DOI: 10.19650/j.cnki.cjsi.J2312096

# 多层非线性局部感受野极限学习机方法用于录井气体分析*

李忠兵[1],袁章雨[1],梁海波[2],谌贵辉[1],蒋川东[1]
(1.西南石油大学电气信息学院 成都 610500; 2.西南石油大学机电工程学院 成都 610500)

**摘 要**:随着我国能源需求的不断提升以及钻探环境的日益复杂化,开展高精度的烷烃类气体浓度检测对于提高油气勘探效率具有重要意义。光谱录井技术具有烃类气体检测快速、准确等优势,已成为石油勘探过程中备受关注的研究热点。针对录井气体红外光谱由于饱和吸收、噪声干扰、基线漂移等方面引起的非线性问题,提出了多层非线性局部感受野极限学习机(NM-LRF-ELM)模型。该模型将一维光谱数据转换为二维矩阵格式,利用局部感受野的数据处理方式在输入与隐藏层之间实现非线性特征提取。同时,引入改进的 T-sigmoid 激活函数,并在全连接层后加入 dropout 层来降低模型的过拟合风险。模型的特征提取与定量分析呈一体化结构,直接输出定量分析预测值。采集了两组共 407 个混合烷烃气体样本的红外光谱作为实验数据集,进行定量分析实验。实验结果表明,相较于滑动窗口类与灰狼优化定量分析模型,该模型的训练时间显著减少了 90%以上。即使在同系物的非线性干扰下,模型的预测精度仍低于系统误差。因此,提出的方法有助于在现场环境变化复杂的情况下,降低未知气体的非线性干扰,提高对目标气体的红外光谱检测精度。

**关键词**:气测录井;红外光谱;定量分析;局部感受野极限学习机
**中图分类号**:TH741 **文献标识码**:A **国家标准学科分类代码**:410.55

# Multi-layer nonlinear local receptive field extreme learning machine method for logging gas analysis

Li Zhongbing[1], Yuan Zhangyu[1], Liang Haibo[2], Chen Guihui[1], Jiang Chuandong[1]
(*1. School of Electrical Engineering and Information, Southwest Petroleum University, Chengdu 610500, China*;
*2. School of Mechatronic Engineering, Southwest Petroleum University, Chengdu 610500, China*)

**Abstract**: With China's increasing energy demand and the complex drilling environment, it is of great significance to carry out high-precision detection of alkane gas concentration to improve oil and gas exploration efficiency. Spectral logging technology has become a research hotspot in the process of oil exploration with the advantages of quick and accurate recording results. In this article, a multi-layer nonlinear local receptive field extreme learning machine (NM-LRF-ELM) model is proposed for resolving nonlinear problems caused by saturation absorption, noise interference, and baseline drift. The model converts one-dimensional spectral data into two-dimensional matrix format and realizes nonlinear feature extraction between input and hidden layer by using local receptive field data processing. Meanwhile, an improved T-sigmoid activation function is introduced and the dropout layer is added after the fully connected layer to reduce the overfitting risk of the model. The feature extraction and quantitative analysis of the model show an integrated structure and directly outputs the predicted value of quantitative analysis. In this article, the infrared spectra of 407 mixed alkane gas samples from two groups are collected as an experimental data set for quantitative analysis. The experimental results show that the training time of this model is reduced by more than 90% compared with the sliding window model and the gray Wolf model, and the prediction accuracy of the model is still lower than the system error under the nonlinear interference of the homolog. Therefore, the proposed method is helpful in reducing the nonlinear interference of unknown gas and improve the infrared spectrum detection accuracy of target gas under the

收稿日期:2023-10-29 Received Date: 2023-10-29
*基金项目:油气藏地质及开发工程国家重点实验室开放基金(PLN2022-42)、国家自然科学基金(52074233)、油气生产安全与风险控制重庆市重点实验室开放基金(cqsrc202101)项目资助

condition of complex field environment changes.

**Keywords**: gas logging; infrared spectrum; quantitative analysis; local receptive field limit learning machine

## 0 引 言

油气是人类赖以生存的重要战略资源,对国家经济发展、社会稳定以及国际外交等都具有至关重要的作用[1]。烷烃混合气体的浓度预估是判断油气层产能的重要评价依据[2],由于油气的形成条件复杂且流动性较强,以往的录井技术已无法满足快速准确反映油气层信息的需求。因此,开展高精度的烷烃类气体浓度检测研究在油气勘探行业具有重要的现实意义和广泛的应用前景[3]。

光谱录井技术具有无需将各组分气体分离便能够同时分析多种气体组分信息的优点[4],迅速发展为新型气测录井技术。其中,红外光谱技术具有快速、高效、准确性好,无损检测等特点,目前被广泛地应用于气测录井领域。Stoil 等[5]介绍了光谱技术在意大利托斯卡纳州阿尔贝格纳河三角洲等地的应用,为沿沉积物岩心连续记录矿物丰度提供简单而省时的方法。荆文峰等[6]介绍了红外光谱技术在长庆油田开发中的创新应用,证明了该技术对于油田"提速、提效、提质、提产"中的显著作用。然而在实际应用过程中,红外光谱检测技术仍面临着检测浓度范围跨度大、特征谱峰重叠现象严重等问题。

烃类气体同系物的分子结构具有较高的相似性,导致烃类物质的特征红外吸收峰会产生严重的重叠和干扰,各物质特征难以识别。由于钻井复杂地质环境的影响,录井过程中采集到的红外光谱信号容易受到大量噪声或者无用信号的干扰。因此,建立一个高精度且高稳定性的红外光谱定量分析模型成为了红外光谱气测录井的关键任务。

传统红外光谱分析过程主要包含预处理、特征提取和定量分析等,这几个过程是独立的。许多学者通过改进特征提取方法来提高定量分析模型的性能。Li 等[7]对统计学的滑动窗口方法进行改进,提出了自适应滑动窗口偏最小二乘建模方法(AMW-PLS),并应用于研制的红外光谱气测录井系统中,对甲烷($C_1$)、乙烷($C_2$)、丙烷($C_3$)、正丁烷($nC_4$)以及正戊烷($nC_5$)5 种烷烃物质进行定量分析,平均相对误差为 3.365 1%。韩建等[8]提出了自-互相关联合(SCC)光谱超稀疏波段筛选方法,在筛选出的两个油水混合物波段上呈显著相关性,对油水混合物测量建模方法具有一定参考价值。基于统计学的特征提取方法[9-10]能够有效地提高定量分析模型的稳定性,但建模效率不高,并且区间内仍存在冗余变量,使得模型精度有待提升。Song 等[11]从仿生学角度利用遗传算法对烷烃混合气体的红外光谱进行特征筛选,定量分析测井气体的组分时平均预测误差为 1.78%。李玉军等[12]提出粒子群优化的最小二乘支持向量机方法,在 0.1% ~ 1% 的甲烷、乙烷及 0.1% ~ 1.5% 的丙烷混合气体定量分析中预测结果均方误差达 $1.82\times10^{-7}$。虽然基于仿生学的特征提取算法[13-14]在一定程度上提高了预测精度,但该类算法往往需要预设大量参数,这些参数会对定量分析模型的性能产生影响。

传统光谱分析将特征提取与定量分析分开,有利于提高模型的稳定性,但不能较好地解决建模效率不足及非线性干扰导致模型定量分析精度不高等问题。随着深度学习的快速发展,神经网络算法逐渐被应用于红外光谱的定量分析中。与过去不同的是,神经网络算法的特征提取不再与定量分析严格分开,而是呈一体化结构,直接通过输出层实现定量分析。随着大量训练的积累,该方法可以兼顾模型的精度及稳定性,且具有较高的建模效率,逐渐受到重视。Song 等[15]建立了基于卷积神经网络(convolutional neural networks, CNN)的录井气体光谱分析模型,完成了对甲烷、乙烷、丙烷、正丁烷及正戊烷的定量分析,对经过预处理的单质气体和混合气体识别准确率可达 99% 和 98%。郝惠敏等[16]采用 SDP-VGG 法建立七组分混合烷烃气体浓度识别模型,对甲烷、乙烷、丙烷、正丁烷、异丁烷($iC_4$)、正戊烷及异戊烷($iC_5$)混合烷烃气体中的甲烷浓度识别准确率达到 91.2%。深度学习算法在红外光谱分析中的应用仍处于初级阶段,数据维度不匹配、网络规模大、参数的最优选择难、模型可解释性差等相关问题还有待进一步探讨。

