infrared spectroscopy with sparse component analysis

ZHOU Yang<sup>1,2\*</sup>, DAI Shu-guang<sup>2</sup>, GE Ding-fei<sup>1</sup>

(1. College of Information and Electronic Engineering,

Zhejiang University of Science and Technology, Hangzhou 310023, China;

2. School of Optical-Electrical and Computer Engineering, Shanghai 200093, China)

\* Corresponding author, E-mail: zybuua@163.com

**Abstract:** Independent Component Analysis(ICA) in a blind source separation method can not obtain ideal prediction results and excellent measuring stability in detecting diesel quality parameters by near-infrared spectroscopy. Therefore, a blind source separation method based on sparse characteristics of near-infrared spectroscopy is proposed. The method is named Near Infrared Spectroscopy with Sparse Component Analysis(NIFS SCA) and is used for the prediction of boiling points, density and total aromatics for the diesel. This method firstly trains the redundant dictionary by spectral samples and finishes the sparse transformation for the calibration sample. Then, it estimates the mixing matrix, and establishes the regression model between mixing matrix and diesel quality parameters. As the cluste-

收稿日期: 2013-05-06; 修订日期: 2013-05-30.

基金项目: 国家自然科学基金资助项目(No. 51376162); 浙江省教育厅资助项目(No. Y201327869); 浙江科技学院交叉预研专项(No. 2013JC07Y)

ring progress is hard to determine the number of clusters when the sparsity of spectral is not equal to one and the clustering feature is fuzzy, the AP clustering algorithm is applied to the clustering process. The experiments show that the Relative coefficient( $R$ ) and the Root Mean Square Error of Prediction (RMSEP) of the NIFS SCA for diesel boiling point, density, total aromatic prediction respectively are 98.91%, 99.68%, 99.43% and 2.84,  $0.88 \times 10^{-3}$ , 0.59, which is better than those of ICA and the full spectrum PLS methods. The proposed method can be an effective blind source separation quantitative analysis method for detecting diesel quality parameters and also can be promoted to other spectral detection applications.

**Key words:** near-infrared spectroscopy; diesel quality detection; Independent Component Analysis (ICA); Sparse Component Analysis(SCA); Blind Source Separation(BSS)

## 1 引言

近红外光谱检测法具有无损、快速、低成本、无污染等众多优点,近年来广泛应用在柴油品质参数的快速分析中<sup>[1-3]</sup>。但柴油近红外光谱检测在采集过程中存在噪声,不同组分的光谱吸收峰间相互混叠,不同波长点含有复共线性关系,这些因素制约了近红外光谱法对柴油品质参数的检测精度<sup>[4-5]</sup>。因此,如何克服上述问题的干扰,提高柴油近红外光谱法的检测精度成为亟待解决的问题。

盲源分离技术在图像、声音等领域能完成降噪、分离、特征提取等多种任务,因而迅速成为信号处理领域的研究热点。Scholz、Shao等<sup>[6-7]</sup>将盲源分离思想引入到近红外光谱分析领域,认为近红外光谱由“纯物质”组分的光谱混合而成,样本光谱中不同“纯物质”组分的含量信息可与对应的物理化学参数建立回归模型,并完成新样本参数的预测。盲源分离思想在近红外光谱分析领域主要使用独立分量分析(Independent Component Analysis, ICA)算法。Toivainen综述了ICA法在近红外光谱分析领域的应用情况<sup>[8]</sup>,方利民等<sup>[9]</sup>利用ICA法完成了柴油中多个品控参数的预测。这些报道表明,盲源分离思路及ICA法在光谱回归校正中具有可行性。

ICA法在光谱独立分量提取过程中,要求参与建模的近红外光谱样本大于“纯物质”组分的数量,即要求分离问题是正定或超定的<sup>[10]</sup>。而柴油组分较为复杂,组成柴油的纯物质数量具有不确定性,校正集样本不足将导致分离失败。ICA法

分离过程还易受到噪声的干扰<sup>[11]</sup>,当柴油光谱噪声过大时,ICA法组分的分离效果并不理想。ICA法在分离过程中要求各组分光谱服从统计独立的假设,而实际柴油光谱的不同组分、不同波长变量间存在一定的相关性,该问题制约了ICA法的性能和检测范围。

