



硕士学位论文
MASTER DISSERTATION

气固两相流中气体组分在线测量方法
研究

Research on On-line Measurement of Gas
Components in Gas-Solid Two-phase Flow

作者 沈永

校内导师 郭天太 副教授

校外导师 赵军 高工

专业领域 仪器仪表工程

中国计量大学

二〇一七年六月

Research on On-line Measurement of Gas
Components in Gas-Solid Two-phase Flow

By
Yong Shen

A Dissertation Submitted to
China Jiliang University
In partial fulfillment of the requirement
For the degree of
Master of Engineering

China Jiliang University

June, 2017

独创性声明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作和取得的研究成果，除了文中特别加以标注和致谢之处外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得中国计量大学或其他教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名：

签字日期： 年 月 日

学位论文版权使用授权书

本学位论文作者完全了解 中国计量大学 有关保留、使用学位论文的规定。特授权 中国计量大学 可以将学位论文的全部或部分内容编入有关数据库进行检索，并采用影印、缩印或扫描等复制手段保存、汇编以供查阅和借阅。同意学校向国家有关部门或机构送交论文的复印件和磁盘。（保密的学位论文在解密后适用本授权说明）

学位论文作者签名：

导师签名：

签字日期： 年 月 日

签字日期： 年 月 日

中图分类号 O657.33

学校代码 10356

UDC 661.7

密级 公开

中國計量大學

CHINA JILIANG UNIVERSITY

硕士学位论文

MASTER DISSERTATION

气固两相流中气体组分在线测量方法
研究

Research on On-line Measurement of Gas
Components in Gas-Solid Two-phase Flow

作者 沈 永

专业领域 仪器仪表工程

校内导师 郭天太 副教授

校外导师 赵军 高工

申请学位 工学硕士

论文类型 应用研究

二〇一七年六月

致 谢

首先感谢郭天太和孔明老师，本文的研究工作是在两位老师的悉心指导下完成的，他们的渊博知识让我感到知识的浩瀚，严谨的治学态度和认真的学术精神是我需要永远学习的榜样。

三年的研究生生活，过的很快，这里真的需要谢谢郭老师，不仅在科研学习上，更是在做人做事读书方面。以前的我不喜欢读书，记得在研一第一次班级聚会上，郭老师就问我有没有读过四大名著，真的很惭愧那个时候真的没读过，后面在郭老师的影响下逐渐喜欢上读书，做人做事和读书是一辈子的事，这也是我在研究生生涯中学到的最宝贵的东西。研究生期间有过迷茫和不知所措，郭老师能够从历史的角度劝导我，让我知道看待很多问题格局要大，眼光要远，要乐观。还记得研三第一学期开学和郭老师一起写国基金申请书的日子，一起熬夜赶写和改写，真的很让人难忘，特别是郭老师认真的态度，对于一个标点都要严格要求，实实在在的事，影响现在的我，也将影响我的将来。

同时，也感谢我们导师组的赵军老师、王道档老师和刘维老师，赵老师的风趣幽默，专业知识渊博，关心关爱学生，给我们创造了和谐快乐的学习氛围，衷心祝愿我们导师组的每个导师都能桃李满天下。

读研的经历让我的人生丰富多彩，收获了知识，收获了友谊，感谢和我一起学习的的小伙伴们，我的师兄师姐洪博、杨遥、吴超、叶婷、丁超、陈茜茜、杨德志、徐寅、王英军，不管是在学习上还是生活方面，各位师兄师姐总能给我很多帮助。感谢我的同窗，沈海栋、陈栋、路首创、吕晨焕、贾茜媛、黄伟、贺楚红和徐杨波。和你们在一块共同学习了三年，玩耍了三年，三年瞬间而逝，希望你们以后工作顺利。

在这里，还要感谢我的父母，谢谢你们对我这么多年来培养，默默为我付出，是你们的支持和鼓励，让我感到家的温暖，谢谢你们。

最后感谢各位专家在百忙之中抽出时间评阅我的论文，你们的意见将为我研究生学习的道路画上圆满的句号。请各位专家不吝赐教，批评指正。

沈永

2017年6月

气固两相流中气体组分在线测量方法研究

摘要：在矿井中，两相流中的两相即易燃易爆气体和煤粉颗粒给井下工作人员的生命健康和安全带来极大威胁。因此，排除井下空气中煤粉颗粒的干扰实现高效、快速以及实时的矿井气体在线检测，对于井下作业人员的安全和防止瓦斯爆炸事故的发生都有重要意义。本文模拟矿井中气固两相流的真实环境，搭建了气体红外光谱数据的采集系统，采集了五种极性矿井气体的红外光谱数据，包括二氧化碳（CO₂）、甲烷（CH₄）、乙烷（C₂H₆）、丙烷（C₃H₈）和丁烷（C₄H₁₀），利用小波变换对采集到的原始光谱数据进行降噪处理，然后采用两种特征提取方法即峰值提取法（PEM）和主成分分析法（PCA）对降噪后的数据进行特征提取和降维，然后建立支持向量机（SVM）及其改进模型和极限学习机（ELM）及其改进模型，以甲烷（CH₄）为例，采用不同参数寻优方法，对 SVM 模型的预测精度进行分析，确定利用 PSO 算法对 SVM 的惩罚参数 c 和核函数参数 g 进行寻优，建立基于 PSO 算法的改进 PEM-SVM 模型和改进 PCA-SVM 模型。结果表明，在预测 C₂H₆、C₃H₈ 和 C₄H₁₀ 三种低浓度气体时，改进 PEM-SVM 模型的预测精度较高，在预测 CH₄ 和 CO₂ 两种高浓度气体时，改进 PCA-SVM 模型预测的精度较高，最后整合两个模型的优点，建立了 PEM-PCA-SVM 集成模型。同时，也建立了基于 D-ELM 的矿井下气固两相流中气体组分的定量分析模型，即 PEM-D-ELM 模型和 PCA-D-ELM 模型，并整合了两个模型的优势，建立了集成 PEM-PCA-ELM 模型。使用测试集样本对 PEM-PCA-ELM 模型进行测试验证并和 PEM-PCA-SVM 模型进行比较，发现 PEM-PCA-ELM 建模速度较快，但在预测精度方面不及 PEM-PCA-SVM 模型。

关键词：D-ELM；定量分析；红外光谱；PSO-SVM；气固两相流

分类号：中图分类号 O657.33；UDC 661.7

Research on On-line Measurement of Gas Components in Gas-Solid Two-phase Flow

Abstract: Two-phase flow in the coal mine includes flammable and explosive gas and coal particles, brings a great threat to workers' health and lives. Therefore, eliminating the interference of coal particles is important for high efficient, rapid and real-time on-line detection of mine gas, also it is significant for the safety of mine workers and can avoid gas explosion accidents. In this thesis, simulating the real environment of mine, and an experiment system was set up to collect the gas spectra of methane(CH_4), ethane(C_2H_6), propane(C_3H_8), butane(C_4H_{10}) and carbon dioxide(CO_2). The original spectral data is denoised by wavelet transform, and then the feature extraction and dimension reduction of the data after noise reduction are carried out by two kinds of feature extraction methods: peak extraction method (PEM) and principal component analysis (PCA). Then, a support vector machine (SVM) and its improved model and extreme learning machine (ELM) and its improved model were established to quantitatively analyze the mixed gas in the gas-solid two-phase flow in the mine, Taking the methane (CH_4) as an example, the prediction accuracy of the SVM model was analyzed by different parameter optimization methods. The PSO algorithm was used to optimize the penalty parameter c and the kernel function parameter g of the SVM algorithm. PEM-SVM model and improved PCA-SVM model was established based on the PSO algorithm. The prediction accuracy of the improved PEM-SVM model is high when the three kinds of low concentration gas of C_2H_6 , C_3H_8 and C_4H_{10} are predicted. The accuracy of the improved PCA-SVM model is high when the high concentration of CH_4 and CO_2 are predicted. Integrating the advantages of the two models, the PEM-PCA-SVM integrated model was built, And in order to compare with PEM-PCA-SVM model, establishing PEM-PCA-ELM model, which integrated the advantages of PEM- D-ELM model and PCA-D-ELM model. The PEM-PCA-ELM model was tested and verified by using the test set samples, the results show that: The PEM-PCA-ELM model is found to be faster, but the prediction accuracy of

PEM-PCA-SVM model is not as good as PEM-PCA-SVM model.

Keywords: D-ELM; quantitative analysis; infrared spectrum; SVM; gas-solid two - phase flow

Classification: O657.33; UDC 661.

目 次

摘要	I
目 次	IV
图清单	VI
表清单	VIII
1 绪论	1
1.1 研究的背景及意义	1
1.2 基于红外光谱法气体定量分析的国内外研究现状	2
1.3 论文主要结构及安排	4
1.4 本章小结	6
2 红外光谱分析技术原理	7
2.1 红外光谱的产生和划分红外光谱区间	7
2.2 傅里叶变换红外光谱仪的工作原理及结构	8
2.3 红外光谱定量分析技术	10
2.3.1 光谱数据预处理	10
2.3.2 定量分析理论	13
2.3.3 模型的验证与评价	22
2.3 本章小结	23
3 气固两相流中气体光谱数据采集	24
3.1 配气装置的研发	24
3.2 实验系统的搭建	31
3.3 实验样本	34
3.4 本章小结	37
4 基于支持向量机的定量分析	38
4.1 光谱数据处理	38
4.2 基于 SVM 气固两相流中气体组分的定量分析	40
4.2.1 基于 PEM-SVM 气固两相流中气体组分的定量分析	40
4.2.2 基于 PCA-SVM 气固两相流中气体组分的定量分析	42
4.3 基于改进 SVM 气固两相流中气体组分的定量分析	43

4.3.1 SVM 参数寻优.....	43
4.3.2 基于改进 PEM-SVM 气固两相流中气体的定量分析	46
4.3.3 基于改进 PCA-SVM 气固两相流中气体的定量分析	50
4.4 结果分析与讨论	54
4.5 本章小结	55
5 基于极限学习机的定量分析	56
5.1 基于 ELM 气固两相流中气体组分的定量分析	56
5.2 基于 D-ELM 气固两相流中气体组分的定量分析	57
5.2.1 基于 PEM-D-ELM 气固两相流中气体组分的定量分析	57
5.2.2 基于 PCA-D-ELM 气固两相流中气体组分的定量分析.....	60
5.4 结果分析与讨论	62
5.5 本章小结	64
6 总结与展望	65
6.1 总结全文	65
6.2 工作展望	66
参考文献	67
作者简介	71