2015 年 Huang 等[17]提出了基于局部感受野的极限学习机(local receptive field extreme learning machine, LRF-ELM)模型,并将其用于图像处理。LRF-ELM 模型采用稀疏的连接方式将输入与隐藏层相连,每个连接都有相应的局部感受野,并且每个池化层节点相当于几个隐层节点的集合。这种设计结构旨在缩短模型的训练时间,同时提高模型的稳定性和对非线性数据的处理能力。此后,Xiao 等[18]将该模型与红外光谱结合应用于煤分类;Zhang 等[19]将该模型与拉曼光谱结合应用于奶酪产品识别。由于该模型在录井气体的红外光谱分析中还没有应用先例,能否解决红外光谱吸收峰重叠和干扰产生的非线性问题仍待研究。

尽管局部感受野极限学习机模型在红外光谱分析中展现出潜力,但仍未较好的解决数据维度不匹配、过拟合、录井气体吸收峰重叠干扰等问题。为此,本文提出了一种多层非线性局部感受野极限学习机(multi-layer

nonlinear local receptive field extreme learning machine, NM-LRF-ELM)模型,旨在解决这些问题并提高模型的性能。该模型将原始一维红外光谱的吸光度数据按一定波长反射比率转换成二维矩阵格式来匹配改进模型的输入,然后经卷积池化等操作实现录井气体红外光谱定量分析模型的建立。本文采集了两组烷烃气体红外光谱数据集进行实验,并讨论模型的定量分析效果。本文通过实验比较不同激活函数及 dropout 率,验证参数设置的合理性;然后,利用另一组数据集验证了所提模型的泛化能力和数据局限性,通过调整参数建立不同物质的高精度定量分析模型;最后,通过与滑动窗口类和灰狼优化定量分析模型进行对比实验,验证本文模型在训练时间和定量分析效果方面的性能。

## 1　基本原理

LRF-ELM 算法中,输入层与隐藏层节点之间的连接权重采用连续概率分布随机生成,这种随机生成的连接构成局部感受野。LRF-ELM 架构如图 1 所示,该算法通过卷积、池化提取输入样本的特征后,再通过全连接层将特征映射到输出空间。

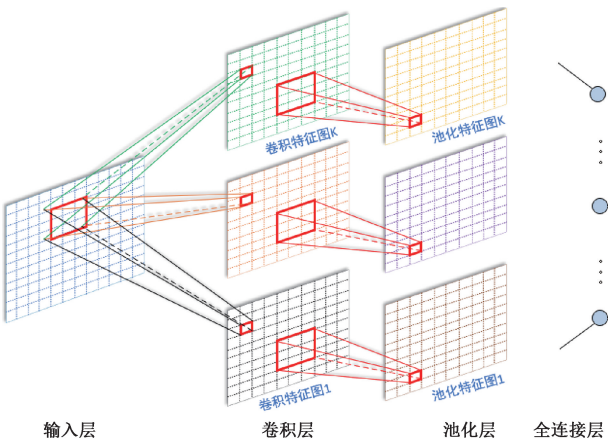



图 1　LRF-ELM 架构

Fig. 1　The architecture of LRF-ELM

该算法采用 $K$ 个不同的输入权值来获得 $K$ 个不同的特征图,并对其进行特征映射和特征组合,其中具体包含了如下 4 个操作。

1)随机生成初始权重并正交化,为了获得输入的充分表示,采用 $K$ 个不同的输入权重,从而得到 $K$ 个互异的特征图,然后随机生成初始权重:$\hat{\boldsymbol{A}}^{init}=[\hat{\boldsymbol{a}}_1^{init},\hat{\boldsymbol{a}}_2^{init},\cdots,\hat{\boldsymbol{a}}_K^{init}]$,$\hat{\boldsymbol{a}}_k^{init}\in\boldsymbol{R}^{r^2}$,$k=1,2,\cdots,K$,其中 $\hat{\boldsymbol{A}}^{init}\in\boldsymbol{R}^{r^2\times K}$。然后采用奇异值分解(singular value decomposition, SVD)正交化初始权重,正交化后的矩阵记为 $\hat{\boldsymbol{A}}$,它的每一列 $\hat{\boldsymbol{a}}_k$ 都是 $\hat{\boldsymbol{A}}^{init}$ 的正交基。当 $r^2<K$ 时,需要完成先转置再正交化再转置的过程。

2)卷积操作,设输入卷积层的大小为 $d\times d$,感受野大小为 $r\times r$,那么卷积的特征图(feature map)的大小为 $(d-r+1)\times(d-r+1)$。第 $K$ 个特征图的输入权重 $\boldsymbol{a}_k\in R^{r\times r}$,由 $\hat{\boldsymbol{a}}_k$ 逐列排成。第 $K$ 个卷积特征图节点$(i,j)$ 的值 $\boldsymbol{c}_{i,j,k}$ 为:

$$\boldsymbol{c}_{i,j,k}=\left(\sum_{m=1}^{r}\sum_{n=1}^{r}\boldsymbol{x}_{i+m-1,j+n-1}\cdot\boldsymbol{a}_{m,n,k}\right)\quad i,j=1,\cdots,(d-r+1)\tag{1}$$

3)平方根池化,池化大小 $e$ 表示池化中心到边的距离,且池化图的大小与卷积特征图的大小相同即$(d-r+1)\times(d-r+1)$。$\boldsymbol{c}_{i,j,k}$ 和 $\boldsymbol{h}_{p,q,k}$ 分别表示第 $K$ 个特征图中的节点$(i,j)$ 和第 $K$ 个池化图中的组合节点$(p,q)$ 的值,其中 $\boldsymbol{h}_{p,q,k}$ 为:

$$\boldsymbol{h}_{p,q,k}=\sqrt{\sum_{i=p-ej=q-e}^{p+eq+e}\boldsymbol{c}_{i,j,k}^2}\quad p,q=1,\cdots,(d-r+1)\tag{2}$$

若节点$(i,j)$不在$(d-r+1)$范围内,则 $\boldsymbol{c}_{i,j,k}=0$。

4)利用隐藏层输出矩阵 $\boldsymbol{H}$ 计算输出权重 $\boldsymbol{\beta}$,将所有组合节点的值组合成一个行向量,并把 $N$ 个输入样本的行向量放在一起,得到全连接矩阵 $\boldsymbol{H}\in\boldsymbol{R}^{N\times K\cdot(d-r+1)^2}$。输出权重 $\boldsymbol{\beta}$ 为:

$$\boldsymbol{\beta}=\begin{cases}\boldsymbol{H}^{\mathrm{T}}\left(\dfrac{\boldsymbol{I}}{\boldsymbol{C}}+\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}}\right)^{-1}\boldsymbol{T}, & N\leqslant K\cdot(d-r+1)^2\\ \left(\dfrac{\boldsymbol{I}}{\boldsymbol{C}}+\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}\right)^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{T}, & N>K\cdot(d-r+1)^2\end{cases}\tag{3}$$

由以上 ELM-LRF 原理可知,该模型的连接采用局部感受野的特征映射方式,并通过卷积和池化操作捕捉局部结构信息,经过全连接输出。ELM-LRF 能够有效地改善分类任务的性能,提高模型的分类准确度,为特征提取及定量分析提供了一种有效的方法。

## 2　多层非线性局部感受野极限学习机

本文在 LRF-ELM 基础上进行改进,形成 NM-LRF-ELM,探讨神经网络类算法在红外光谱定量分析中的应用。传统的 LRF-ELM 模型主要用于图像处理,对应的数据一般是二维数据,而本文的研究对象近红外光吸收光谱为一维数据。为此针对原网络结构进行了如下改进:1)近红外光谱数据为一维向量,与二维图像数据有较大差异,因此本文模型网络在原网络结构的输入层前增加一维光谱转二维矩阵操作;2)为了更好地提取原数据的非线性特征信息,将原网络的一层卷积池化层改为两层;3)为了降低模型在对光谱数据处理时的过拟合风险,在卷积与池化层间增加 T-sigmoid 函数作为激活函数,且在全连接层与输出层间增加 dropout 层,其结构如图 2 所

示。NM-LRF-ELM 网络结构首先将一维红外光谱的吸光度数据转换为二维矩阵格式;然后,经过两层卷积层、两层池化层,每层依据原算法选用 48 个 3×3 的卷积核,池化层选用 3×3 的平方根池化,在卷积层与池化层之间使用 T-sigmoid 函数作为激活函数;最后经全连接层、dropout 层和输出层得出预测结果。