随着稀疏表示理论的发展,稀疏分量分析方法(Sparse Component Analysis, SCA)因其分离的源信号无需服从统计独立假设且观测信号数(校正集样本数)少于源信号数(“纯物质”组分的数量),取得了比ICA法更好的分离性能<sup>[12]</sup>。为进一步提升柴油近红外法的检测性能和克服ICA法的限制,本课题组首次将SCA法应用于柴油近红外光谱分析。SCA法实施的前提是完成光谱信号的有效稀疏化,稀疏化后的光谱信号应最大程度地保留样本成分特征并降低噪声的干扰。课题组在光谱稀疏表示领域已经做了先期研究,并形成了一套近红外光谱的稀疏化方法<sup>[13]</sup>。

SCA法实现稀疏变换后,近红外光谱的分量提取主要使用基于稀疏特性的两步法<sup>[14]</sup>,该方法先进行混合矩阵估计,再实施源信号估计。由于稀疏变换后的校正集光谱在同一个波长点上并非只有一个分量为零,传统聚类方法如K均值聚类算法、模糊C均值聚类算法并不能完成混合矩阵散点方向的准确分割,同时也无法准确估算聚类数。人为指定聚类数将影响近红外法的预测精度。仿射传播(Affinity Propagation, AP)聚类算法根据近邻传播特性搜索最优的类代表集合,能按照稀疏后光谱散点在最优点上的内在特性自动确定聚类数,即“纯物质”的组分数<sup>[15]</sup>。本文提

出将 AP 聚类算法应用到混合矩阵估计步骤中,以克服光谱稀疏度不足导致的聚类模糊问题,进一步提高光谱盲源分离的稳定性。

## 2 近红外光谱稀疏分量分析方法

### 2.1 冗余字典及光谱的稀疏化

在有限维空间  $\mathbf{R}^n$  中,近红外光谱  $\mathbf{y} \in \mathbf{R}^n$  可由一组单位正交稀疏基  $\mathbf{D} = [d_1 \ d_2 \ \cdots \ d_k]$  线性表示,即  $\mathbf{y} = \mathbf{D}\mathbf{x}$ ,  $\mathbf{d}_k \in \mathbf{R}^k$ 。若表示系数  $\mathbf{x} \in \mathbf{R}^k$  中只有少数几个非零元素,则为稀疏表示。矩阵  $\mathbf{D}$  中的列向量代表稀疏基,称为原子。对于  $n$  维信号,如果  $k = n$ ,即字典中含有  $n$  个线性无关原子,称这个字典张成空间是完备的;如果  $k > n$ ,则认为  $\mathbf{D}$  是冗余的(过完备、超完备),称为冗余字典。

光谱稀疏化的首要前提是为近红外光谱寻找能够匹配光谱信号固有特征的冗余字典。目前,尚无研究明确哪种固定基字典适合近红外光谱的稀疏表示,近年发展的字典训练算法为近红外光谱冗余字典的形成提供了可行的方法。字典训练过程就是利用校正集光谱信号进行迭代运算,最终生成匹配光谱吸收峰特征的冗余字典。典型的字典训练算法有 K 均值奇异值分解(K-SVD)算法、RLS-DLA 等<sup>[16-17]</sup>。在已知光谱信号  $\mathbf{y}$  和冗余字典  $\mathbf{D}$  的情况下,直接求解其稀疏表示系数  $\mathbf{x}$  是病态的,需要采用次优的逼近方法来求解。当前的主流算法是贪婪追踪算法<sup>[18]</sup>,贪婪追踪算法通过迭代过程,每次从字典中选取与光谱信号残差最相关的原子作为局部最优解,如匹配追踪算法(MP)、正交匹配追踪算法(OMP)。

课题组已经对近红外光谱在冗余字典稀疏表示下的回归校正性能做了初步研究,并确定了一套在冗余字典下近红外光谱的稀疏表示方法<sup>[13]</sup>。该方法先用近红外光谱结合 K-SVD 算法完成对冗余字典的训练,再用 OMP 算法寻找光谱在该训练字典下的稀疏表示系数。研究表明,稀疏表示系数包含了光谱样本的物理化学特征信息,并能显著提高近红外光谱的物理化学特征预测性能。该方法能完成光谱的有效稀疏化处理,而稀疏化处理使光谱具有稀疏特性是运用稀疏盲源分离算法的前提。