图清单

图 2.1 迈克尔逊干涉仪示意图	9
图 2.2 Mallat 算法信号分解与合成图	12
图 2.3 光谱定量分析流程	13
图 2.4 支持向量机网络结构	16
图 2.5 SVM 分类超平面示意图	17
图 2.6 ELM 网络结构图	20
图 2.7 D-ELM 算法流程	21
图 3.1 5 路高精度配气装置	24
图 3.2 5 路高精度配气内部结构图	25
图 3.3 气体输入输出接口图	25
图 3.4 接口堵头	26
图 3.5 钢瓶开关图	26
图 3.6 浮子流量计旋钮图	27
图 3.7 装置前面板视图	27
图 3.8 启动界面	28
图 3.9 手动模式界面	28
图 3.10 自动模式—参数设置界面	29
图 3.11 操作流程圖	29
图 3.12 自动模式—运行界面	31
图 3.13 数据采集系统	32
图 3.14 数据采集系统实物图	32
图 3.15 比色皿	33
图 3.16 采集的气体光谱图	34
图 3.17 五种单组份气体的红外光谱图	35
图 4.1 建模流程图	38
图 4.2 原始光谱图	40
图 4.3 小波降噪后的光谱图	40
图 4.4 PEM-SVM 的原理	41

图 4.5 PCA-SVM 原理图.....	43
图 4.6 PSO 寻优流程图.....	44
图 4.7 GA 寻优流程图.....	45
图 4.8 PSO 算法的参数优化结果.....	45
图 4.9 GA 算法的参数优化结果.....	46
图 4.10 CH ₄ 测试集改进 PEM-SVM 的预测结果.....	48
图 4.11 C ₂ H ₆ 测试集改进 PEM-SVM 的预测结果.....	48
图 4.12 C ₃ H ₈ 测试集改进 PEM-SVM 的预测结果.....	48
图 4.13 C ₄ H ₁₀ 测试集改进 PEM-SVM 的预测结果.....	49
图 4.14 CO ₂ 测试集改进 PEM-SVM 的预测结果.....	49
图 4.15 PCA 的降维结果.....	51
图 4.16 CH ₄ 测试集改进 PCA-SVM 的预测结果.....	52
图 4.17 C ₂ H ₆ 测试集改进 PCA-SVM 的预测结果.....	52
图 4.18 C ₃ H ₈ 测试集改进 PCA-SVM 的预测结果.....	52
图 4.19 C ₄ H ₁₀ 测试集改进 PCA-SVM 的预测结果.....	53
图 4.20 CO ₂ 测试集改进 PCA-SVM 的预测结果.....	53
图 5.1 CH ₄ 测试集 PEM-D-ELM 的预测结果.....	57
图 5.2 C ₂ H ₆ 测试集 PEM-D-ELM 的预测结果.....	58
图 5.3 C ₃ H ₈ 测试集 PEM-D-ELM 的预测结果.....	58
图 5.4 C ₄ H ₁₀ 测试集 PEM-D-ELM 的预测结果.....	58
图 5.5 CO ₂ 测试集 PEM-D-ELM 的预测结果.....	59
图 5.6 CH ₄ 测试集 PCA-D-ELM 的预测结果.....	60
图 5.7 C ₂ H ₆ 测试集 PCA-D-ELM 的预测结果.....	61
图 5.8 C ₃ H ₈ 测试集 PCA-D-ELM 的预测结果.....	61
图 5.9 C ₄ H ₁₀ 测试集 PCA-D-ELM 的预测结果.....	61
图 5.10 CO ₂ 测试集 PCA-D-ELM 的预测结果.....	62

表清单

表 2.1 不同红外区对应的波长和波数	8
表 2.2 支持向量机常用核函数	19
表 3.1 详细操作说明表	30
表 3.2 每种气体的浓度范围	35
表 3.3 各气体组份特征峰吸收范围/cm ⁻¹	36
表 3.4 部分多组分气体浓度配置	36
表 4.1 四种小波去噪结果	39
表 4.2 特征提取波段及个数	41
表 4.3 GA 和 PSO 寻优结果	46
表 4.4 改进 PEM-SVM 模型结果	47
表 4.5 两种模型结果比较	47
表 4.6 改进 PEM-SVM 模型对五种混合气体的预测结果	50
表 4.7 改进 PCA-SVM 模型结果	51
表 4.8 改进 PCA-SVM 模型对五种混合气体的预测结果	54
表 4.9 PEM-PCA-SVM 模型对五种混合气体的预测结果	55
表 5.1 ELM 模型对五种混合气体的预测结果	56
表 5.2 PEM-D-ELM 模型对五种混合气体的预测结果	59
表 5.3 PCA-D-ELM 模型对五种混合气体的预测结果	62
表 5.4 PEM-PCA-ELM 模型对五种混合气体的预测结果	63
表 5.5 PEM-PCA-SVM 和 PEM-PCA-ELM 模型结果比较	64

1 绪论

1.1 研究的背景及意义

我国煤炭资源储量丰富与长期的旺盛需求，导致煤炭依旧是我国主要的能源资源和工业原料，其在我国一次能源生产和消费结构中比重分别高达 80% 和 70% 左右^[1]。以“十二五”时期为例，2010 年全国煤炭产量为 32.4 亿吨左右，一次煤炭消费总量为 32.5 亿吨左右；2015 年全国煤炭产量为 41 亿吨左右，年均增长率为 4.9%，一次煤炭消费总量为 40 亿吨左右，年均增长率为 4.3% 左右。可以看出在未来的数十年内，煤炭仍将长期作为我国不可替代的常规能源基础。

然而，我国煤炭产量的 95% 以上是人工开采的，矿井下的生产者普遍受到瓦斯爆炸和煤自燃等灾害的严重威胁。2005 至 2014 年，不管在事故死亡人数上还是事故起数上，由于瓦斯爆炸引起的死亡人数和事故起数都是最多的，分别占 64.05% 和 50.83%^[2]。虽然“十二五”期间，全国煤矿事故平均死亡人数和重特大事故年均起数分别比“十一五”期间下降 64.4% 和 53.3%，但煤矿事故仍然频发，煤矿生产的安全形势依然十分严峻。2015 年 12 月 16 日，黑龙江鹤岗市向阳区煤矿发生瓦斯爆炸事故，造成 19 人死亡。2016 年 12 月 3 日，内蒙古赤峰市一煤矿发生瓦斯爆炸事故，造成 32 人遇难。

矿井下的气固两相流中的两相即为易燃易爆的有毒气体和空气中的煤粉颗粒，易燃易爆的有毒气体的主要成分是甲烷（ CH_4 ）、乙烷（ C_2H_6 ）等瓦斯气体。煤层是一个典型的裂隙—空隙系统，瓦斯主要以游离态和吸附态存在于裂隙—空隙内，在开采过程中，这些游离态和吸附态的瓦斯气体就会从煤层中释放出来，当达到一定浓度时，如存在高温火源和充足的氧气，就会发生瓦斯爆炸。瓦斯爆炸的浓度界限为 5%~16%，低于 5% 时，通常不会发生爆炸事故；当瓦斯浓度在 9.5% 左右时，如遇明火，爆炸威力巨大；当瓦斯浓度高于 16% 时，通常不会发生爆炸，但遇明火仍会燃烧，后期仍可能引起爆炸^[3-6]。因此，排除空气中煤粉颗粒的干扰，实现高效、快速以及实时的矿井气体在线检测，对于井下作业人员的安全和防止瓦斯爆炸事故的发生都有重要意义。

目前，应用最广的气体检测器是热导元件和载体催化元件相结合的催化燃烧式气体传感器，其原理是当待测的矿井气体接触传感器时，传感器上的催化剂会和待测气体发生反应产生热量，使得感应电阻的阻值发生变化，通过电信号的变化量和对应气体浓度进行标定，测得待测气体的浓度^[7-9]。催化燃烧式传

传感器的优点是响应速度快、测量范围大。缺点是由于使用催化剂，传感器的活性会受到硫化物的影响，使得其灵敏度下降，即抗干扰能力差，测量误差大，且稳定性不好，维护起来比较困难，寿命仅 1~2 年^[10]。相比于催化燃烧式传感器，红外气体传感器以其稳定性好、选择性强、灵敏度高、扫描速度快、测量精度高等特点受到国内外的高度重视^[11-15]。由于不同气体只吸收特定波段的红外光谱，并且吸收峰不同，所以测定某种气体时，很少会受到其他气体的干扰。根据郎伯—比尔定律^[16-20]，气体对红外光谱能量的吸收量和其浓度有关，所以可通过气体吸收红外光谱的波段对其进行定性分析，通过吸收红外光谱能量的多少对其进行定量分析。本文主要用红外光谱法对矿井下气固两相流中气体组分进行定性和定量分析。

1.2 基于红外光谱法对气体定量分析的国内外研究现状

由于红外光谱法检测技术的原理及其众多优点，气体红外检测技术得到了广泛研究，取得了不错的成绩。国外的红外气体检测技术比国内早，在硬件研究方面比国内有很大优势，主要应用于空气监测和汽车尾气检测^[21-22]。由于国内矿井众多，空气污染比较严重，红外检测器的市场非常大，导致红外气体检测技术在近几年得到长足发展。

(1) 国外研究现状

1983 年，H. Inaba 等^[23]利用光谱吸收式气体检测器在近红外波段范围内，对甲烷气体浓度进行检测，检测下限达到 2000×10^{-6} 。

2005 年，N. Makhoukhi 等^[24]基于红外光谱法的偏最小二乘法模型对天然气中的主要气体成分进行了定量分析。结果表明，虽然该方法的预测精度相比色谱分析结果没有太大提高，但检测速度有明显优势。

2007 年，英国城市技术有限公司推出了 IRcel 电子红外系列气体传感器，可测 $0 \sim 1000 \times 10^{-6}$ 范围内的二氧化碳和 $0 \sim 100\%$ 范围内的甲烷。在一定温度范围内输出稳定，抗损能力强，寿命长，无需校准。

2010 年，Zakia Sebih 等^[25]针对汽车尾气污染环境等问题，设计了一种基于傅里叶变换的红外光谱仪器，可用于定量分析多种空气污染物系统。该系统的最小检测分辨率可达到 80×10^{-6} 。

2011 年，Anna A. Stec 等^[26]利用 FTIR 技术对火灾中的有毒气体检测做了深入研究，对火灾产生的气体进行了定性和定量分析，然后分析了压强、温度等因素对结果的影响。

2013 年, Ritobrata Sur 等^[27]使用 TDLAS 检测技术, 利用不同波长的激光器, 在高粉尘、高噪声的环境下对甲烷、二氧化碳、一氧化碳和水汽进行了同时检测。

2015 年, P. Andrea 等^[28]使用 DFB 和 VCSEL 激光器, 利用直接吸收光谱技术对水汽的多根吸收谱线进行了检测, 分析多根谱线对水汽的吸收强度。

(2) 国内研究现状

2006 年, 中国科学院安徽光学精密机械研究所的孙桂玲等^[32]采用小波分析方法^[29-31]对特征波段进行分解, 提取分解后高频分量中的有效信息, 同时去除冗余无用的信息, 根据局部相邻正负极值确定原始光谱曲线上每个吸收带的区间范围, 通过局部过零点计算出各个吸收带的中心波长。实验证明, 该方法在光谱特征提取方面比传统方法更简洁有效。