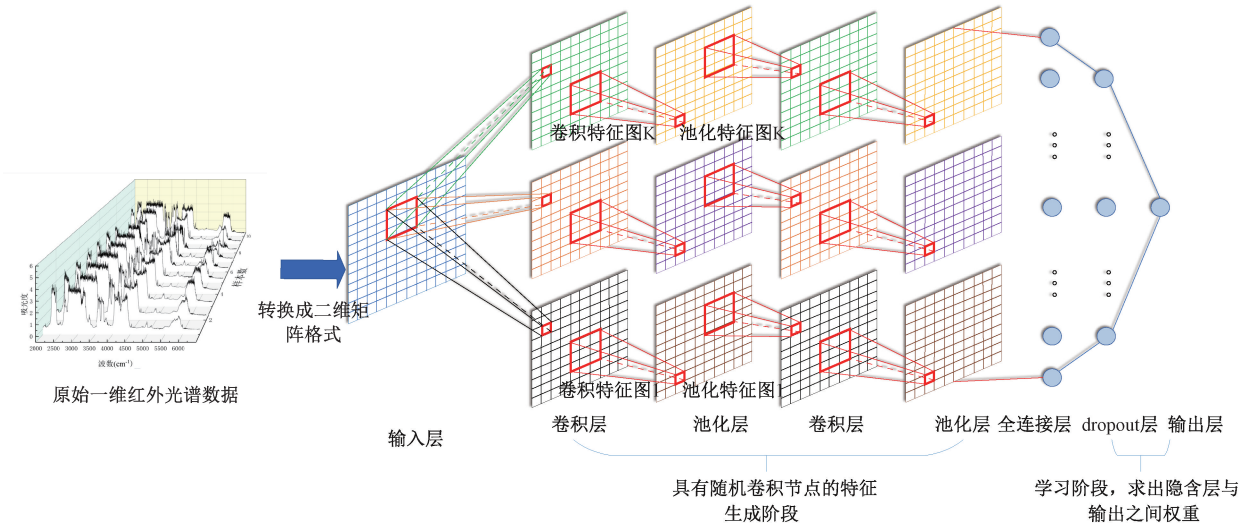



图 2 NM-LRF-ELM 网络结构

Fig. 2 The network structure of NM-LRF-ELM

## 2.1 光谱转换

为了适应改进模型二维矩阵格式的数据输入,需要在输入层前将原始一维红外光谱的吸光度数据转换为二维矩阵格式。将一维光谱转换为二维矩阵格式可以提供空间信息,从而提供更丰富的光谱特征,实现更有效的特征提取。

假设 $\boldsymbol{x}_1 = \{x_1, x_2, \cdots, x_N\}$ 为经过式(4)进行相对转换后的一维红外光谱。

$$x_i = \frac{A_i}{A_m} \quad i = 1, 2, \cdots, N \tag{4}$$

式中:$N$ 为一维光谱数据长度;$A_i$ 为第 $i$ 个波长处的吸光度值;$A_m$ 为选定的特征波长处的吸光度值。取最大的整数使 $M^2 \leqslant N$,根据式(5)转换得到二维光谱数据。

$$\boldsymbol{x}_2 = \begin{bmatrix} x_1 & x_2 & \cdots & x_M \\ x_{M+1} & x_{M+2} & \cdots & x_{2M} \\ \vdots & \vdots & & \vdots \\ x_{(M-1)M+1} & x_{(M-1)M+2} & \cdots & x_{MM} \end{bmatrix} \tag{5}$$

图 3 所示为一维光谱到二维的转换,二维光谱数据可以呈现在一维数据中无法观察到的纹理特征,这种改进使得模型能够在原始吸光度数据上进行卷积的特征提取操作,能更有效地提取原始数据的非线性特征,从而提高定量分析模型的精度。

## 2.2 特征提取

本文模型采用两层卷积池化层,用于提取原数据的非线性特征,确保模型能取得较好的预测效果。同时,为了降低模型的过拟合风险,NM-LRF-ELM 模型在卷积层与池化层之间加改进的 sigmoid 函数,并将其命名为 T-sigmoid。T-sigmoid 函数经过原点,可以将很大范围内的输入特征值压缩到-2~2,使得模型在深层网络中能够保持数据幅度不会出现较大的变化,其表达式如下:

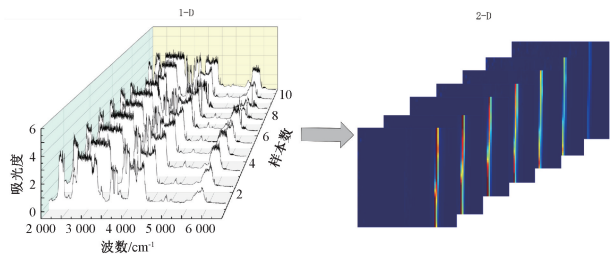



图 3 一维光谱数据到二维矩阵数据的转换

Fig. 3 The conversion of one-dimensional spectrum to two-dimensional matrix data

$$\text{T-sigmiod}(x) = 4 \times \frac{1}{1 + e^{-x}} - 2 \tag{6}$$

T-sigmoid 函数接近 $f(x) = x$ 这样一个线性模型,在一定程度上解决了 sigmoid 函数输出较为平缓现象导致模型线性解释率不高的问题;同时还能解决 sigmoid 函数的输出值在 0~1 可能会导致梯度消失的问题。

## 2.3 浓度预测

本文在全连接层与输出层间引入 dropout 层,有效缓解由全连接层参数较多导致的过拟合,在一定程度上增强模型的泛化能力。

NM-LRF-ELM 的训练目标是得到最小训练误差和最小范数的输出权重$\boldsymbol{\beta}$,输出权重矩阵通过式(7)计算得出。

$$\boldsymbol{\beta}=\begin{cases}\boldsymbol{H}^{\mathrm{T}}\left(\dfrac{\boldsymbol{I}}{\rho}+\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}}\right)^{-1}\boldsymbol{T}, & N\leqslant K\cdot(d-r+1)^2\\ \left(\dfrac{\boldsymbol{I}}{\rho}+\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}\right)^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{T}, & N>K\cdot(d-r+1)^2\end{cases} \tag{7}$$

式中:$\boldsymbol{T}$ 是训练样本的标签矩阵;$\rho$ 是正则化系数;$\boldsymbol{I}$ 是单位矩阵;全连接层矩阵 $\boldsymbol{H}\in\boldsymbol{R}^{N\times K\cdot(d-r+1)^2}$。

最终可以得到多样品、多组分混合气体的红外光谱 NM-LRF-ELM 模型计算公式如下:

$$\boldsymbol{C}=\boldsymbol{H}\boldsymbol{\beta} \tag{8}$$

式中:$\boldsymbol{H}$ 表示全连接层矩阵;$\boldsymbol{\beta}$ 表示输出层与隐藏层之间的权重矩阵;$\boldsymbol{C}$ 表示未知气体的浓度矩阵。

### 2.4　录井气体分析流程

本文提出了 NM-LRF-ELM 方法用于录井气体分析,其分析流程如下。

1)将原始一维红外光谱的吸光度数据转换为二维矩阵格式。

2)在 $K$ 个互异的特征图上产生随机输入权值 $a_k$,并采用 SVD 正交化初始权重,以正交化后的初始权重 $\hat{\boldsymbol{A}}$ 对单隐层前馈神经网络的输入层的特征图和隐藏层的特征图进行相连,生成二维矩阵特征。

3)使用随机生成的局部感受野,对步骤 2)生成的二维矩阵特征进行卷积特征映射。

4)卷积层提取特征后,将输出特征映射传递给激活函数,实现网络层的非线性变换。使用 T-sigmoid 函数作为激活函数,保证组合节点数据幅度在深层网络中不会出现较大的变化。