### 2.2 稀疏盲源分离机理

定义  $\mathbf{X}_{n \times l}$  为稀疏处理后的近红外光谱矩阵,它由  $n$  个训练集光谱样本组成且具有  $l$  个波长点信息。 $\mathbf{S}_{m \times l}$  为光谱  $m$  个独立稀疏源即样本的某种“纯物质”组分光谱。原始光谱可认为是独立稀疏源(“纯物质”组分)的线性叠加。 $\mathbf{A}_{n \times m}$  是不同样本光谱在不同独立稀疏源的混合矩阵。根据上述定义,近红外光谱可以表示为:

$$\mathbf{X} = \mathbf{A}\mathbf{S}, \quad (1)$$

$$\mathbf{X}_i = \mathbf{A}_{i1}\mathbf{S}_1 + \mathbf{A}_{i2}\mathbf{S}_2 + \cdots + \mathbf{A}_{in}\mathbf{S}_n. \quad (2)$$

光谱盲源分离问题就是已知稀疏变换后的光谱  $\mathbf{X}$ ,求解混合矩阵  $\mathbf{A}$  和稀疏信号源  $\mathbf{S}$ 。当光谱具有稀疏特征时,对任意特定的波长点  $t$ ,式(1)、式(2)可以改写成:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} \begin{pmatrix} s_{1t} \\ s_{2t} \\ \vdots \\ s_{mt} \end{pmatrix} = \begin{pmatrix} x_{1t} \\ x_{2t} \\ \vdots \\ x_{nt} \end{pmatrix}. \quad (3)$$

由于独立信号源  $s$  的稀疏性和样本数量充足,所以  $s_{it} \mid i=1,2,\dots,m$  中只有一个变量非零将以大概率事件出现。假定非零变量为  $s_{*t}$ ,则:

$$\begin{cases} x_{1t} = a_{1*} \times s_{*t} \\ x_{2t} = a_{2*} \times s_{*t} \\ \cdots \\ x_{nt} = a_{n*} \times s_{*t} \end{cases}. \quad (4)$$

从式(4)中可以看出,稀疏化后光谱某次采样即某个波长点的信息是混合矩阵  $\mathbf{A}$  的列向量与某个常数  $s_{*t}$  的乘积。在  $\mathbf{R}^n$  维空间中,光谱  $\mathbf{X}$  的列向量聚集分布在混合矩阵  $\mathbf{A}$  列向量组成的  $m$  个直线方向上。对光谱信号的列向量分布进行聚类就能确定混合矩阵的列向量方向,完成光谱混合矩阵的估计。

目前,常用的混合矩阵的估计方法有 K 均值和模糊 C 均值聚类算法。这两种算法的计算效率较高,当样本类内聚特征明显时,将达到较好的聚类效果。但在工程实际中,光谱在执行稀疏变换时因冗余字典中的单一原子不能代表所有的光谱信息,故稀疏度不能设为 1,这使得  $s_{it} \mid i=1,2,\dots,m$  中非零分量的个数大于 1。此时聚类特征是模糊的,获得样本类内聚状特征是较为困难的,且传统聚类方法不能准确判定聚类数目,严重制

约了估计精度。针对此问题,本文将 AP 聚类算法应用到混合矩阵估计中,以改善估计性能。

AP 聚类算法由 Frey 提出<sup>[19]</sup>,该算法的显著优点在于聚类数目不需要一开始就确定,而是在消息传递和数据迭代过程中自动生成,且其收敛速度相对较快。在 AP 算法中,使用负欧式距离定义相似值:

$$S(i, j) = -\|x_i - x_j\|^2. \quad (5)$$

所有数据点间的相似值可以组成一个相似度矩阵,相似度矩阵中每个点代表光谱数据样本间的相似程度,即样本  $j$  成为样本  $i$  聚类中心的合适程度。相似度矩阵对角线上的点  $S(i, i)$  称为倾向值,它表示第  $i$  个样本点作为样本聚类中心的概率。在运算前可以人工指定倾向值,用于控制和影响最后的聚类中心数目,但倾向值在迭代和消息传递过程中会逐步改变并展现最终的特性。初始倾向值设置越大,聚类数目越多;相反,则聚类数目越小。