2007 年, 西安交通大学的白鹏等^[35,36]提出了基于支持向量机^[33,34]分析混合气体红外光谱的新方法。由于混合气体各种成分在相同的波段吸收红外光谱能量, 导致光谱曲线发生严重重叠, 该方法利用核函数将光谱数据在高维空间进行变换, 建立支持向量机回归预测模型, 对混合气体进行预测和分析。结果表明, 经过对光谱数据的特征提取, 选择合适的核函数和各个参数, 实验结果的最大平均绝对误差为 0.132%; 混合气体识别的准确率高于 94%。2008 年, 他们在此基础上提出了分布模式的红外光谱在线识别方法。研究确定了真实环境中可能出现的 60 种混合气体分布模式。实验结果表明, 该方法可以在小样本的条件下对混合气体进行分布模式的在线识别, 识别精度为 98.8%。

2009 年, 郝慧敏等^[39]提出了一种基于核偏最小二乘^[37,38]结合支持向量机的红外光谱分析方法。首先利用核偏最小二乘方法对甲烷(CH_4)、乙烷(C_2H_6)、丙烷(C_3H_8)、异丁烷($\text{iso-C}_4\text{H}_{10}$)、正丁烷($\text{n-C}_4\text{H}_{10}$)等七种严重交叉重叠的红外特征吸收谱线进行特征提取, 然后将提取到的特征数据作为支持向量机模型的输入, 建立预测模型。结果表明, 核偏最小二乘法能够有效提取混合气体的 FTIR 光谱数据特征, 降低数据维度, 减少数据噪声, 同时模型的预测精度有所提高, 预测时间也大幅减少。

2010 年, 中国科学技术大学高明亮等^[43]搭建 FTIR 混合气体分析系统, 对 8 种低浓度有毒有害气体进行检测, 通过交叉组合建立三种模型: 多项式偏最小二乘法、偏最小二乘结合 BP 神经网络法^[40-42]和最小二乘结合支持向量机法。并与传统方法进行比较, 同时分析各个模型本身的性能和对气体浓度预测的精度。实验结果表明, 偏最小二乘结合 BP 神经网络模型的性能最佳。

2011 年,杜涛等^[44]基于 FTIR 技术对矿井下 11 种极性混合气体进行定量分析研究,利用偏最小二乘法对光谱数据进行降维,同时结合化学计量学对光谱数据进行特征分析,通过粒子群优化算法^[45-47]寻找支持向量机的最优参数。之后在模型中引进非线性校正算法,使得模型更加稳定,同时也从误差角度进行了分析,提高实验的可靠性。

2012 年,李玉军等^[48]提出 DOSC 算法对烷烃类混合气体进行定量分析。他们针对烷烃类多组分气体的红外光谱的基线漂移问题,提出了利用直接正交信号校正算法对红外光谱数据进行预处理,并和导数算法进行对比,采用偏最小二乘建立预测模型,实验结果表明,DOSC 算法对红外光谱基线的校正效果最佳。

2013 年,张爱菊^[49]提出基于数据挖掘技术对瓦斯气体的红外光谱数据进行分析研究。她在处理红外光谱数据时发现,增加 K 平均非分层聚类分析结合偏最小二乘算法比单纯采用最小二乘法处理数据在精度上明显占有优势。结果表明,采用 K-means^[50-52]聚类方法对数据进行预处理对预测的精度有明显提高。

2014 年,西安交通大学赵安新等^[53]基于 2DCOS 对混合气体 FTIR 中同分异构体谱峰进行了识别。他们采用广义二维相关光谱(2DCOS)和 FTIR 光谱技术对多组分烃类气体中的同分异构体进行辨别,解决了气体吸收峰交叉重叠、红外光谱特征相似和特征变量选择困难等问题。实验结果初步表明了,2DCOS 在多组分烃类气体 FTIR 定量分析中光谱分辨率得到加强,使预测精度得到提高。

2015 年,陈媛媛等^[54]提出基于思维进化蝙蝠算法处理混合气体各成分吸收峰相互交叉重叠的光谱特征选择困难的问题,结合浓度反演模型对混合气体进行定量分析。其通过基于特征波长点建立浓度反演模型,测试集的均方误差可达 0.155。结果表明,思维进化蝙蝠算法全局寻优能力强,算法的收敛速度快,预测精度高。

1.3 论文主要结构及安排

近年来,红外光谱分析技术在很多领域中得到越来越多的应用,也给混合气体的定量研究带来了新的契机。FTIR 技术目前在固体和液体领域应用比较多,也更成熟。而由于气体的特殊性质,其检测精度很容易受到温度、压强等噪声的干扰,气体在红外领域的更深入研究目前还有一定难度。本文模拟矿井中气固两相流的真实环境,搭建了气体红外光谱数据的采集系统,采集了五种

极性矿井气体的红外光谱数据,包括二氧化碳(CO_2)、甲烷(CH_4)、乙烷(C_2H_6)、丙烷(C_3H_8)和丁烷(C_4H_{10}),然后建立支持向量机(SVM)及其改进模型和极限学习机(ELM)及其改进模型,对矿井下气固两相流中的混合气体进行定量分析研究,并对各种模型的结果进行比较分析,提出了一种基于 FTIR 技术的矿井下气固两相流中气体组分的在线测量方法。

本文的结构安排如下:

第一章,主要介绍本文的研究背景和目前主流的气体检测方法以及其各自的优点和缺点。论述了在气体检测领域,红外光谱技术的国内外发展现状、研究重点和未来发展趋势,同时介绍了本文的研究目的和研究意义。

第二章,主要介绍红外光谱分析技术的基本原理,包括红外光谱的产生及如何划分红外光谱、傅里叶红外光谱仪的基本原理和红外光谱定量分析技术。阐述了红外光谱应用于多组分混合气体定量分析的相关理论、光谱数据预处理方法和各种模型的原理和评价标准。

第三章,详细介绍气固两相流中气体组分红外光谱数据采集系统,以及其中最为重要的配气装置的研发,设计了合理的配气方案。根据本文的研究目标研发了简单、易操作的配气装置,可以根据配气方案精确地计量输入到气体池中气体的量,通过红外光谱仪得到每个样本的红外光谱图,分析每种气体的特征吸收峰、吸收波段等,为气体定量分析模型提供可靠的实验数据。

第四章,利用小波变换对原始光谱数据进行减噪处理,然后采用两种特征提取方法即峰值提取法(PEM)和主成分分析法(PCA)对降噪后的数据进行特征提取和降维,最后建立基于 SVM 的矿井下气固两相流中气体组分的定量分析模型。以甲烷(CH_4)为例,采用不同参数寻优方法,对 SVM 模型的预测精度进行分析,确定利用 PSO 算法对 SVM 的惩罚参数 c 和核函数参数 g 进行寻优,建立基于 PSO 算法的改进 PEM-SVM 模型和改进 PCA-SVM 模型。结果表明,在预测 C_2H_6 、 C_3H_8 和 C_4H_{10} 三种低浓度气体时,改进 PEM-SVM 模型的预测精度较高,在预测 CH_4 和 CO_2 两种高浓度气体时,改进 PCA-SVM 模型预测的精度较高,最后整合两个模型的优点,建立了 PEM-PCA-SVM 集成模型。

第五章,利用极限学习机(ELM)对矿井下气固两相流中气体组分进行定量分析,采用峰值提取法(PEM)和主成分分析法(PCA)对小波降噪后的数据进行特征提取和降维处理,提高了建模效率,并对极限学习机算法做出了改进,建立了基于 D-ELM 的矿井下气固两相流中气体组分的定量分析模型,即 PEM-D-ELM 模型和 PCA-D-ELM 模型,并整合了两个模型的优势,建立了集

成 PEM-PCA-ELM 模型。使用测试集样本对 PEM-PCA-ELM 模型进行测试验证并和 PEM-PCA-SVM 模型进行比较,发现 PEM-PCA-ELM 建模速度较快,但在预测精度方面不及 PEM-PCA-SVM 模型。

1.4 本章小结

本章主要介绍了本文的研究背景和目前主流的检测气体的方法以及本文的研究目的和研究意义。论述了在气体检测领域,红外光谱技术的国内外发展现状、研究重点和未来发展趋势,同时说明了本文的结构和安排。

2 红外光谱分析技术原理

2.1 红外光谱的产生和红外光谱区间的划分

采用傅里叶变换红外（Fourier transform infrared, FTIR）光谱仪测定样品的红外光谱时，使用的红外光源是连续波长的光源。连续波长光源照射红外样品后，样品中的分子会吸收某些波长的光。没有被吸收的光到达检测器，检测器将检测到的光信号经过模/数转换，再经过傅里叶变换，即可得到样品的单光束光谱。为了得到样品的红外光谱，需要从样品的单光束光谱中扣除掉背景的光谱光束，也就是需要测试红外光不经过样品的情况下得到的背景单光束光谱。这样得到的背景单光束光谱中包含了仪器内部各种零部件和空气的信息。在测试样品的单光束和测试背景的单光束光谱是，这些信息是完全相同的。所以，样品的单光束光谱中扣除背景的单光束光谱后就得到了样品的红外投射光谱。在红外光谱中，在被吸收的光的波长或波数位置会出现吸收峰。某一波长的光被吸收得越多，透射率就越低，吸收峰就越强。当样品分子吸收很多种波长的光时，在测得的红外光谱中就会出现许多吸收峰。

红外透射光谱的纵坐标有两种表示方法，即透射率 T （%，Transmittance）和吸光度 A （Absorbance）。纵坐标采用透射率 T 表示的光谱称为透射率光谱，纵坐标采用吸光度 A 表示的光谱称为吸光度光谱。

某一波长（或波数）光的透射率 $T_{(\nu)}$ 是红外光透过样品后的光强 $I_{(\nu)}$ 和红外光透过背景（通常是空光路）的光强 $I_{0(\nu)}$ 的比值。即

$$T_{(\nu)} = \frac{I_{(\nu)}}{I_{0(\nu)}} \times 100\% \quad (2-1)$$

某一波长（或波数）光的吸收强度即吸光度 $A_{(\nu)}$ 是透射率 $T_{(\nu)}$ 倒数的对数，即

$$A_{(\nu)} = \lg \frac{1}{T_{(\nu)}} \quad (2-2)$$

透射率光谱和吸光度光谱之间可以相互转换。透射率光谱虽然能直观地看出样品对不同波长红外光的吸收情况，但是透射率光谱的透射率与样品的含量不成正比关系，即透射率光谱不能用于红外光谱的定量分析。要进行定量分析，必须将透射率转换成吸光度光谱。根据朗伯—比尔定律（Lambert-Beer's Law）

可知，吸光度光谱的吸光度值 A 在一定范围内与样品厚度和样品浓度成正比关系，即吸光度光谱能用于红外光谱的定量分析，所以现在的红外光谱图大都采用吸光度光谱表示。

光谱图的横坐标通常采用波数 (cm^{-1}) 表示，也可以采用波长 (μm) 或 (nm) 表示。

红外光谱通常分为三个区域，即近红外区、中红外区和远红外区。测试这三个区间的红外光谱所用的红外仪器或仪器内部的配置是不相同的，这三个区间所获得的光谱信息也不相同。表 2.1 列出了这三个红外区所对应的波长和波数。