5)对卷积特征映射形成的组合节点的值进行平方根池化,然后将光谱数据划分多个区域并将各区域的代表特征提取出来,从而得到二维矩阵的底层特征。

6)重复步骤 3)~ 5),深入提取原数据的非线性特征。

7)在全连接层将所有得到的二维矩阵的底层特征组合到一起,获得二维矩阵数据的底层特征矩阵。

8)在 dropout 层删掉部分特征信息,增加每层各特征之间的正交性,有效缓解模型过拟合。

9)由式(7)计算输出权重$\boldsymbol{\beta}$,由式(8) 得出多样品、多组分混合录井气体浓度 $\boldsymbol{C}$。

## 3　实验数据与评价指标

### 3.1　实验数据

本文采用自制烷烃气体红外光谱采集系统采集了两组共 407 个混合烷烃气体样本的红外光谱作为实验数据集。如图 4 所示,待测实验样品以氮气为背景载气,由甲烷、乙烷、丙烷、正丁烷、异丁烷、正戊烷、异戊烷、二氧化碳($CO_2$)等标准气源(大连计量检验检测研究院)输入到 6 通道混合配气系统 LFIX-7000(莱峰、成都、中国)进行混合配比形成,输出气体浓度误差为气源浓度的±1%。混合配气系统输出的样品气体以 1 000 mL/min 的流量经可以反向吹扫气体的干燥管除湿后进入有效光程长度为 4.8 m、体积为 400 mL 的光程池,光程池外部套有恒温装置,可使光程池内部温度恒定在 27.5℃。最后通过计算机控制红外光谱仪 ALPHA Ⅱ(Bruker, Germany)采集得到波数范围为 2 000~6 500 $cm^{-1}$ 共 4 356(66×66)个波数点的光谱数据。

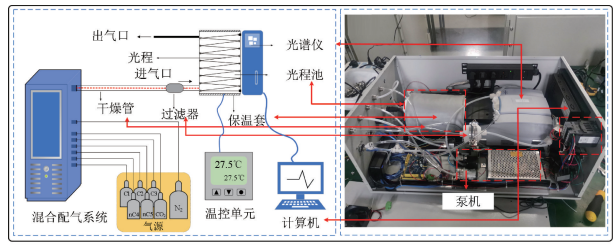



图 4　录井气体红外光谱采集系统示意及硬件结构

Fig. 4　Schematic diagram and hardware structure diagram of the infrared spectrum acquisition system for logging gas

本文使用上述采集系统所采集的 407 个红外光谱数据集按照不同的混合类型分为“双组分实验数据集”(48 个)和“7 组分实验数据集”(359 个)两个部分。

双组分数据集由标准气源浓度为 $500\times10^{-6}$ 的正丁烷和正戊烷混合组成,共 48 个样本。双组分数据集中各样本的浓度分布如图 5 所示。其中,1~22 号样本中两种物质的浓度同时递增;23~35 号样本中正丁烷的浓度继续递增,正戊烷的浓度取随机值;36~48 号样本中正戊烷的浓度递增,正丁烷的浓度取随机值。

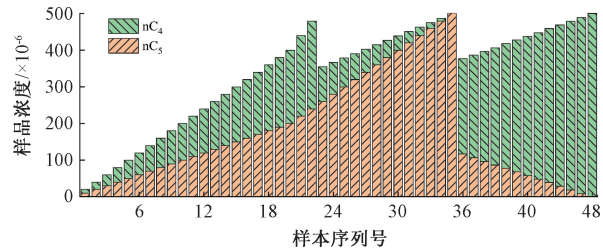


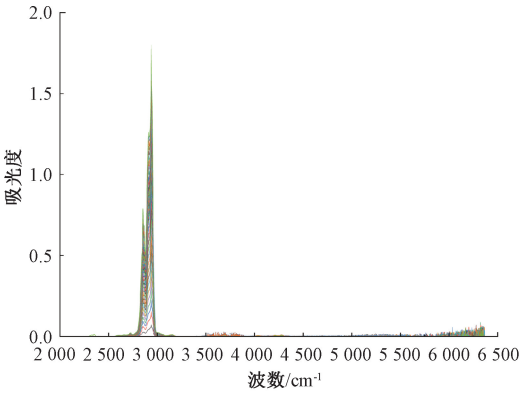

图 5　双组分数据集中各样本浓度分布

Fig. 5　Concentration distribution of samples in the two-component dataset

7 组分数据集由标气浓度为 100% 的甲烷、乙烷、丙烷、二氧化碳及标气浓度为 5% 的正丁烷、标气浓度为

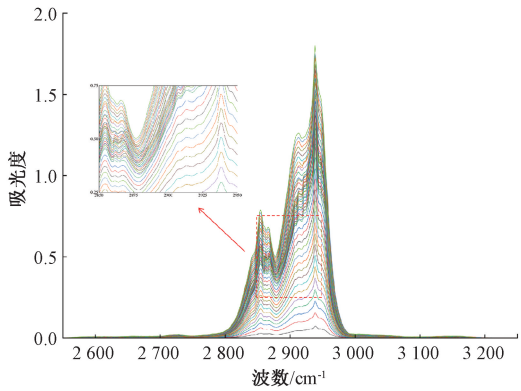
10%的异丁烷、标气浓度为4%的异戊烷等7种物质共359个样本组成。由于各组分气源标气浓度存在较大差异性，甲烷、乙烷、丙烷和二氧化碳的浓度普遍较高，远大于其他物质，对其余物质的定量分析形成较强干扰。

双组分和7组分样品的真实红外光谱曲线如图6所示，不同的样品使用不同的颜色进行标记。图6(a)和(c)分别是双组分数据集和7组分数据集的真实光谱曲线。图图6(b)和(d)分别是双组分数据集和7组分数据集的光谱局部放大图(2 550~3 250 $cm^{-1}$)。



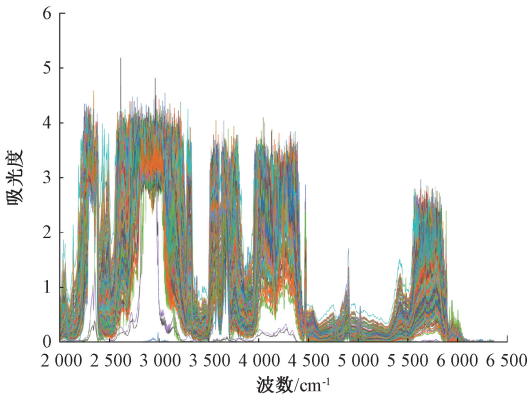

(a) 双组分数据集光谱

(a) Spectrum of two-component dataset



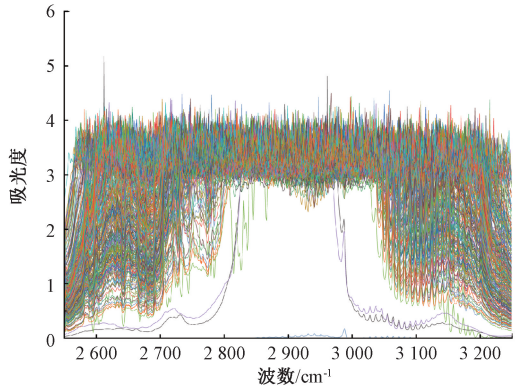

(b) 图(a)局部放大图

(b) Partial enlarged detail of (a)




(c) 七组分数据集光谱

(c) Spectrum of seven-component dataset




(d) 图(c)局部放大图

(d) Partial enlarged detail of (c)

图6 原始光谱

Fig. 6 Original spectrogram

在双组分数据集中，当两种物质组分的浓度均递增时(样本1~22)，光谱谱线依次增高，如图6(b)所示，随着其中一种组分浓度持续增加而另一种组分浓度逐渐减小，光谱谱线部分波段会出现交叉，总体上呈现较为理想的线性吸收。然而，在7组分数据集中，组分类型更为复杂，不同组分的吸收峰被严重遮蔽和重叠，组分之间的谱图呈现较强的非线性干扰，如图6(a)所示，建立高精度的定量分析模型是面向复杂录井组分检测的关键，利用机器学习模型具有较强的非线性特征提取的能力，有望建立满足现场应用的高精度定量分析模型。

### 3.2 评价指标

本文定量分析模型使用的评价指标为均方根误差(root mean squared error，RMSE)、决定系数 $R^2$ 和相对预测偏差(relative prediction deviation，RPD)。其中RMSE数值越小代表模型精度越高，反之则代表模型精度较差；$R^2$ 数值在0~1，数值越接近于1则表示模型线性相关度越高；RPD值与RMSE值呈负相关关系，其数值越大表示模型预测效果越好，3个指标公式[20]如下：

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N-1}} \tag{9}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{10}$$

$$RPD = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \bar{y})^2}{N-1}} / RMSE \tag{11}$$

式中：$i$ 为测量样本序号；$N$ 为测量样本数量；$y_i$ 为第 $i$ 个测量样本实际浓度值；$\hat{y}_i$ 为第 $i$ 个测量样本预测浓度；$\bar{y}$ 为 $N$ 个样本真实浓度的均值。本文所使用的训练集均方根误