AP 聚类算法在数据迭代当中主要传递两类信息:一类是吸引度信息,另一类是归属度信息。吸引度定义为样本  $i$  向其候选聚类中心样本  $j$  传递的信息,标记为  $r(i, j)$ ,反映样本  $j$  作为样本  $i$  聚类中心的适合程度。归属度定义为样本聚类中心  $j$  向样本  $i$  传递的信息,标记为  $a(i, j)$ ,反映样本对其聚类中心  $j$  的支持程度。它们分别表示为:

$$r(i, j) = s(i, j) - \max_{j' \neq j} \{a(i, j') + s(i, j')\}, \quad (6)$$

$$a(i, j) = \max\{0, r(j, j) + \sum_{i' \neq (i, j)} \max\{0, r(i', j)\}\}. \quad (7)$$

吸引度和归属度存在相互竞争,吸引度使用样本中其它的候选聚类中心来触发竞争,而归属度是通过考察其它数据点的吸引度来触发竞争。进一步按式(8)定义自发归属度,用以反映样本点作为聚类中心的能力。

$$a(j, j) = \sum_{i' \neq (i, j)} \max\{0, r(i', j)\}. \quad (8)$$

综上所述,整个 AP 算法流程如下:

- (1) 人为确定倾向值  $S(i, i)$ , 计算相似值  $S(i, j)$ , 形成相似度矩阵  $S$ ;
- (2) 计算吸引度  $r(i, j)$ ;
- (3) 计算归属度  $a(i, j)$  和自发归属度  $a(j, j)$ ;

(4) 重复步骤(2)~(3)并进行迭代,直到聚类中心不再变化或达到停止迭代次数;

(5) 如果  $a(j, j) + r(i, j) > 0$ , 则样本点  $j$  为自动涌现的聚类中心;对于样本  $i$ , 寻找样本  $j$ , 使  $a(i, j) + r(i, j)$  最大, 则对应的样本  $j$  为样本真正属于的聚类中心。

## 2.3 稀疏分量回归预测模型

定义  $\mathbf{Y}^{\text{cal}}$  为样本校正集近红外光谱矩阵,  $\mathbf{Y}^{\text{pred}}$  为验证集光谱矩阵,  $\mathbf{X}^{\text{cal}}$  为稀疏化后校正集光谱矩阵,  $\mathbf{X}^{\text{pred}}$  为稀疏化后验证集光谱矩阵,  $\mathbf{C}^{\text{cal}}$  为校正集某一化学物理参数的测量值,  $\mathbf{A}^{\text{cal}}$  为样本校正集光谱混合矩阵,  $\mathbf{A}^{\text{pred}}$  为验证集光谱混合矩阵。SCA 回归校正模型如图 1 所示。光谱 SCA 回归校正方法首先用近红外光谱仪对校正集和验证集样本进行光谱采样, 获得  $\mathbf{Y}^{\text{cal}}$  和  $\mathbf{Y}^{\text{pred}}$ 。然后使用  $\mathbf{Y}^{\text{cal}}$  结合 K-SVD 字典训练算法完成对冗余字典的训练, 并使用 OMP 算法求得稀疏化后的光谱  $\mathbf{X}^{\text{cal}}$  和  $\mathbf{X}^{\text{pred}}$ 。再用 SCA 算法估计校正集和验证集混合矩阵  $\mathbf{A}^{\text{cal}}$ ,  $\mathbf{A}^{\text{pred}}$ , 最后使用  $\mathbf{A}^{\text{cal}}$  和  $\mathbf{C}^{\text{cal}}$  建立偏最小二乘(PLS)回归校正模型, 将  $\mathbf{A}^{\text{pred}}$  作为模型输入, 即可计算得验证集样本某一物理化学参数的预测值。

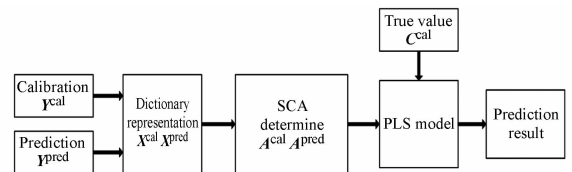


图 1 稀疏分量分析回归校正模型

Fig. 1 Regression model for Sparse Component Analysis(SCA)

## 3 预测柴油品质参数及结果分析

### 3.1 柴油样本及近红外光谱采集

使用 Nicolet 公司生产的 6700 型傅里叶变换近红外光谱仪(FT-NIRS)对同一个批次的 200 个柴油样本采集近红外光谱, 选取 150 个作为校正集, 50 个作为验证集。光谱采集的波长为 750~1 550 nm, 扫描间隔为 6 nm。采集到的柴油近红外光谱如图 2 所示。实验采用 SCA 方法对柴油近红外光谱进行解析并预测柴油的沸点、密度、芳烃总量 3 个品控参数, 全部算法在 Matlab2010a