表 2.1 不同红外区对应的波长和波数

区间	波长/ μm	波数/ cm^{-1}
近红外区	0.78 ~ 2.5	12800 ~ 4000
中红外区	2.5 ~ 25	4000 ~ 400
远红外区	25 ~ 100	400 ~ 10

2.2 傅里叶变换红外光谱仪的工作原理及结构

红外光谱仪中所使用的红外光源发出的红外光是连续的，从远红外到中红外到近红外区间，是由无数个无限窄的单色光组成的。当红外光源发出的红外光通过迈克尔逊干涉仪时，每一种单色光都发生干涉，产生干涉光。红外光源的干涉图就是由这些无数个无限窄的单色干涉光组成的。也可以说，红外干涉图是由多色光干涉组成的。

为了更好地理解在迈克尔干涉仪中多色光的干涉情况，首先考虑单色光的干涉情况。如果一个单色光源在理想状态下能发出一束无限窄的理想的准直光，即单色光，假设单色光的波长为 λ (单位: cm)，波数 ν (单位: cm^{-1})。波长 λ 和波数 ν 的关系为：

$$\nu = \frac{1}{\lambda} \quad (2-3)$$

假定分束器是一个不吸收光的薄膜，它的反射率和透射率各为 50%。当单色光照射到干涉仪中的分束器后 50% 的光发射到固定镜，又从固定镜发射回分束器，另外 50% 的光透射过分束器到达动镜，又从动镜反射回到分束器。这

两束光从离开分束器到重新回到分束器，所走过的距离的差值叫做光程差。如图 2.1 所示，光程差 δ 即

$$\delta = 2(OM - OF) \quad (2-4)$$

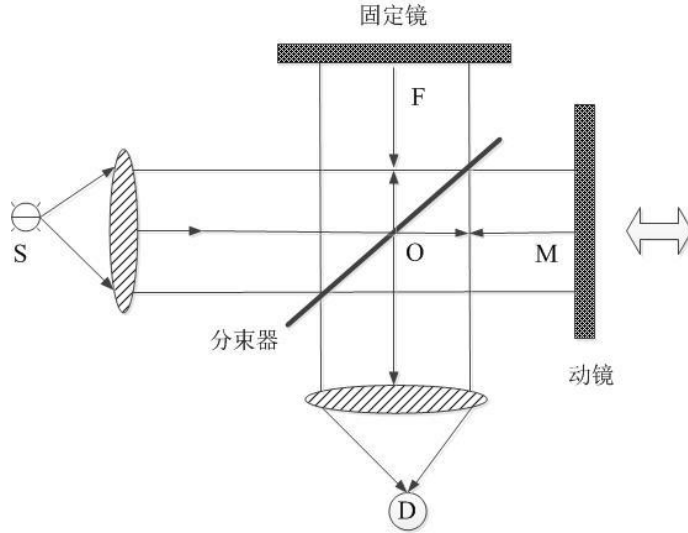


图 2.1 迈克尔逊干涉仪示意图

当固定镜和动镜到分束器的距离相等时，称此时的光程差为零光程差 ($\delta=0$)。在零光程差时，从固定镜和动镜反射回到分束器上的两束光，他们的相位完全相同，这两束光相加后并没有发生干涉，相加后光的强度等于这两束光的强度之和。如果从固定镜发射回来的光全部投射过分束器，从动镜反射回来的光也全部在分束器上反射，那么，检测器检测到的光强就等于单色光源发出的光强。如果动镜以匀速移动，检测器检测到的信号强度呈余弦波变化，也就是说，单色光的干涉图是一个余弦波。每当光程差等于单色光波长的整数倍时，到达检测器的信号最强。检测器检测到的干涉光的光强 $I'(\delta)$ 由下式给出：

$$\begin{aligned} I'(\delta) &= 0.5I(\nu)[1 + \cos(2\pi \frac{\delta}{\lambda})] \\ &= 0.5I(\nu)[1 + \cos(2\pi\nu\delta)] \end{aligned} \quad (2-5)$$

式中， $I(\nu)$ 表示波数为 ν 的单色光光源的光强。从式(2-5)可以看出，当光程差 δ 等于波长的整数倍时，即 $\cos(2\pi \frac{\delta}{\lambda})=1$ 时，有

$$I'(\delta) = I(\nu) \quad (2-6)$$

式中，光强 $I'(\delta)$ 由两部分组成：一部分是常数项 $0.5I(\nu)$ ，另一部分是余弦调制项 $0.5I(\nu)\cos(2\pi\delta\nu)$ 。在光谱测量中，只有余弦调制项的贡献是主要的。干涉图就是由余弦调制项产生的。单色光通过理想的干涉仪得到的干涉图 $I(\delta)$ 由下面的方程给出：

$$I(\delta) = 0.5I(\nu)\cos(2\pi\nu\delta) \quad (2-7)$$

检测器检测到的干涉图强度不仅正比于光源的强度，而且正比于分束器的效率、检测器的响应和放大器的特性。以上三个因素对于某一特定仪器的影响会保持不变，是一个常量。因此，方程(2-7)应该乘以一个与波数有关的因子 $H(\nu)$ ，变为：

$$I(\delta) = 0.5H(\nu)I(\nu)\cos(2\pi\nu\delta) \quad (2-8)$$

设 $0.5H(\nu)I(\nu) = B(\nu)$ ，则式(2-8)变为：

$$I(\delta) = B(\nu)\cos(2\pi\nu\delta) \quad (2-9)$$

式(2-9)是干涉图的方程， $I(\delta)$ 即为 $B(\nu)$ 的 cosine 傅里叶变换。为了得到红外光谱图还要对式(2-9)进行逆变换：

$$B(\nu) = 2\int_{-\infty}^{+\infty} I(\delta)\cos(2\pi\nu\delta)d\delta \quad (2-10)$$

2.3 红外光谱定量分析技术

2.3.1 光谱数据预处理

实验采集得到的红外光谱通常都需要进行数据处理，但是要得到高质量的红外光谱并不是一件容易的事情，尤其是矿井中的气固两相流的环境会给光谱处理带来更多的困难，因此，需要对光谱数据进行预处理。目前，常用的光谱数据预处理的方法包括：基线校正、光谱归一化及乘谱、小波变换、光谱导数和傅里叶退卷积光谱等。

(1) 基线校正

基线校正就是将吸光度光谱的基线人为地拉回到 0 基线上。在进行基线校正之前，通常都将光谱转换成吸收光谱。当然也可以对投射率光谱进行基线校正，校正后的光谱基线与 100%线重合。基线校正有两种方法：一种是谱进行基线校正 (baseline correction)。对于倾斜的基线和漂移的基线可选择自动基线校正方法，当然也可以采用逐点校正方法。对于出现干涉条纹的基线不

能选择自动基线校正方法，只能通过手动逐点地对基线进行校正。利用红外光谱法进行定量分析时，如需要计算吸收峰高或峰面积，最好将吸光度光谱进行基线校正。

（2）光谱归一化及乘谱

光谱归一化（normalize scale）是将光谱的纵坐标进行归一化。对于吸光度光谱，光谱归一化是将光谱中最大吸收峰的吸光度归一化为1，将光谱的基线归一化为0。其计算方式如下：

$$x_s = \frac{x - \bar{x}}{s} \quad (2-11)$$

式中， x_s 是归一化后的光谱值， x 是当前光谱值， \bar{x} 是样品数为 n 的光谱平均值， s 为样品数为 n 的标准偏差，即：

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (2-12)$$

乘谱（multiply）是将红外光谱乘以一个系数得到的光谱，也就是将光谱中所有数据点的 y 值都乘以同一个系数得到的光谱。在对红外光谱进行定量分析时，需要测试标准样品的光谱。但是有时很难找到标准样品，这时需要有一个内标。所谓内标，就是在光谱中寻找一个参比峰，将各个光谱中的参比峰归一化，即将各个光谱的参比峰乘以不同的系数，使得所有光谱的参比峰吸光度都一样，然后在比较其他吸收峰的强度，达到定量分析的目的。

（3）小波变换

小波变换是同傅里叶变换相似的分析方法，它同时具有频域和二维分辨率的变换，利用小波正交基多尺度的分解一系列不同的单一频率信号，所以这是一种精细分析信号的方法^[55]。小波变换可以处理在复杂环境下采集到的光谱数据，即使一些光谱特征被烟雾和粉尘等噪声干扰的情况下，依然能够被小波分类和识别，使得它在光谱数据处理方面有巨大优势。小波变换是把采集到的红外光谱信号 $x(t)$ 可以通过一簇函数加权求和得到，这簇函数由小波正交基 $\psi(t)$ 经过平移和伸缩形成。时间平移为 b ，伸缩尺度为 a 的小波 $\psi_{a,b}$ 定义为：

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right), a \neq 0 \quad (2-13)$$

定义 $W_x(a,b)$ 为连续的小波变换：

$$W_x(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t) \psi\left(\frac{t-b}{a}\right) dt \quad (2-14)$$

设 $a = 2^j, b = k2^j$ ，离散化连续的小波，则离散后的小波为：

$$\psi_{j,k}(t) = 2^{-\frac{j}{2}} \psi(2^{-j}t - k) \quad (2-15)$$

基于小波变换原理的 Mallat^[56-57]算法将信号处理分为两部分：信号的分解与信号的合成。其中 $c_{j,k}$ 是平滑后的原始信号， $d_{j,k}$ 为细节信号， g_n 是与小波相关的带通滤波器的脉冲响应。

信号分解：

$$\begin{cases} c_{j,k} = \sum h_{n-2k} c_{j-1,n} \\ d_{j,k} = \sum g_{n-2k} c_{j-1,n} \end{cases}, j = 1, 2, \dots, J \quad (2-16)$$

信号合成：

$$c_{j-1,k} = \sum [h_{n-2k} c_{j,n} + g_{n-2k} d_{j,n}], j = J, J-1, \dots, 1 \quad (2-17)$$

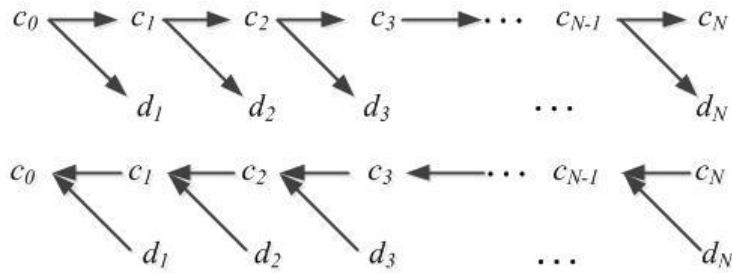


图 2.2 Mallat 算法信号分解与合成图

从图 2.2 信号的分解与合成图中可看出，Mallat 算法将原始信号 $c_0(n)$ 分解为某一尺度的平滑信号和多尺度的细节信号。通过小波变换，可以将复杂环境下的干扰信号除去，从而更好的提取红外光谱数据的特征。