差(RMSEC)和测试集均方根误差(RMSEP)都由式(9)得出,二者的样本分别为训练集和测试集。RMSE 通过计算预测误差的平方和的均方根来衡量模型的性能,侧重于关注预测误差的绝对大小,本文将其用于评价自制的烷烃气体红外光谱采集系统采集的双组分与 7 组分实验效果。

此外,本文采用平均相对误差 MRE 评估模型在标准工业样气上的测试效果,通过计算预测误差与真实值之间的相对差异来衡量模型的性能,侧重于相对误差的度量,其数值越小代表模型精度越高,反之则代表模型精度较差。MRE 公式如下:

$$MRE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \tag{12}$$

## 4　实验与讨论

### 4.1　模型参数有效性验证实验

为了验证引入的激活函数及 dropout 层对模型效果的影响,在物质类型更少的双组分数据集上通过对比实验来确定模型的最优参数。其中,正则化系数 $\rho$ 设置为 0.01,训练集、测试集的比例由 Kennard-Stone(KS)算法[21]划分为 75% 和 25%。

1)不同激活函数的影响

图 7 所示为 sigmoid、ReLU、T-sigmoid 以及 sin 4 种激活函数下模型用于分析正丁烷、正戊烷时的 RMSEC 和 RMSEP 变化情况。可以看出,4 种激活函数的 RMSEC 和 RMSEP 值的差值均不大,没有出现过拟合的现象。总体表现为 T-sigmoid< sin< ReLu< sigmoid,证实了引入的改进 T-sigmoid 作为激活函数时模型效果最好,能取得较好的定量预测效果。此时,正丁烷的 RMSEC 和 RMSEP 分别为 $2.76\times10^{-6}$、$2.92\times10^{-6}$,正戊烷的 RMSEC 和 RMSEP 分别为 $3.27\times10^{-6}$、$3.05\times10^{-6}$,均低于配气系统的输出误差,证明提出的模型在浓度范围较低时(待测样本浓度均低于 $500\times10^{-6}$)时具有优秀的分析能力。

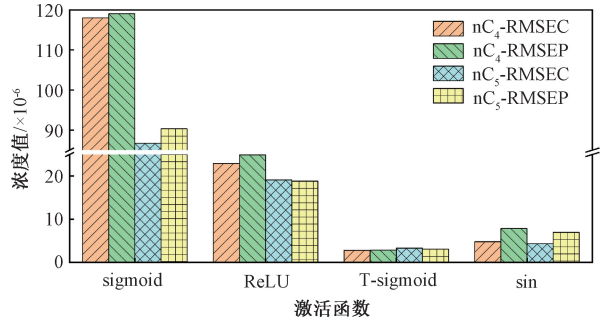



图 7　不同激活函数下模型的 RMSEC 和 RMSEP 值

Fig. 7　The values of RMSEC and RMSEP with different activation functions

2)dropout 率的影响

图 8 所示为无 dropout 层、有 dropout 层且 dropout 率分别为 0.2、0.5、0.8 时模型用于分析正丁烷、正戊烷的 RMSEC 和 RMSEP 变化情况。可以看出,无 dropout 层时,模型在训练集及测试集上的均方根误差均大于 $5\times10^{-6}$(标气气源浓度为 $500\times10^{-6}$ 时混合配气系统的输出误差),增加 dropout 层后,模型的均方根误差低于配气系统的输出误差,证实了改进模型引入的 dropout 层对于预测效果以及过拟合都有较好的改善能力。当 dropout 率设置为 0.2 时、0.5 时,模型效果相近,均方根误差值均在 $3\times10^{-6}$ 左右。

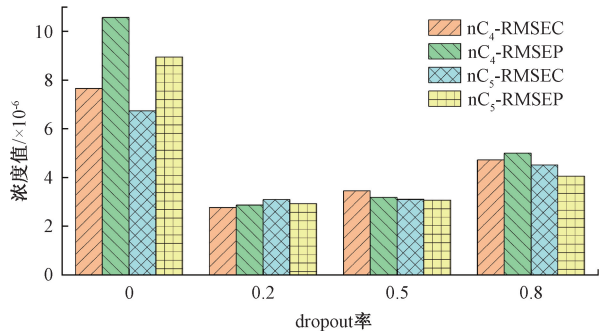
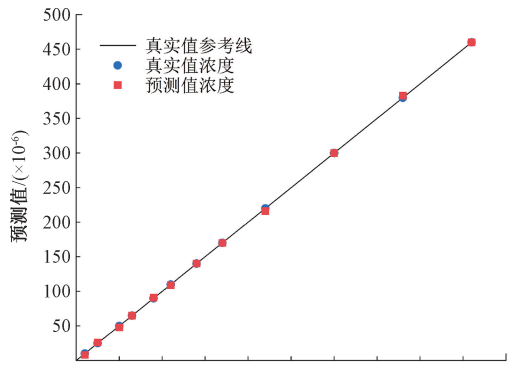



图 8　不同 dropout 率下模型的 RMSEC 和 RMSEP 值

Fig. 8　The values of RMSEC and RMSEP with different dropout rates

3)最优参数下的模型定量分析

本文 NM-LRF-ELM 模型的网络结构采用输入层、两层卷积池化层、全连接层、dropout 层和输出层构成,激活函数使用 T-sigmoid 函数。在双组分数据集上,模型的 dropout 率设为 0.2。图 9 所示为双组分中正丁烷和正戊烷的预测结果。可以看出在给定的最优参数条件下,两种物质的预测值与真实值差距不大,正丁烷、正戊烷的 RMSEP 值分别为 $2.92\times10^{-6}$、$3.05\times10^{-6}$,取得了不错的定量预测效果。




(a) 正丁烷

(a) N-butane

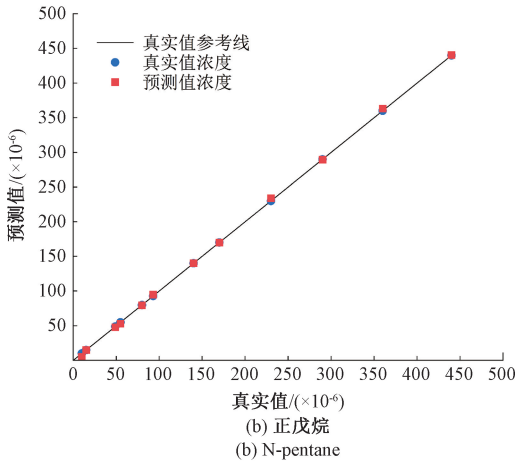



图 9 双组分数据集预测结果

Fig. 9 The predicted results on the two-component dataset

## 4.2 不同数据实验效果

在双组分数据集上确定的模型结构及最优参数,直接作用于 7 组分数据集,其统计结果如表 1 所示。

**表 1 7 组分数据集上模型的 RMSEC、RMSEP、$R^2$ 和 RPD 值($\rho=0.01$, $dropout=0.2$)**

**Table 1 The values of RMSEC, RMSEP, $R^2$, and RPD on the seven-component dataset ($\rho=0.01$, $dropout=0.2$)**

| 物质 | RMSEC/($\times10^{-6}$) | RMSEP/($\times10^{-6}$) | $R^2$ | RPD |
|---|---|---|---|---|
| $C_1$ | 53.93 | 9 009.04 | 0.997 3 | 19.39 |
| $C_2$ | 34.88 | 7 968.46 | 0.997 9 | 21.79 |
| $C_3$ | 36.82 | 7 706.24 | 0.998 2 | 23.44 |
| $nC_4$ | 16.81 | 2 606.20 | 0.292 7 | 1.12 |
| $iC_4$ | 14.79 | 1 742.43 | 0.869 2 | 2.64 |
| $iC_5$ | 7.46 | 1 352.38 | 0.225 5 | 0.96 |
| $CO_2$ | 42.82 | 8 016.48 | 0.997 1 | 18.98 |

可以看出,甲烷、乙烷、丙烷和二氧化碳 4 种物质获得了较好的预测效果,RMSEP 值均低于配气系统的固有误差 10 000×$10^{-6}$(标准气体浓度为 100%,配气系统输出误差为 10 000×$10^{-6}$),而正丁烷、异丁烷和异戊烷的 RMSEP 值分别为 2 517.91×$10^{-6}$、2 136.62×$10^{-6}$、1 555.89×$10^{-6}$,均远高于各自配气系统的输出误差(正丁烷、异丁烷和异戊烷的标准气源浓度分别为 5%、10%、4%,配气系统输出误差分别为 500×$10^{-6}$、1 000×$10^{-6}$、400×$10^{-6}$),预测效果很差。可能的原因为正丁烷、异丁烷和异戊烷 3 种组分的浓度远低于其余 4 种组分的浓度,由于烷烃类气体分子结构的高度相似性,光谱谱峰重叠严重,高浓度气体组分将会对低浓度气体组分造成较强的非线性干扰。