平台上编程实现。校正集样本的沸点值用 ASTM D86 方法测定,为 197~293 °C。密度用 ASTM D4052 方法测定,为 0.804 3~0.869 5 g/mL。芳烃总量采用 ASTM D518 规定方法测定,含量为 14.3%~44.2%。

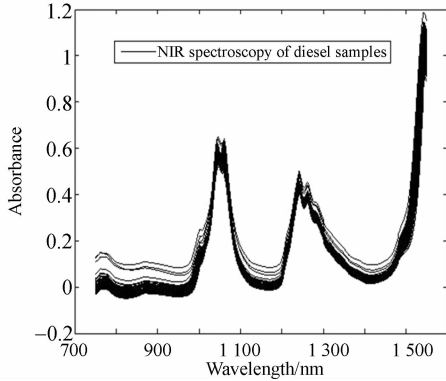


图 2 柴油近红外光谱

Fig. 2 Near infrared spectrum of diesel

### 3.2 稀疏度对预测结果的影响

在近红外光谱稀疏化的过程中,稀疏度即光谱中非零元素的个数,关系到后期预测的精度和稳定性。在式(4)的聚类求解中,理论上要求稀疏度为 1,这在实际应用中是很难达到的。稀疏度为 1 意味着只用字典中一个原子来表示光谱信息,这必然损失光谱的特征信息。当稀疏度过多时,虽然表面上增加了信息细节,但稀疏降噪理论<sup>[20]</sup>认为细节部分混有大量的噪声,降低了光谱的解析度。另外,原子数过大还会给混合矩阵的聚类估计造成障碍。故合理选择稀疏度是预测的关键。图 3~图 5 给出了稀疏度为 31~40 时沸

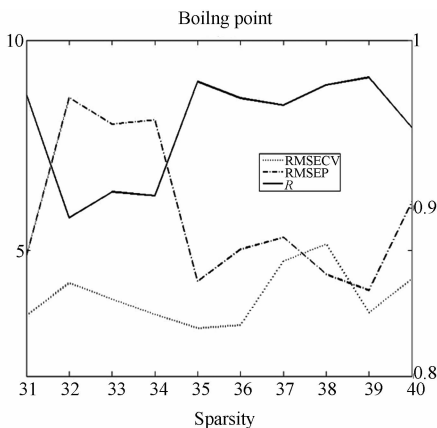


图 3 不同稀疏度下沸点的预测性能

Fig. 3 Prediction of boiling point with different sparsities

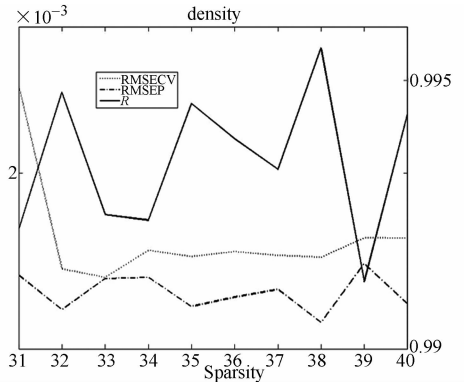


图 4 不同稀疏度下密度的预测性能

Fig. 4 Prediction of density with different sparsities

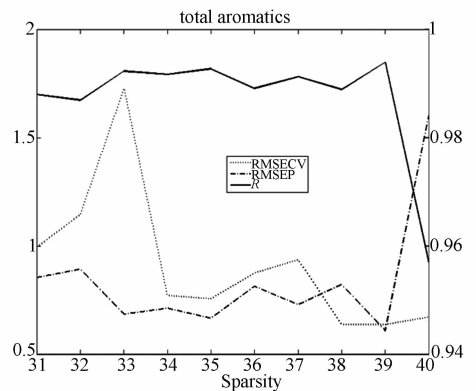


图 5 不同稀疏度下芳烃总量的预测性能

Fig. 5 Prediction of total aromatics with different sparsities

点、密度、芳烃总量预测模型的评价参数。模型评价参数选用校正集“留一”交叉验证标准差(RMSECV)、验证集预测相关系数(R)和预测均方根误差(RMSEP)<sup>[21-22]</sup>。