(4) 光谱导数

数学上，曲线上某一点的一阶导数是这一点的切线的斜率。红外光谱转换为一阶导数光谱，就是计算出光谱中每个数据点处切线的斜率，连成曲线就成一阶导数光谱。一阶导数光谱能够显示出原光谱中的吸收峰和肩峰。

二阶导数属于高阶导数，将红外光谱的一阶导数光谱再求一次一阶导数得到的光谱就是原光谱的二阶导数光谱。二阶导数光谱能够找到原光谱中吸收峰和肩峰的准确位置（中心位置）。

(5) 傅里叶退卷积光谱

利用傅里叶退卷积 (Fourier self-deconvolution) 光谱数据处理技术可以增强红外光谱的分辨能力, 它可以将严重的谱带分开。在红外光谱中, 很多谱带是由两个以上的窄谱带合成的来的。对于这种谱带, 采集光谱数据时, 无论分辨率设的多么高, 都无法将这些窄谱带分辨开。而傅里叶退卷积光谱数据处理技术却能将这窄谱带有效分开。

傅里叶变换红外光谱是对有限长光程差的干涉图进行傅里叶变换得到的。在对有限长光程差干涉图进行傅里叶变换时, 要用一个切趾函数 $D(\delta)$ 乘以无限长光程差测得的干涉图, 如根据式 (2-13) 可得:

$$B_m(\nu) = 2 \int_{-\infty}^{+\infty} I(\delta) D(\delta) \cos(2\pi\nu\delta) d\delta \quad (2-18)$$

实测光谱 $B_m(\nu)$ 是干涉图函数和切趾函数分别进行傅里叶变换的卷积, 所以所谓傅里叶退卷积光谱, 就是将卷积得到的实测光谱退卷积, 及将实测光谱重新变成干涉图, 然后在选择一个合适的切趾函数与干涉图相乘, 在重新进行傅里叶变换就完成了全部傅里叶退卷积运算。

2.3.2 定量分析理论

气体的红外光谱定量分析主要是研究气体浓度与傅里叶红外光谱仪的响应 (光谱数据) 之间的关系, 并构建二者之间的校正模型, 用于预测待测气体的浓度。构建气体定量分析模型流程如图 2.3 所示。

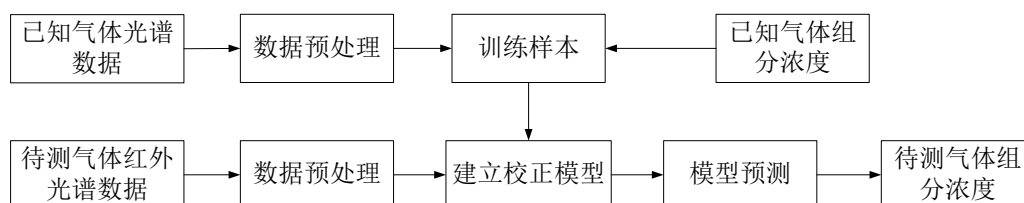


图 2.3 光谱定量分析流程

目前, 比较经典的气体定量分析方法有: 朗伯-比尔定律法 (Lambert-Beer)、多元线性回归 (MLR)、偏最小二乘法 (PLS)、支持向量机 (SVM) 和极限学习机 (ELM) 等。每一种定量分析模型都有自身优势和局限性, 所以为了提高模型的综合性能, 往往在实际应用中要根据模型的特点进行权衡, 从而选择出最佳的定量分析模型。

(1) 朗伯-比尔定律法 (Lambert-Beer)

红外光谱定量分析的依据是朗伯—比尔定律，简称为比尔定律。比尔定律表述为：当一束光通过样品时，任一波长光的吸收强度（吸光度）与样品中各组分的浓度成正比，与光程长成正比。对于组分单一的红外吸收光谱，在任一波数(ν)的吸光度为：

$$A(\nu) = \lg \frac{1}{T(\nu)} = a(\nu)bc \quad (2-19)$$

式中， $A(\nu)$ 和 $T(\nu)$ 分别表示在波数(ν)处的吸光度和透射率； $A(\nu)$ 是没有单位的； $a(\nu)$ 表示在波数(ν)处的吸光度； b 表示光程长； c 表示样品浓度。

由于红外光谱的吸光度具有加和性，对于 N 个组分的混合样品，在波数(ν)处的总吸光度为：

$$A(\nu) = \sum_{i=1}^N a_i(\nu)bc_i \quad (2-20)$$

由式（2-20）可知，吸光度可能在波数(ν)处出现叠加，而根据式（2-19）建立的定量分析模型，对于多组分混合气体，是不能准确表征混合气体中具体组分浓度的。综合分析来看，根据朗伯—比尔定律建立的定量分析模型虽然结构相对简单，但是在使用已经建立的定量分析模型对未知样本进行预测时，必须要确定未知样本中的具体组分，且要求不能有非线性因素等干扰。这使得朗伯—比尔定律法在建立混合气体定量分析模型方面有诸多的局限性。

（1）多元线性回归（MLR）

多元线性回归又称为 P 矩阵法，是通过将朗伯—比尔定律式（2-19） $A(\nu) = a(\nu)bc = kc$ 做变换，得到：

$$\mathbf{c} = \mathbf{pA} \quad (2-21)$$

式中， $p = 1/k$ 。引入服从 $\mu(0, \sigma)$ 的正态分布的随机误差 e_i ，得到推广矩阵：

$$\mathbf{C} = \mathbf{AP} + \mathbf{E} \quad (2-22)$$

其中， \mathbf{C} 为 $n \times q$ 维，是 n 个由 q 种组分组成的混合气体样品的浓度数据矩阵。 \mathbf{A} 为 $n \times m$ 维，是在 m 个红外光谱波数处测得的 n 个混合气体样品的红外光谱数据矩阵。 \mathbf{P} 为 $m \times q$ 维，是一个系数矩阵。 \mathbf{E} 为 $n \times q$ 维，是一个随机误差矩阵。针对某一特定混合气体样品，由最小二乘原理可知：

$$\mathbf{C} = \mathbf{A}\hat{\mathbf{P}} \quad (2-23)$$

$$\hat{\mathbf{P}} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{C} \quad (2-24)$$

矩阵 $\hat{\mathbf{P}}$ 代表的是矩阵 \mathbf{C} 和矩阵 \mathbf{A} 之间的对应的线性关系。利用矩阵 $\hat{\mathbf{P}}$ 可以

根据未知混合气体样本的光谱数据 A ，预测其浓度：

$$\hat{c} = A\hat{p} \quad (2-25)$$

但在求解 P 矩阵时，根据线性方程组的求解原则，当系数矩阵的秩等于增广矩阵的秩时，方程有解；而当系数矩阵的秩小于方程组未知数个数 N 时，方程有无穷多解。这就要求在求解关系矩阵 \hat{P} 时，样本数一定要大于等于波段数。

(3) 偏最小二乘法 (PLS)

1983 年，化学家伍德和阿巴诺等人最早提出了偏最小二乘法，它主要研究的是多因变量对多自变量的回归建模，具体在气体定量分析领域，主要研究的是多组分的红外光谱数据（自变量）和其浓度（因变量）之间的关系，并建立回归模型。PLS 很好地解决了多组分气体红外谱线重叠（多重共线性）问题，同时也克服了样本个数少于变量个数等问题。

PLS 的算法原理：假有定 $n \times q$ 维的浓度矩阵 Y 及其 $n \times p$ 维的红外光谱矩阵 X ， q 代表混合气体的组分种类， q 代表在 q 个波数处。不失一般性，首先对光谱矩阵 X 和浓度矩阵 Y 进行标准化处理，利用主成分分析思想提取光谱数据 X 的最大主成分，将其作为因变量，并对数据矩阵 X 和 Y 进行如下变换：

$$X = TP' + E \quad (2-26)$$

$$Y = UQ' + F \quad (2-27)$$

式中， T 和 U 分别是矩阵 X 、 Y 的得分矩阵， P 和 Q 是其负载矩阵， E 和 F 为其残差矩阵。其中 $T = (t_1, t_2, \dots, t_h)$, $U = (u_1, u_2, \dots, u_l)$ 。

然后，利用多元线性回归原理对矩阵 T 和 U 做线性回归，可得：

$$U = TB + W \quad (2-28)$$

式中， U 和 T 都是 $n \times h$ 维矩阵， R 为浓度矩阵 X 的秩，一般有 $h \leq R$ ； W 是矩阵 U 、 T 的拟合残差矩阵。根据最小二乘求得矩阵 B 的估计矩阵为：

$$\hat{B} = (T'T)T'U \quad (2-29)$$

最后通过数据重构，建立浓度矩阵 Y 和红外光谱矩阵 X 之间的关系模型，即可通过式 (2-30) 来预测未知混合气体的个组分浓度：

$$\hat{Y} = UQ' = T\hat{B}Q' \quad (2-30)$$

在使用偏最小二乘法对混合气体进行定量分析过程中，由于光谱数据矩阵 X 的各组分存在一定的线性相关性，所以只要选取带有足够信息的主因子即可。在 PLS 法中，一般采用交叉有效验证来确定主成分个数，即通过增加一个主成

分并且计算出此时均方根误差或预测误差平方和来确定这个因子对模型的影响程度。其计算公式如下：

$$RMSECV = \sqrt{\frac{PRESS}{n-1}} \quad (2-31)$$

$$PRESS = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2-32)$$

式中， y_i 、 \hat{y}_i 代表增加某一主因子式的实际值和预测值。 $RMSECV$ 值和 $PRESS$ 值小，表明模型的预测结果的精度更加精确。

(4) 支持向量机 (SVM)

支持向量机 (Support Vector Machine, SVM) 是一种有监督学习模型，由 Vapnik^[58]在 1995 年所提出，是一种较为常用的机器学习方法。SVM 多用于分类，其通过选择不同的惩罚因子和损失函数可以很好地解决函数回归等问题，相比于其他很多分类方法具有很好的泛化能力且适用于任何种类的数据^[59]的优点。图 2.4 为支持向量机的网络结构图，其中 X 为输入， y 为输出， K 为核函数， b 为偏置项。

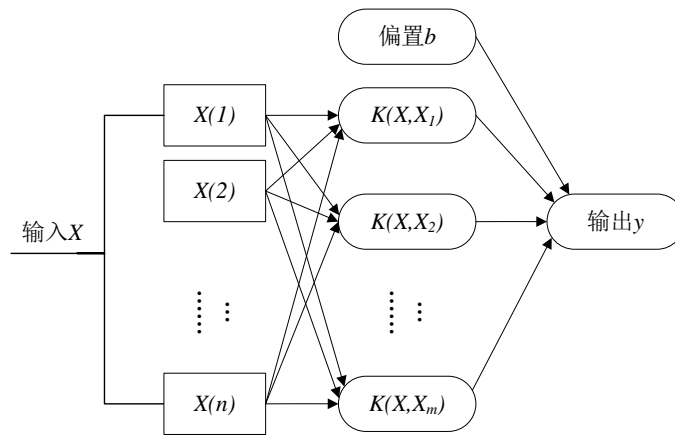


图 2.4 支持向量机网络结构

SVM 算法的原理是将低维输入空间变换到一个高维特征空间，通过合理选择核函数完成非线性转换，最终找到一个平面使特征向量中各点与该平面之间的距离尽可能达到最大，称此平面为划分超平面，以该超平面为界限将不同类别的样本分开。图 2.5 反映的是 SVM 两类问题模型。其中，“+”和“-”为待分的两类训练样本， x_1 、 x_2 为样本的两个特征项，从图中可看出存在多个划分超平面能将训练样本分开，例如 H_1 、 H_4 ， H_2 和 H_3 分别为过两类样本中离最优划分超平面最近且平行的平面，称为支持平面， H_2 和 H_3 之间的距离被称为间