此外,7 种组分的 RMSEC 和 RMSEP 差距过大,出现了明显的过拟合现象,说明物质种类增多和数据复杂性增强,导致模型的参数不再适用。通过改变正则化系数 $\rho$ 与 dropout 率,得到适用于 7 组分数据集的正则化系数 $\rho$ 和 dropout 率分别为 0.000 07 和 0.1,其预测结果统计如表 2 所示。

**表 2 7 组分数据集上模型的 RMSEC、RMSEP、$R^2$ 和 RPD 值($\rho=0.000\ 07$, $dropout=0.1$)**

**Table 2 The values of RMSEC, RMSEP, $R^2$ and RPD on the seven-component dataset ($\rho=0.000\ 07$, $dropout=0.1$)**

| 物质 | RMSEC/($\times10^{-6}$) | RMSEP/($\times10^{-6}$) | $R^2$ | RPD |
|---|---|---|---|---|
| $C_1$ | 3 155.03 | 8 669.89 | 0.997 9 | 19.72 |
| $C_2$ | 2 845.67 | 7 298.28 | 0.997 8 | 21.53 |
| $C_3$ | 2 333.64 | 7 130.95 | 0.998 0 | 22.25 |
| $nC_4$ | 842.52 | 1 753.89 | 0.918 3 | 2.89 |
| $iC_4$ | 746.77 | 1 259.41 | 0.978 0 | 6.21 |
| $iC_5$ | 403.04 | 985.42 | 0.911 8 | 2.78 |
| $CO_2$ | 2 335.83 | 7 668.31 | 0.997 6 | 19.60 |

与表 1 相比,表 2 中 7 种组分的 RMSEP 值均减小,甲烷、乙烷、丙烷、正丁烷异丁烷、异戊烷、异丁烷、异戊烷和二氧化碳的 RMSEP 值分别降低为表 1 中的 96.2%、91.6%、92.5%、67.3%、72.3%、72.9%、95.7%,说明经过参数优化后,模型能更好地用于预测混合气体中各组分的含量。正丁烷、异丁烷和异戊烷的预测效果有明显的提升,但仍高于配气系统的输出误差,说明浓度差异较大的混合气体中,低浓度气体的高精度预测仍然是一项十分艰巨的任务。总体上,本文模型用于预测浓度差异较大的混合气体中的各组分时,其决定系数 $R^2$ 均达到 0.91 以上,能够在同系物的强干扰下,仍有效进行高精度高稳定性的定量分析,可较好地适应复杂录井环境。

## 4.3 不同模型对比实验

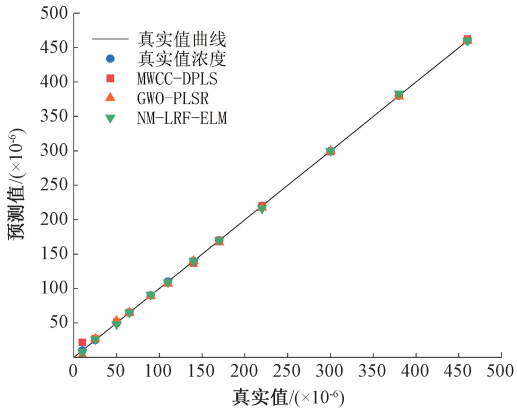
为了进一步评估本文模型的性能,滑动窗口相关系数差分偏最小二乘法(moving window correlation coefficient differences partial least squares, MWCC-DPLS)[22]和灰狼优化偏最小二乘回归算法(grey wolf optimization partial least squares regression algorithm, GWO-PLSR)[23]被用于在浓度差异较小的双组分数据集和浓度差异较大的 7 组分数据集上进行对比实验。其中,MWCC-DPLS 采用滑动窗口算法提取有效波段,再利用改进的偏最小二乘算法实现定量分析;GWO-PLSR 采用灰狼优化算法提取有效特征波长,再结合偏最小二乘回归算法实现定量分析。3 种模型采用均不经过预处理的原始光谱数据进行训练和预测。

表 3 为 3 种模型在双组分数据集上的算法训练时间 $t_c$ 和评价指标。可以看出,3 种模型在双组分数据集上都有较好的定量分析效果,其中,本文模型用于预测正丁烷、正戊烷时,RMSEP 值均低于配气系统输出误差,而 MWCC-DPLS、GWO-PLSR 模型 RMSEP 值在配气系统输出误差为 $5\times10^{-6}\pm0.7\times10^{-6}$ 内,由此可见,通过将一维光谱数据转换为二维矩阵数据用于充分提取光谱的有效信息,并通过引入改进 T-sigmoid 激活函数和 dropout 层,能有效提升模型性能。而 3 种模型的训练时间表现为 NM-LRF-ELM << GWO-PLSR < MWCC-DPLS,本文模型 NM-LRF-ELM 有效利用了局部感受野极限学习机网络算法,一体化实现特征提取及定量分析,可以有效缩短模型训练时间,大幅度缩短录井气体的预测时间。图 10 所示为 3 种模型针对正丁烷和正戊烷的直观预测效果。

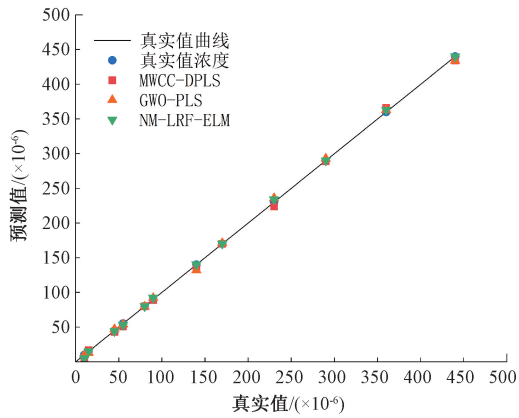
**表 3　3 种模型下双组分的 RMSEC、RMSEP、$R^2$ 和 RPD 的值**

**Table 3　Values of RMSEC, RMSEP, $R^2$, and RPD for the two-components under the three models**

| 物质 | 模型 | $t_c$/s | RMSEC/($\times10^{-6}$) | RMSEP/($\times10^{-6}$) | $R^2$ | RPD |
|---|---|---|---|---|---|---|
| $nC_4$ | MWCC-DPLS[22] | 117.28 | 2.33 | 5.70 | 0.998 4 | 25.07 |
| | GWO-PLSR[23] | 104.59 | 0.66 | 4.30 | 0.999 1 | 33.55 |
| | NM-LRF-ELM | 10.14 | 2.76 | 2.92 | 0.999 6 | 49.84 |
| $nC_5$ | MWCC-DPLS | 112.34 | 2.32 | 5.55 | 0.998 4 | 25.40 |
| | GWO-PLSR | 97.94 | 0.81 | 5.00 | 0.998 7 | 28.21 |
| | NM-LRF-ELM | 8.80 | 3.27 | 3.05 | 0.999 6 | 49.58 |




(a) 正丁烷
(a) N-butane




(b) 正戊烷
(b) N-pentane

图 10　双组分数据集上三种模型的预测效果

Fig. 10　The prediction view of the models on the two-component dataset

表 4 为 3 种模型在 7 组分数据集上的算法训练时间 $t_c$ 和其他评价指标。可以看出,3 种模型在 7 组分数据集的甲烷、乙烷、丙烷和二氧化碳四种物质都取得较好的定量分析效果(RPD 值较大),其中 RMSEC 值与 RMSEP 值均低于配气系统的固有误差 $10\ 000\times10^{-6}$。3 种模型对正丁烷、异丁烷和异戊烷这 3 种物质的预测效果则高于配气系统的输出误差,但本文模型 NM-LRF-ELM 预测效果明显优于 MWCC-DPLS、GWO-PLSR 模型。本文模型在预测甲烷、乙烷、丙烷、正丁烷异丁烷、异戊烷、异丁烷、异戊烷和二氧化碳时,在使用的 5 个评价指标中分别有 4、2、5、5、4、2、3 个为最优,总体优于另外两个模型。值得注意的是,3 种模型在 7 种物质上的训练时间均表现为 NM-LRF-ELM << GWO-PLSR < MWCC-DPLS。