分析图 3~图 5 可以发现, RMSECV 的变化趋势与 R 的变化具有极大的负相关性,和 RMSEP 的变化具有极大的正相关性,这与文献[13]的结论吻合。由此表明, RMSECV 可作为建模时稀疏度选择的依据。

### 3.3 不同聚类方法对 SCA 算法预测结果的影响

用 SCA 算法建立柴油芳烃总量预测模型,在混合矩阵估计步骤中,分别使用了 K 均值聚类(k-means cluster)、模糊 C 均值聚类(fcm cluster)和 AP 聚类(AP cluster)算法。表 1 给出了不同聚类算法的验证集预测相关系数。其中, k-means cluster 和 fcm cluster 算法的类别数由人为指定,而 AP cluster 的类别数在指定倾向值后由算法自动确定,倾向值的计算可参见文献[19]。

表1 不同聚类方法的预测相关系数

Tab.1 Prediction correlations with different cluster methods

No. of classes	k-means	fcm	AP
5	77.82%	85.80%	
10	fail	96.36%	
20	fail	98.24%	
30	fail	99.35%	99.43%
40	fail	99.03%	
50	fail	68.53%	
60	fail	98.35%	

k-means cluster 算法在人为指定类别数大于 5 时,部分迭代计算得到的聚类中心并没有被分配样本,即出现了空类,最终导致预测失败。而 fcm cluster 算法性能要优于 k-means 算法,在指定类别数后,算法完成了分类,但是性能尚未超越 AP 聚类算法。实验中发现,k-means cluster 和 fcm cluster 算法的稳定性有待改进,在聚类数指定不准的极端情况下,预测性能会大大下降。而 AP 聚类算法中类别数在聚类过程中自动生成,克服了传统算法缺点,性能最优,后续品质参数预测实验均采用了 AP 聚类算法。

### 3.4 柴油近红外光谱稀疏分量预测结果分析

实验中除用校正集光谱结合提出的 SCA 算法建立了柴油品控参数的稀疏分量预测模型外,同时采用全谱 PLS<sup>[23]</sup>、ICA 法(实施算法为 FastICA)<sup>[24]</sup>建立回归模型并对验证集样本的沸点、密度、芳烃总量 3 个参数进行了预测,其预测结果用于和 SCA 回归模型进行对比。表 2 给出了 3 种算法的验证集的最优预测结果。

表2 不同方法的预测结果

Tab.2 Prediction results with different methods

criteria		PLS	ICA	SCA
Boiling point	R	94.27%	91.42%	98.91%
	SEP	5.71	6.97	2.84
	ICs	6	12	69
Density	R	97.36%	97.81%	99.68%
	SEP	$2.46 \times 10^{-3}$	$2.25 \times 10^{-3}$	$0.88 \times 10^{-3}$
	ICs	5	13	47
Total aromatics	R	99.02%	99.15%	99.43%
	SEP	0.85	0.79	0.59
	ICs	5	13	52

由表 2 可知,柴油的近红外全谱 PLS、ICA 和 SCA 算法对沸点预测的 R,SEP 分别为 94.27%,91.42%,98.91%和 5.71,6.97,2.84,对密度预测的 R,SEP 分别为 97.36%,97.81%,99.68%和  $2.46 \times 10^{-3}$ , $2.25 \times 10^{-3}$ , $0.88 \times 10^{-3}$ ,对芳烃总量预测的 R,SEP 分别为 99.02%,99.15%,99.43%和 0.85,0.79,0.59。对比分析可知,SCA 法明显地提升了油品控参数的预测性能。SCA 算法中 AP 聚类优化估计的分量数在 3 个参数检测中均超过了 40,远大于 ICA 算法估计的分量数。ICA 算法提取的分量数不能超过训练的样本数,其计算过程进行了白化降维,不能对光谱成分进行更为精细的解析;而 SCA 从光谱的稀疏特征出发,一方面克服了 ICA 算法要求分量独立假设的制约,另一方面弥补了 ICA 算法降维过程的信息损失,提高了近红外光谱的预测性能和稳定性。