隔，位于 H_2 和 H_3 上的训练样本点称为支持向量（Support Vector, SV）， H 为理想划分超平面，该划分超平面对训练样本局部的抗扰动性最好，与两类样本间的距离最大，产生的分类结果的鲁棒性最好。

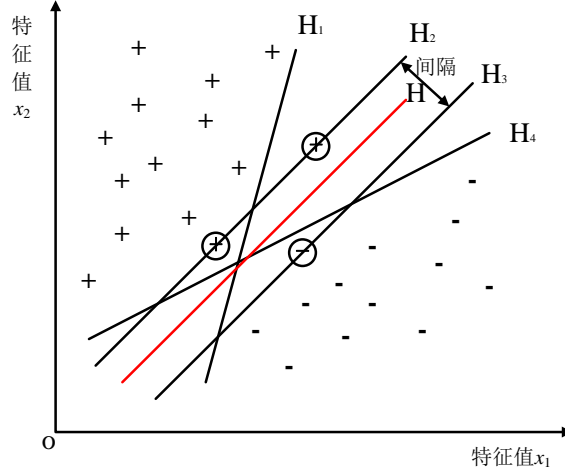


图 2.5 SVM 分类超平面示意图

样本空间 (x_i, y_i) ，其中 $i=1, 2, 3, \dots, l$ ， $x_i \in R^D$ ， $y_i \in R$ ， x_i 表示输入量， y_i 表示输出量，划分超平面可由如下线性方程定义：

$$f(x) = w^T x + b \quad (2-33)$$

式中， w 为法向量，决定了超平面的方向； b 为位移项，决定了超平面与原点间的距离。可见，划分超平面可由法向量 w 和位移 b 所确定。

在样本空间中，任意点到超平面的距离可写为：

$$r = \frac{|\omega^T + b|}{\|\omega\|} \quad (2-34)$$

假设训练样本可被超平面正确分类，对于 (x_i, y_i) ，

$$\begin{cases} w^T x_i + b > 0, y_i = +1 \\ w^T x_i + b < 0, y_i = -1 \end{cases} \quad (2-35)$$

间隔为：

$$r = \frac{2}{\|w\|} \quad (2-36)$$

式(2-36)可直观理解为：间隔越大，由最优划分超平面所定义的线性分类

器分类的准确率也就越高。欲找到最优划分超平面，也就是要找到满足式(2-37)中约束的参数 w 和 b ，使 w 最大，即

$$\max_{w,b} \frac{2}{\|w\|} \quad (2-37)$$

$$s.t. \quad y_i(w^T x_i + b) \geq 1, i = 1, 2, \dots, m \quad (2-38)$$

可见，为了使间隔达到最大，仅需最大化 $\|w\|^{-1}$ ，等价于最小化 $\|w\|^2$ ，于是，式(2-37)可改写为：

$$\max_{w,b} \frac{1}{2} \|w\|^2 \quad (2-39)$$

式(2-39)是一个凸二次规划问题，问题可转化为对 w 和 b 求下面 Lagrange 函数的极小值。

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 + \sum_{i=1}^m \alpha_i (1 - y_i(w^T x_i + b)) \quad (2-40)$$

式(2-40)中， $\alpha > 0$ 为 Lagrange 系数，问题是对 w 和 b 求式(2-43)的极小值。令 $L(w, b, \alpha)$ 对 w 和 b 偏导为零，得：

$$w = \sum_{i=1}^m \alpha_i y_i x_i \quad (2-41)$$

$$\sum_{i=1}^m \alpha_i y_i = 0, (\alpha_i \geq 0, i = 1, 2, \dots, m) \quad (2-42)$$

将式(2-41)代入式(2-40)，可将 $L(w, b, \alpha)$ 中的 w 和 b 消去，再考虑到式(2-42)的约束，问题转化为在此约束条件球下式的最大值：

$$Q(\alpha) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j x_i^T x_j \quad (2-43)$$

解出 α 后，求出 w 和 b 即可得到模型：

$$f(x) = w^T x + b = \sum_{i=1}^m \alpha_i y_i x_i^T x + b \quad (2-44)$$

SVM 在求解过程中,通过核函数的引入,解决了在高维空间中关于内积的运算问题,将样本从低维度的空间映射到一个高维度的特征空间,非线性分类问题得到了解决。利用 SVM 进行分类问题处理的性能好坏很大程度依赖于对模型的选择,优化分类模型的方法可分为:1) 构造一个全新的模型;2) 对当前已有模型的参数重新进行智能寻优。考虑到构造全新模型的难度较大,故本文将重点放在优化已有模型参数上。优化模型参数分为两部分:核函数的类型选取和核函数的参数选取。核函数的类型选取对分类器性能好坏起决定作用,它可直接决定分类的最终结果;而核函数的参数选取不当将导致分类器不能够正确分类。表 2.2 列出了几种常用的核函数。

表 2.2 支持向量机常用核函数

名称	表达式	参数
多项式核	$K(x_i, x_j) = (x_i^T x_j)^d$	$d \geq 1$, 为多项式的系数
拉普拉斯核	$K(x_i, x_j) = \exp(-\frac{\ x_i - x_j\ }{\sigma})$	$\sigma > 0$
线性核	$K(x_i, x_j) = x_i^T x_j$	-
RBF 核	$K(x_i, x_j) = \exp(-\frac{\ x_i - x_j\ ^2}{2\sigma^2})$	$\sigma > 0$, 为 RBF 核的宽带
Sigmoid	$K(x_i, x_j) = \tanh(\beta x_i^T x_j + \theta)$	\tanh 为双曲正切函数, $\beta > 0, \theta < 0$

(5) 极限学习机 (ELM)

2006 年,新加坡南洋理工大学副教授 Huang 提出了一种新的前馈网络,极限学习机 (Extreme Learning Machine, ELM) 概念,并介绍了其基本原理和诸多优点^[60-61]。

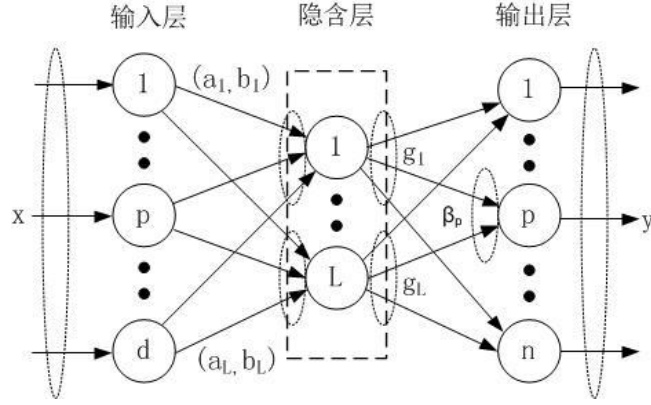


图 2.6 ELM 网络结构图

如图 2.6 所示，ELM 网络由输入层、隐含层、和输出层组成。若有 N 个训练样本 $\{(x_j, t_j) | x_j \in X \subset R^d, t_j \in R^m, j=1, 2, \dots, N\}$ 从输入端输入，对于含有 L 个隐节点且激活函数为 $g(x)$ 的单隐层向前神经网络 (SLFNs) 的输出函数可以表示为：

$$f(x) = \sum_{i=1}^L \beta_i g_i(x_i) = \sum_{i=1}^L \beta_i g_i(\alpha_i x_i + b_i), j=1, 2, \dots, N \quad (2-45)$$

式中， (α_i, b_i) 为第 i 个隐节点与输入层之间产生的随机参数， α_i 为输入权值， b_i 为偏置， β_i 是该节点与输出层之间的权值。若 $f(x)$ 的期望输出为 t ，则：

$$f(x) = \sum_{i=1}^L \beta_i g_i(\alpha_i x_i + b_i) = t_j, j=1, 2, \dots, N \quad (2-46)$$

式(2-46)可简化为：

$$\mathbf{H}\beta = \mathbf{T} \quad (2-47)$$

其中，

$$\mathbf{H} = \begin{pmatrix} g(\alpha_1 \mathbf{x}_1 + b_1) & \cdots & g(\alpha_L \mathbf{x}_1 + b_L) \\ \vdots & \ddots & \vdots \\ g(\alpha_1 \mathbf{x}_N + b_1) & \cdots & g(\alpha_L \mathbf{x}_N + b_L) \end{pmatrix}_{N \times L}, \quad \beta = \begin{pmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{pmatrix}_{L \times m}, \quad \mathbf{T} = \begin{pmatrix} t_1^T \\ \vdots \\ t_L^T \end{pmatrix}_{N \times m}$$

\mathbf{H} 为 ELM 的隐含层输出矩阵，其是由隐含层参数 (α_i, b_i) 确定，所以当 (α_i, b_i) 随机产生后， \mathbf{H} 即为已知。

D-ELM (Dynamic ELM) 算法是在 ELM 和 AG-ELM^[62-63]、EM-ELM^[64]等算法基础上改进而来。该算法中隐节点是动态增长的，初始隐节点 $L_0=1$ ，随后可以每次增加 $\delta L (\delta L \geq 1)$ 个，且隐节点参数 (α_i, b_i) 均是随机产生的，输出权值 β_i

则是通过最小误差原则迭代得到。初始化网络时，隐层输出矩阵为：

$$\mathbf{H} = \mathbf{H}(\alpha_1, \dots, \alpha_{l_0}; b_1, \dots, b_{l_0}; x_1, \dots, x_N) \quad (2-48)$$

则：

$$\hat{\beta} = \mathbf{H}_1^+ \mathbf{T} \quad (2-49)$$

式中， \mathbf{H}_1^+ 表示 \mathbf{H}_1 的 Moore-Penrose 广义逆^[65]，则误差为：

$$E(\mathbf{H}_1) = \|\mathbf{H}_1 \hat{\beta} - \mathbf{T}\|^2 = \|\mathbf{H}_1 \mathbf{H}_1^+ \mathbf{T} - \mathbf{T}\|^2 \quad (2-50)$$