通过对比实验可以看出,当被测混合气体组分较少且浓度跨度较小时,本文模型比对比模型具有明显优越的预测能力;当被测混合气体组分较多且浓度跨度较大时,本文模型与对比模型在高浓度组分上均具有较好的定量分析效果,在低浓度组分上则具有更优秀的分析能力,能更好地应用于组分未知且多样的录井现场环境。在算法的训练时间上,本文模型则具有远小于其他模型的训练时间,可以进一步用于光谱的及时学习模型[24],实现在线模型训练集更新。

### 4.4　标准工业样气测试

本文选择斯伦贝谢石油天然气服务公司的三套标准工业气体混合物用于实验,以从客观的角度评估 NM-LRF-ELM 模型用于实际数据时的预测性能。3 组样品含有异构体异丁烷、异戊烷、正丁烷和正戊烷(样品 C 还

表 4 3 种模型下 7 组分的 RMSEC、RMSEP、$R^2$ 和 RPD 的值

Table 4 Values of RMSEC, RMSEP, $R^2$, and RPD for the seven-components under the three models

| 物质 | 模型 | $t_c$/s | RMSEC/($\times10^{-6}$) | RMSEP/($\times10^{-6}$) | $R^2$ | RPD |
|---|---|---|---|---|---|---|
| $C_1$ | MWCC-DPLS | 205.86 | 5 080.69 | 5 750.22 | 0.996 3 | 18.27 |
| | GWO-PLSR | 214.60 | 8 661.71 | 6 892.97 | 0.953 1 | 16.98 |
| | NM-LRF-ELM | 25.68 | 3 155.03 | 8 669.89 | 0.997 9 | 19.72 |
| $C_2$ | MWCC-DPLS | 232.99 | 5836.01 | 4 917.69 | 0.998 6 | 22.06 |
| | GWO-PLSR | 171.08 | 6586.81 | 7 173.30 | 0.996 3 | 20.90 |
| | NM-LRF-ELM | 27.21 | 2 845.67 | 7 298.28 | 0.997 8 | 21.53 |
| $C_3$ | MWCC-DPLS | 217.45 | 8 402.03 | 8 451.55 | 0.995 8 | 15.72 |
| | GWO-PLSR | 184.69 | 4 970.64 | 8 548.17 | 0.991 9 | 17.52 |
| | NM-LRF-ELM | 26.66 | 2 333.64 | 7 130.95 | 0.998 0 | 22.25 |
| $nC_4$ | MWCC-DPLS | 469.15 | 2 579.44 | 3 325.61 | 0.895 7 | 2.15 |
| | GWO-PLSR | 442.93 | 2 291.70 | 1 973.32 | 0.908 3 | 2.47 |
| | NM-LRF-ELM | 26.60 | 842.52 | 1 753.89 | 0.918 3 | 2.89 |
| $iC_4$ | MWCC-DPLS | 436.84 | 1 758.59 | 1 056.17 | 0.960 4 | 4.61 |
| | GWO-PLSR | 406.35 | 2 573.74 | 2 630.61 | 0.931 5 | 3.93 |
| | NM-LRF-ELM | 23.96 | 746.77 | 1 259.41 | 0.978 0 | 6.21 |
| $iC_5$ | MWCC-DPLS | 483.67 | 773.81 | 921.48 | 0.926 3 | 4.09 |
| | GWO-PLSR | 527.18 | 1 171.55 | 1 310.25 | 0.910 7 | 2.26 |
| | NM-LRF-ELM | 24.13 | 403.04 | 985.42 | 0.911 8 | 2.78 |
| $CO_2$ | MWCC-DPLS | 258.42 | 7 627.89 | 7 576.56 | 0.997 1 | 21.08 |
| | GWO-PLSR | 195.23 | 5 117.08 | 8 853.03 | 0.995 3 | 17.75 |
| | NM-LRF-ELM | 23.32 | 2 335.83 | 7 668.31 | 0.997 6 | 19.60 |

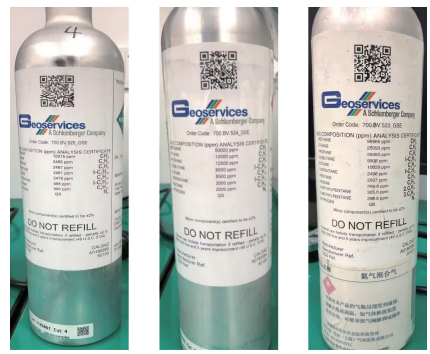
含有 $C_6$,预测时视为未知干扰物质),浓度真实值及基于本文模型的预测值如表 5 所示。

表 5 3 组标准工业气体混合物的测量结果

Table 5 Measurement results for three sets of standard industrial gas mixtures

| 物质 | 样品 A/($\times10^{-6}$) | | 样品 B/($\times10^{-6}$) | | 样品 C/($\times10^{-6}$) | |
|---|---|---|---|---|---|---|
| | 真实值 | 预测值 | 真实值 | 预测值 | 真实值 | 预测值 |
| $C_1$ | 10 015 | 7 690 | 50 000 | 38 296 | 98 966 | 91 991 |
| $C_2$ | 2 495 | 1 938 | 12 500 | 9 276 | 25 093 | 16 981 |
| $C_3$ | 2 487 | 2 170 | 12 500 | 11 194 | 25 063 | 24 479 |
| $nC_4$ | 2 478 | 2 583 | 8 500 | 10 825 | 10 029 | 10 409 |
| $nC_5$ | 995 | 1 094 | 2 000 | 2 487 | 2 507 | 2 668 |
| $C_6$ | | | | | 905.2 | |
| MRE/% | 2.880 9 | | 5.739 8 | | 6.444 6 | |

从表 5 可以看出,甲烷、乙烷、丙烷 3 种物质预测效果较好,均低于样气真实值,在有其他烷烃物质干扰的情况下仍能获得较好分析结果;然而,正丁烷、异戊烷两种物质与在 7 组分数据集上表现相似,无法取得低于真实值的误差,进一步说明在浓度差异较大的真实混合气体中,低浓度气体的高精度预测仍然是一项十分艰巨的任务。




(a) 样品A (b) 样品B (c) 样品C

(a) Sample A (b) Sample B (c) Sample C

图 11 3 组标准工业气体混合物

Fig. 11 Three sets of standard industrial gas mixtures

随着测试样品浓度的增加,MRE 值开始增加,模型的分析精度呈现一定的下降趋势,在存在干扰物质的情况下,样品 B 的准确度下降了近两倍,而样品 C 的准确度仅下降了 0.7%,这是由于浓度超过一定值时红外光谱中会出现饱和吸收现象。对于低于阈值浓度的样品,模型的预测精度受到重叠耦合和饱和吸收的影响,而在浓度高于该阈值时,饱和吸收普遍存在,并且由于光谱特征的重叠耦合导致精度下降占主导地位。需要注意的是,浓度约为 100 000×$10^{-6}$ 的样品 MRE 值仍在 6.5% 以内,在某些应用场景可以认为在合理误差范围内。

## 5　结　　论

针对红外光谱数据中噪声干扰和饱和吸收等原因引起的非线性干扰问题,本文提出了一种 NM-LRF-ELM 模型。该模型将一维光谱数据转换为二维矩阵数据,通过调整网络层数来增强对红外光谱数据的特征提取能力。此外,引入 T-sigmoid 激活函数和 dropout 层以减少过拟合现象,并通过实验验证了该模型的有效性。首先,本文在双组分数据集上确定了最佳激活函数和 dropout 率,并验证了模型在最优参数下的有效性和稳定性;其次,以 7 组分红外光谱数据集为实验对象验证了所提算法的泛化能力;最后,在双组分数据集上与基于滑动窗口类、灰狼优化特征提取的定量分析模型进行对比试验,通过模型训练时间及定量分析模型评价指标证明了本文模型具有高效率、高精度的优势。

本文提出的 NM-LRF-ELM 模型通过局部感受野极限学习机可以有效地将光谱特征提取与定量分析进行一体化处理,避免复杂的特征提取过程,实现非线性干扰下多组分混合气体的高精度快速定量分析。这对于在录井复杂现场环境下提高红外光谱定量分析的精度和效率具有重要的现实意义。进一步提高浓度差异大等因素引起的强干扰条件下微量气体的高精度定量分析,仍然是光谱气体录井技术的主要研究方向。

## 参考文献


[1] 曹博涵. 全球油气勘探开发趋势及对我国油气行业的启示[J]. 当代石油石化, 2023, 31(1):31-37.
CAO B H. Global oil and gas E&P trend and implications to China[J]. Petroleum & Petrochemical Today, 2023, 31(1): 31-37.