## 4 结 论

本文提出了一套近红外光谱结合 SCA 检测柴油品质参数的方法。该方法利用校正集光谱样本训练了冗余字典,并使用训练后的冗余字典对光谱进行了稀疏变换。使用 AP 聚类算法对变换后的光谱实施了盲源分离,完成了混合矩阵估计。最后利用混合矩阵与品质参数建立了 PLS 预测模型,用于柴油品质参数的预测。后期实验采集了柴油样本 200 个,利用本文提出方法快速预测了它们的沸点、密度和芳烃总量。

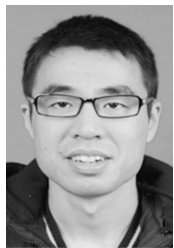
实验结果表明,SCA 方法能在柴油近红外光谱回归预测中有效提取与品控参数有关的分量,建立的回归模型对沸点、密度、芳烃总量预测的 R,RMSEP 分别达到了 98.91%,99.68%,99.43%和 2.84, $0.88 \times 10^{-3}$ ,0.59,性能超过了 ICA 和全谱 PLS 等方法,不仅验证了光谱稀疏特征与物质内在特性具有相对较强的相关性,还克服了 ICA 光谱分析算法受众多假设前提的制约。SCA 法作为新的光谱盲源分析手段在光谱定量分析领域的应用,给近红外光谱快速定量分析提供了有效参考。

## 参考文献:

- [1] JULIO C L A, CLAUDETE B H, RONEI J P. Determination of diesel quality parameters using support vector regression and near infrared spectroscopy for an in-line blending optimizer system[J]. *Fuel*, 2012, 97: 710-717.
- [2] WANG S J, WU D, LIU K S. Semi-supervised machine learning algorithm in near infrared spectral calibration; a case study on diesel fuels[J]. *Advanced Science Letters*, 2012, 11(1): 416-419.
- [3] ZHANG W B, YUAN W Q, ZHANG X M, et al.. Predicting the dynamic and kinematic viscosities of biodiesel-diesel blends using mid- and near-infrared spectroscopy[J]. *Applied Energy*, 2012, 98: 122-127.
- [4] LI H D, LIANG Y Z, LONG X X, et al.. The continuity of sample complexity and its relationship to multivariate calibration; a general perspective on first-order calibration of spectral data in analytical chemistry [J]. *Chemometrics and Intelligent Laboratory System*, 2013, 122(3): 23-30.
- [5] de CARVALHO ROCNA W F, NOGUEIRA R, VAZ B G. Validation of model of multivariate calibration: an application to the determination of biodiesel blend levels in diesel by near-infrared spectroscopy[J]. *Journal of Chemometrics*, 2012, 26(8-9): 456-461.
- [6] SCHOLZ M, GAATZEK S, STERLINGA, et al.. Metabolite fingerprinting; detecting biological features by independent component analysis[J]. *Bioinformatics*, 2004, 20(15): 2447-2454.
- [7] SHAO X G, WANG W, HOU Z Y, et al.. A new regression method based on independent component analysis[J]. *Talanta*, 2006, 69: 676-680.
- [8] TOIVIAINEN M, CORONA F, PAASO J, et al.. Blind source separation in diffuse reflectance NIR spectroscopy using independent component analysis [J]. *Journal of Chemometrics*, 2010, 24(10): 514-522.
- [9] 方利明, 林敏. 柴油近红外光谱的独立分量分析方法[J]. *石油学报(石油加工)*, 2008, 24(6): 726-732.
- FANG L M, LIN M. Near infrared spectra analysis of diesel fuel by independent component analysis [J]. *Acta Petrolei Sinica (Petroleum Processing Section)*, 2008, 24(6): 726-732. (in Chinese)
- [10] HYVÄRINEN A. Independent component analysis; recent advances[J]. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2013, 371: 1-20.
- [11] WALDMANN I P, TINETTI G, DEROO P, et al.. Blind extraction of an exoplanetary spectrum through independent component analysis[J]. *The Astrophysical Journal*, 2013, 766(1): 7-15.
- [12] ASAEI A, DAVIES M E, BOURLARD H, et al.. Computational methods for structured sparse component analysis of convolutive speech mixtures[C]. *Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on IEEE*, 2012: 2425-2428.
- [13] 周扬, 戴曙光, 吕进, 等. 训练字典及其稀疏表示在近红外光谱法检测柴油中的应用[J]. *化学学报*, 2012, 70(18): 1969-1973.
- ZHOU Y, DAI SH G, LÜ J, et al.. Application of train dictionary and sparse representation in diesel property detection using near-infrared spectroscopy[J]. *Acta Chimica Sinica*, 2012, 70(18): 1969-1973. (in Chinese)
- [14] 董天宝, 杨景曙. 稳健的欠定混合矩阵盲辨识[J]. *宇航学报*, 2013, 34(3): 426-433.
- DONG T B, YANG J SH. A robust underdetermined mixing matrix estimation algorithm [J]. *Journal of Astronautics*, 2013, 34(3): 426-433. (in Chinese)
- [15] 杨合龙, 祝磊, 韩斌, 等. 运用近邻传播聚类分析进行 SELDI-TOF 蛋白质谱特征选择[J]. *中国生物医学工程学报*, 2013, 32(1): 14-20.
- YANG H L, ZHU L, HAN B, et al.. Feature selection and sample classification for SELDI-TOF mass spectrometry data based on affinity propagation clustering[J]. *Chinese Journal of Biomedical Engineering*, 2013, 32(1): 14-20. (in Chinese)
- [16] SMITH L N, OLSON C C, JUDDK P, et al.. Denoising infrared maritime imagery using tailored dictionaries via modified K-SVD algorithm [J]. *Applied Optics*, 2012, 51(17): 3941-3949.
- [17] 焦李成, 杨淑媛, 刘芳, 等. 压缩感知回顾与展望[J]. *电子学报*, 2011, 39(7): 1651-1662.
- JIAO L CH, YANG SH Y, LIU F, et al.. Development and prospect of compressive sensing[J]. *Acta Electronica Sinica*, 2011, 39(7): 1651-1662. (in Chinese)