令 ξ 为系统给定期望误差，若 $E(\mathbf{H}_1) < \xi$ ，则学习结束。否则，在隐含层增加 δL 个隐节点，此时输出矩阵变为：

$$\mathbf{H}_2 = [\mathbf{H}_1, \delta \mathbf{H}_1] \quad (2-51)$$

D-ELM 算法流程图如 2.7 所示。

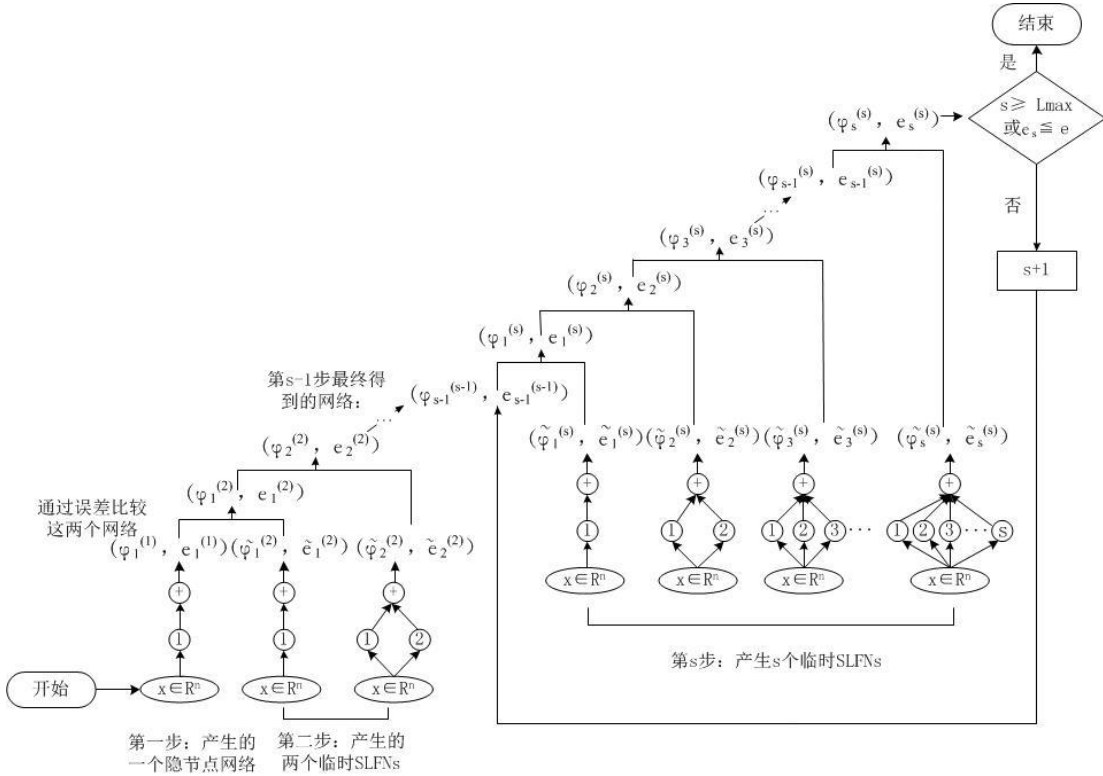


图 2.7 D-ELM 算法流程

具体步骤如下：

已知：训练样本集为 $\mathfrak{N} = \{(x_i, t_i)_{i=1}^N | x_i \in \mathbb{R}^n, t_i \in \mathbb{R}^m\}$ ，激活函数为： $g(x)$ ，最

大隐含节点为: L_{max} , 期望误差为: ξ 。

Step 1: 初始化产生带 $L_0=1$ 个隐节点的网络, 并求得输出矩阵的 Moore-Penrose 广义逆 $\mathbf{H}_1^{+(1)}$ 和输出权值 $\beta_1^{(1)}$, 得到一个 SLFNs $\varphi_1^{(1)}$: $\varphi_1^{(1)} = \beta_1^{(1)} \mathbf{H}_1^{+(1)}$ 。

Step 2: 比较当前网络输出与目标函数误差可得到两个 SLFNs $\varphi_1^{(1)}$ 和 $\varphi_2^{(2)}$:

$$\varphi_1^{(2)} = \begin{cases} \varphi_1^{(1)}, & \text{when } \|\varphi_1^{(1)} - f\|^2 \leq \|\tilde{\varphi}_1^{(2)} - f\|^2 \\ \tilde{\varphi}_1^{(2)}, & \text{others} \end{cases} \quad (2-52)$$

$$\varphi_2^{(2)} = \begin{cases} \varphi_1^{(2)}, & \text{when } \|\varphi_1^{(2)} - f\|^2 \leq \|\tilde{\varphi}_2^{(2)} - f\|^2 \\ \tilde{\varphi}_2^{(2)}, & \text{others} \end{cases} \quad (2-53)$$

其中, $\tilde{\varphi}_1^{(2)} = \beta_1^{(2)} \mathbf{H}_1^{+(2)}$, 以及 $\tilde{\varphi}_2^{(2)} = \beta_1^{(2)} \mathbf{H}_1^{+(2)} + \beta_2^{(2)} \mathbf{H}_2^{+(2)}$ 分别表示 $L=1$ 和 $L=2$ 的两个 SLFNs。

Step 3: 重复第二步骤到 j 步, 此时隐节点超过 L_{max} 或小于 ξ , 保留此网络

$$\tilde{\varphi}_k^{(j)} = \sum_{i=1}^k \beta_{i-1}^{(j)} \mathbf{H}_{j-1}^{+(j)}, (1 \leq k \leq j)。$$

2.3.3 模型的验证与评价

当混合气体预测模型建立之后, 还需要验证光谱数据和浓度数据之间的关系模型是否准确, 即需要对所建立的模型进行评价。因为本文是气固两相流中气体组分在线测量方法研究, 所以评价的指标有时间 t 、误差 e 、相关系数 R (Correlation Coefficient) 等。

(1) 模型预测所用时间 t :

$$t = t_{end} - t_{start} \quad (2-54)$$

式中, t_{end} 代表模型运行结束时的时间, t_{start} 代表模型运行开始时的时间。

(2) 误差 ($error$) 和平均误差 ($Mean_error$):

$$error = \hat{y}_i - y_i \quad (2-55)$$

$$Mean_error = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i) \quad (2-56)$$

式中， \hat{y}_i 和 y_i 分别代表第 i 样品的预测值和标准值， n 为样品数。

(3) 相关系数 R :

$$R = \frac{\sum_{i=1}^n (y_i - \bar{y}_i)(\hat{y}_i - \bar{\hat{y}}_i)}{\sqrt{\sum_{i=1}^n (y_i - \bar{y}_i)^2 \sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}}_i)^2}} \quad (2-57)$$

式中， $\bar{\hat{y}}_i = \frac{\sum_{i=1}^n \hat{y}_i}{n}$ 和 $\bar{y}_i = \frac{\sum_{i=1}^n y_i}{n}$ 分别代表第 i 样品的预测值和标准值的平均值， n 为样品数。

(4) 均方根误差 ($RMSEC$) 和预测均方根误差 ($RMSEP$):

$$RMSEC = \sqrt{\frac{\sum_{i=1}^{n_c} error_i^2}{n_c - 1}} \quad (2-58)$$

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n_p} error_i^2}{n_p - 1}} \quad (2-59)$$

式中， n_c 和 n_p 分别为测试集和验证集的样品数，均方根误差和预测均方根误差的值越小，模型效果越好，结果越精确。

2.3 本章小结

本章主要介绍了红外光谱分析技术的基本原理，主要包括红外光谱的产生及如何划分红外光谱、傅里叶红外光谱仪的基本原理和红外光谱定量分析技术。阐述了红外光谱在多组分混合气体定量分析的相关理论、光谱数据预处理的方法和各种模型的原理和评价标准。

3 气固两相流中气体光谱数据采集

为了采集到矿井下气固两相流中气体组分的红外光谱数据，首先要搭建安全可靠的实验数据采集系统，然后设计合理的实验和数据采集方案。实验系统一般包括配气装置、傅里叶红外光谱仪、计算机等。其中配气装置至关重要，它通过精确计量输入到傅里叶红外光谱仪的气体池中的气体样本各组分的量，来保证所测样本气体浓度数据的准确性，基于此建立的定量分析模型才有较好的可靠性。

3.1 配气装置的研发

如图 3.1 所示仪器为计算机智能 5 路混气系统，所有操作功能都通过显示屏上的触摸屏完成，只需输入需要配置和稀释的气体浓度，通过点击触摸屏便可配置和显示所要配置的气体浓度。在保障浓度精度的前提下，可实现混合气体精密、连续供给。



图 3.1 5 路高精度配气装置

配气装置系统配有液晶触摸屏上位机，直观显示配比过程中的具体参数。带有 MCGS 组态软件，能够在配气现场调试过程中，修正、存储程序、一键启动和预保存需要的配比比例。配比过程中，也可现场通过触摸屏，微调质量流量计，达到更高的精度。选用高精度质量流量控制器，可实现稳定、可靠、高精度的配气要求。

如图 3.2 所示为配气装置的内部结构图，其主要由以下几部分组成：5 个高

精度质量流量控制器 (HORIBA METRON)、5 个 PID 流量控制器控制 MFC 仪表 (宇电)、5 个浮子流量计 (DWYER) 5 个气体过滤器 (FITOK)、精密全不锈钢气体稳压阀 (调压型)、RS485 通讯线、10 寸液晶触屏机 1062K+组态软件和执行器等。其中 5 个高精度质量流量控制器包括 2 个 50ml/min、2 个 100ml/min 和 1 个 500ml/min 量程的质量流量计, 用于精确计量不同气体输入到气体池中的量。配气装置的设计和使用方法如下所述。



图 3.2 5 路高精度配气内部结构图

(1) 气路设计

1) 如图 3.3 所示, 椭圆虚线框内标记, 5 路气路进口, 1 路混合气出口, 位于仪器背面。聚四氟乙烯管路快插连接, 规格 6mm 外径, 壁厚至少 1mm。为了整个气路的密封性, 插入时, 需要一定力度, 进入时应有阶梯顿挫手感。