[2] ROBERT F, DAVID R, SUN B Q, et al. Novel method for evaluating shale-gas and shale-tight-oil reservoirs using advanced well-log data [J]. SPE Reservoir Evaluation and Engineering, 2018, 22(1): 282-301.

[3] 李国欣, 朱如凯. 中国石油非常规油气发展现状、与关注问题[J]. 中国石油勘探, 2020, 25(2): 1-13.
LI G X, ZHU R K. Progress, challenges and key issues of unconventional oil and gas development of CNPC[J]. China Petroleum Exploration, 2020, 25 (2): 1-13.

[4] 朱华东, 周理, 常宏岗, 等. 激光拉曼光谱法测定天然气组成标准研究[J]. 光谱学与光谱分析, 2018, 38(10): 3286-3294.
ZHU H D, ZHOU L, CHANG H G, et al. Study on standardization of natural gas composition analysis by laser Raman spectroscopy[J]. Spectroscopy and Spectral Analysis, 2018, 38(10): 3286-3294.

[5] STOIL C, KEVIN J, GILLES B, et al. Calibration of short-wave InfraRed (SWIR) hyperspectral imaging using diffuse reflectance infrared fourier transform spectroscopy (DRIFTS) to obtain continuous logging of mineral abundances along sediment cores [J]. Sedimentary Geology, 2021, 428(1):106062.

[6] 荆文峰, 阎荣辉, 陈中普, 等. 红外光谱录井技术在长庆油田的创新应用[J]. 录井工程, 2019, 30(3): 124-130, 189-190.
JING W F, YAN R H, CHEN ZH P, et al. Innovative application of infrared spectrum logging technology in Changqing Oilfield [J]. Mud Logging Engineering, 2019, 30(3): 124-130, 189-190.

[7] LI Z B, PANG W, LIANG H B, et al. Multicomponent Alkane IR measurement system based on dynamic adaptive moving window PLS[J]. IEEE Transactions on Instrumentation and Measurement, 2022, DOI: 10.1109/TIM.2022.3195257.

[8] 韩建, 李雨昭, 曹志民, 等. 原油含水率红外光谱测量的超稀疏表示方法[J]. 仪器仪表学报, 2019, 40(6): 78-85.
HAN J, LI Y ZH, CAO ZH M, et al. Ultra-sparse representation method for measuring crude oil water content using infrared spectroscopy technique [J]. Chinese Journal of Scientific Instrument, 2019, 40(6): 78-85.

[9] 张峰, 汤晓君, 仝昂鑫, 等. 一种基于频率与回归系数相结合的自举柔性收缩变量选择方法[J]. 仪器仪表学报, 2020, 41(1): 64-70.

ZHANG F, TANG X J, TONG A X, et al. Bootstrapping soft shrinkage variable selection method based on the combination of frequency and regression coefficient[J]. Chinese Journal of Scientific Instrument, 2020, 41(1): 64-70.

[10] YANG W Y, WANG W M, ZHANG R Q, et al. A modified moving window partial least squares method by coupling with sampling error profile analysis for variable selection in near infrared spectral analysis[J]. Analytical Sciences: The International Journal of the Japan Society for Analytical Chemistry, 2020, 36(3): 303-309.

[11] SONG L M, GUO S Q, YAN Y G, et al. Quantitative analysis of multicomponent mud logging gas based on infrared spectra [J]. Optoelectronics Letters, 2019, 15(4): 312-316.

[12] 李玉军, 汤晓君, 刘君华. 基于粒子群优化的最小二乘支持向量机在混合气体定量分析中的应用[J]. 光谱学与光谱分析, 2010, 30(3): 774-778.
LI Y J, TANG X J, LIU J H. Application of least square support vector machine based on particle swarm optimization in quantitative analysis of gas mixture[J]. Spectroscopy and Spectral Analysis, 2010, 30(3): 774-778.

[13] BASSET M A, SALLAM K M, MOHAMED R, et al. An improved binary grey-wolf optimizer with simulated annealing for feature selection[J]. IEEE Access, 2021, 9: 139792-139822.

[14] BIAN X H, ZHANG R L, LIU P, et al. Near infrared spectroscopic variable selection by a novel swarm intelligence algorithm for rapid quantification of high order edible blend oil[J]. Spectrochimica Acta. Part A, Molecular and Biomolecular Spectroscopy, 2022, 284: 121788.

[15] SONG L M, WU H, YAN Y G, et al. Application of deep learning in quantitative analysis of the infrared spectrum of logging gas[J]. Applied Optics, 2020, 59(17): E9-E16.

[16] 郝惠敏, 梁永国, 武海彬, 等. 对称点模式-深度卷积神经网络的红外光谱识别方法[J]. 光谱学与光谱分析, 2021, 41(3): 782-788.
HAO H M, LIANG Y G, WU H B, et al. Infrared spectrum recognition method based on symmetrized dot patterns coupled with deep convolutional neural network[J]. Spectroscopy and Spectral Analysis, 2021, 41(3): 782-788.

[17] HUANG G B, BAI Z, et al. Local receptive fields based extreme learning machine [J]. IEEE Computational Intelligence Magazine, 2015, 10(2): 18-29.

[18] XIAO D, LI H Z, SUN X Y. Coal classification method based on improved local receptive field-based extreme learning machine algorithm and visible-infrared spectroscopy[J]. ACS Omega, 2020, 5(40): 25772-25783.

[19] ZHANG Z Y, JIANG M Q, XIONG H M. Optimized identification of cheese products based on Raman spectroscopy and an extreme learning machine[J]. New Journal of Chemistry, 2023, 47(14): 6889-6894.

[20] SUN H X, ZHANG S J, REN R, et al. Detection of soluble solids content in different cultivated fresh jujubes based on variable optimization and model update[J]. Foods, 2022, 11(16): 2522.

[21] MORAIS C L M, SANTOS M C D, LIMA K M G, et al. Improving data splitting for classification applications in spectrochemical analyses employing a random-mutation Kennard-Stone algorithm approach [J]. Bioinformatics (Oxford, England), 2019, 24: 5257-5263.

[22] WANG H P, CHU X L, CHEN P, et al. Moving window correlation coefficient differences partial least squares (MWCC-DPLS) quantitative calibration method based on spectral differences between calibration samples: Application to the fast determination of gasoline octane number with near-infrared spectroscopy [J]. Fuel Processing Technology, 2023, 240: 107583.

[23] HOU Y Z, GAO X, LI S N, et al. Variable selection based on gray wolf optimization algorithm for the prediction of saponin contents in Xuesaitong dropping pills using NIR spectroscopy [J]. Journal of Pharmaceutical Innovation, 2022, 18(1): 43-59.

[24] 施锦涛, 陈磊, 秦凯, 等. 基于时空相似性的即时学习在线建模[J]. 仪器仪表学报, 2022, 43(6): 185-193.
SHI J T, CHEN L, QIN K, et al. Online modeling of just-in-time learning based on spatial-temporal similarity[J]. Chinese Journal of Scientific Instrument,

2022, 43(6): 185-193.

## 作者简介

**李忠兵**(通信作者),2009 年于武汉大学获得学士学位,2014 年于武汉大学获得博士学位,现为西南石油大学讲师,主要研究方向为图像处理技术、机器学习和光谱分析。

E-mail: lzb@ swpu. edu. cn

**Li Zhongbing** (Corresponding author) received his B. Sc. degree from Wuhan University in 2009, and Ph. D. degree from Wuhan University in 2014. He is currently a lecturer at Southwest Petroleum University. His main research interests include image processing technology, machine learning, and spectrum analysis.

**梁海波**,2000 于西南石油大学获得学士学位,2003 于西南石油大学获得硕士学位,2008 于西南石油大学获得博士学位,现为西南石油大学教授,主要研究方向为油气井随钻测量新装备、智能测控系统、油气地质导向分析和信息交互分析。

E-mail: secondbo@ 126. com

**Liang Haibo** received his B. Sc. degree from Southwest Petroleum University in 2000, M. Sc. degree from Southwest Petroleum University in 2003, and Ph. D. degree from Southwest Petroleum University in 2008. He is currently a professor at Southwest Petroleum University. His main research interests include new equipment for measuring oil and gas wells while drilling, intelligent measurement and control system, oil and gas geological guidance analysis, and information exchange analysis.