- [18] 王勇,王雪,孙欣尧. 基于稀疏信号重构的无线传感网络目标定位[J]. 仪器仪表学报, 2012, 33(2): 362-368.  
WANG Y, WANG X, SUN X Y. Target location in wireless sensor networks using spares signal reconstruction[J]. *Chinese Journal of Scientific Instrument*, 2012, 33(2): 362-368. (in Chinese)
- [19] FREY B J, DUECK D. Clustering by passing messages between data points[J]. *Science*, 2007, 972(315): 972-976.
- [20] MAHMOUDI M, SAPIRO G. Sparse representations for range data restoration[J]. *IEEE Transactions on Image Processing*, 2012, 21(5): 2909-2915.
- [21] 刘燕德,周延睿,彭彦颖. 基于近红外漫反射光谱检测鸡蛋品质[J]. 光学精密工程, 2013, 21(1): 40-45.  
LIU Y D, ZHOU Y R, PENG Y Y. Detection of egg quality by near infrared diffuse reflectance spectroscopy[J]. *Opt. Precision Eng.*, 2013, 21(1): 40-45. (in Chinese)
- [22] 曹璞,潘涛,陈星旦. 小型近红外玉米蛋白质成分分析仪器设计的波段选择[J]. 光学精密工程, 2007, 15(12): 1951-1958.  
CAO P, PAN T, CHEN X D. Choice of wave band in design of minitype near-infrared corn protein content analyzer[J]. *Opt. Precision Eng.*, 2007, 15(12): 1951-1958. (in Chinese)
- [23] 黄富荣,潘涛,张甘霖,等. 应用近红外漫反射光谱快速测定土壤锌含量[J]. 光学精密工程, 2010, 18(3): 586-592.  
HUANG F R, PAN T, ZHANG G L, et al.. Rapid measurement of zinc contents in soils by near-infrared diffuse reflectance spectroscopy[J]. *Opt. Precision Eng.*, 2010, 18(3): 586-592. (in Chinese)
- [24] VAN L D, WU D Y, CHEN C S. Energy-efficient fast ICA implementation for biomedical signal separation[J]. *IEEE Transactions on Networks*, 2011, 22(11): 1809-1822.

## 作者简介:



周扬(1983—),男,浙江宁波人,博士研究生,讲师,2009年于北京航空航天大学获得硕士学位,主要从事光谱检测、光电信息技术方面的研究。E-mail: zybuaa@163.com

## 导师简介:



戴曙光(1979—),男,上海人,教授,博士生导师,1982年于上海交通大学获得学士学位,1988年于上海机械学院获得硕士学位,主要从事光学信号处理,在线检测方面的研究。E-mail: usstdsg@163.com

(版权所有 未经许可 不得转载)