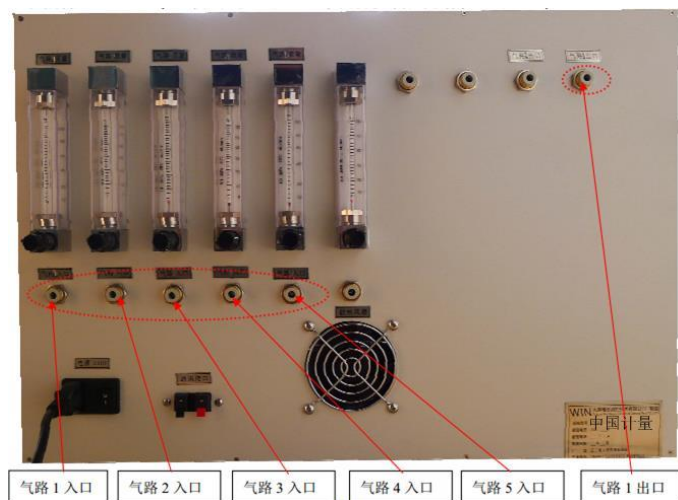


图 3.3 气体输入输出接口图

在实际使用时，可能存在未使用的气路入口，稳妥考虑，应采用堵头封住（设备内部流量计密封，前面板稳压阀门，后面板流量计阀门，三层密封）如图 3.4 所示。



图 3.4 接口堵头

2) 开启气瓶。管路连接完成后，如图 3.5 所示，钢瓶开关。然后，旋转调节螺杆（或减压阀旋钮），将低压表读数，调节至 2~3kgf。

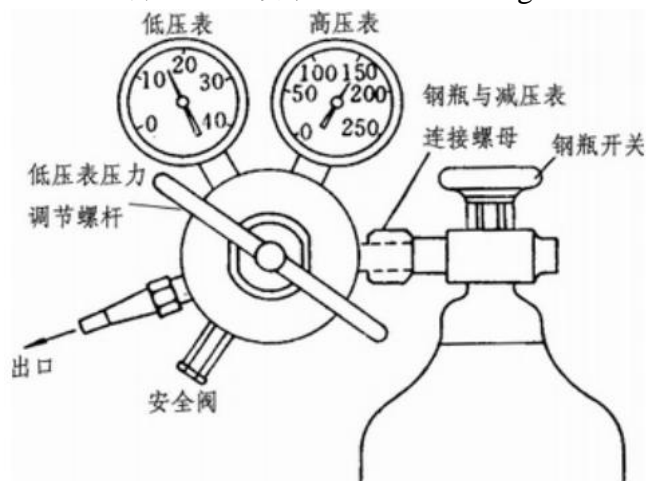


图 3.5 钢瓶开关图

3) 浮子流量计开启。逆时针旋转黑色浮子流量计旋钮，如图 3.6 所示，椭圆内所示，开至最大即可。（未连接管路的支路，顺时针旋转浮子流量计旋转至截止位置）

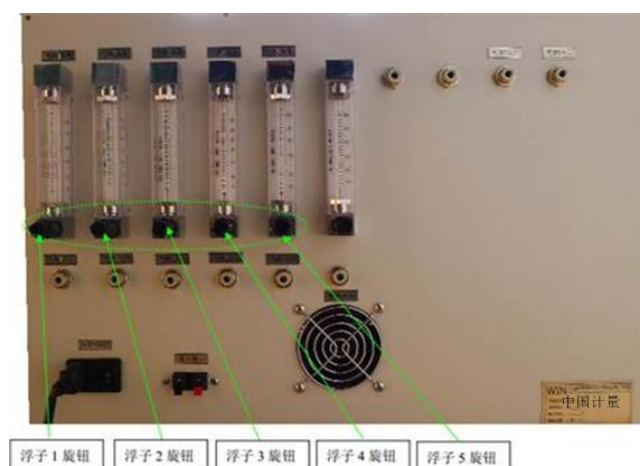


图 3.6 浮子流量计旋钮图

4) 稳压阀开启。用于精密稳定质量流量控制器入口处的气体压力。顺时针旋转，开启稳压阀按钮，如图 3.7 所示的椭圆虚框内。同时观察与其对应的耐震压力表读数，调节至 1.5~2.5kgf 即可。(未连接管路的支路，逆时针旋转旋钮，至截止位置)



图 3.7 装置前面板视图

上述 4 步流程，如果不更换气瓶、增减气体支路，进行一次即可。但是钢瓶的开关，每次均需要开启或关闭。

(2) 电源设计

1) 连接后面板电源，采用 220V/50HZ 交流电源，功率 2KW。启动前面板设备电源，首先，按下最左端“总电源”按钮，如图 3.7 所示，若无异常，则继续按下“仪表电源”和电脑电源。(按钮处于凸起状态为断开；处于凹陷状态

为连接)

2) 预热及稳定。电源开启。系统进入预热状态, 无需任何操作, 等待 10 分钟即可, 在此期间可以进行软件操作、设定、参数校核等, 不会干扰预热。

(3) 软件设计及操作

软件包括 4 个界面: 启动界面、手动界面、自动模式—参数设置、自动模式—运行。

1) 启动界面: 配气装置启动后即进入此界面。如图 3.8 所示, 界面上有自动模式、手动模式和关机, 用手点击一下即可进入相应界面。进入自动模式, 主要是自动配气流程, 各项操作均由此完成; 手动模式仅用于检验气路, 以及全自动配气; 关机, 仅用于当采用“通讯接口”与外部连接时, 关闭本电脑主机, 本装置“通讯接口”采用的是 RS485 通讯协议。



图 3.8 启动界面

2) 手动模式界面: 在“启动界面”, 点击“手动模式”触屏按钮, 即可进入图 3.9 所示的手动模式界面。



图 3.9 手动模式界面

- 本界面主要用于检验气路，比如系统密封性测试，为纯手动流量设定界面。
 - 如图 3.9 所示，全部关闭流量计（触屏按键），将所有流量计设定为 0 ml/min，即全部关闭。流量计关闭返回启动界面，即有图 3.9 返回启动界面图 3.8。
 - 在图 3.9 最下面椭圆框内，五个红色指示灯，代表 5 个气路通讯状态，红绿闪烁，则通讯正常；红色不变化，则通讯异常。
 - 流量输入有两种方法，一是手动触摸滑块输入；二是触摸屏软键盘输入。
 - 进行流量实测时，测量每路气体流量，并以刻度条及数值方式予以显示。
- 3) 自动模式—参数设置界面：在“启动界面”，点击“自动模式”触屏按钮，即可进入图 3.10 所示的自动模式—参数设置界面。本界面用于设定各项参数以及对参数进行校验，最终开启运行。气路 5 为总平衡气，为稀释气体，通常为氮气或者空气。



图 3.10 自动模式—参数设置界面

本界面操作流程如图 3.10 椭圆流程图所示，按照箭头所示方向依次执行。具体的操作流程如图 3.11 所示。



图 3.11 操作流程圖

- 详细操作流程说明如表 3.1 所示。

表 3.1 详细操作说明表

序号	名称	说明
1	气体参数设计	气体参数 输入各路气体（非氮气组分）的参数
		气源成分 输入对应气体分子式，仅作为标识用
		上述参数 仅跟换气体种类时，需要修订该项
2	浓度参数	浓度参数 输入非氮气组分浓度（纯氮气，输入 0）
		气源浓度 输入气瓶标签浓度值
		目标浓度 输入所需气体组分浓度值
		气路 5 总平衡气 目标浓度不需要输入
		上述参数 仅变更时，需重新输入
3	请输入：总流量	请输入 5 路气体的总流量，单位 ml/min
4	检验—确认	点击“检验—确认”，进行总体校验
		1、图 3.10 中标识 3 下方的 5 个框内，为计算后的该路气体流量设定值，如果此值不在该路流量计量程内，则其后面的指示灯将变 成红色。此时，应调整“总流量”或“目标浓度”两项参数，并再次点击“检验—确认”。
		如此反复。
5	开启—运行	2、如点击“检验-确认”，图 3.10 标识 3 下方的指示灯均为绿色，则可进行下一步“开启-运行”
		点击“开启—运行”，关闭“自动模式—参数配置”，进入“自动—运行”界面

4) 自动模式—运行界面在“自动模式—参数设置”界面，“检验—确认”通过后，点击“开启—运行”触屏按钮，即可进入图 3.12 所示的自动模式—运行界面。



图 3.12 自动模式—运行界面

本界面为运行监控界面，只有操控按钮，无需输入任何参数。各区域的功能及说明如下：

- 1) 总流量设定值：如图 3.12 所示，标号 1，用于显示 5 个气路，气源成分，气源浓度、目标浓度，以及 5 路总流量。
- 2) 每路气体流量设定值：标号 2，用于显示 5 个气路。
- 3) 流量实测：标号 3，用于测量每路气体流量，并以刻度条及数值方式予以显示。
- 4) MFC 通讯状态：标号 4，由 5 个红色指示灯代表 5 个气路通讯状态：红绿闪烁，则通讯正常；红色不变化，则通讯异常。
- 5) 操控按钮：标号 5，含义如按钮文字。

3.2 实验系统的搭建

矿井下的气固两相流气体组分数据采集系统主要由傅里叶红外光谱仪、配气装置、气体池、温控仪、压力表、计算机、比色皿和连接管路等组成。图 3.13 所示为整个系统的结构简图。

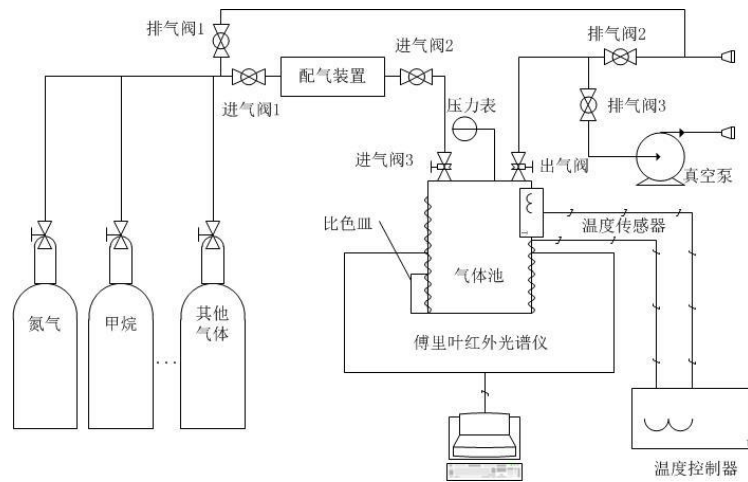
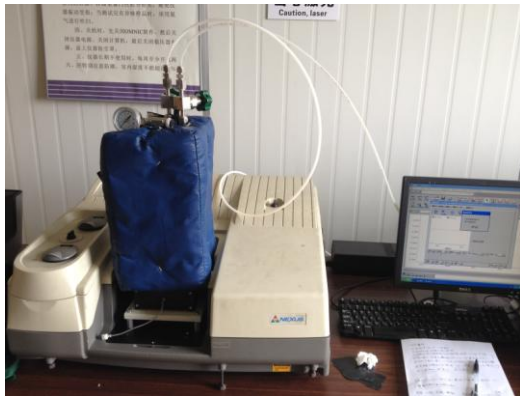


图 3.13 矿井下的气固两相流气体组分数据采集系统

该系统的实物图如图 3.14 所示 (a) 为红外光谱仪与气体池, (b) 为配气装置与计算机, (c) 和 (d) 为各种气体的气瓶。



(a) 红外光谱仪与气体池



(b) 配气装置与计算机

(c) CH_4 、 CO_2 、 N_2 气瓶(d) C_2H_6 、 C_3H_8 、 $\text{n-C}_4\text{H}_{10}$ 等气瓶

图 3.14 数据采集系统实物图

该系统的实验原理是各组分气体通过配气装置按照实验计划向红外光谱的气体池中输入定量气体，最后通入氮气来稀释其他气体，完成浓度比配置。通过在红外光谱仪的光路中加入带有煤颗粒的比色皿来模拟矿井下的气固两相流环境，如图 3.15 所示。利用压力表和温控仪来监测气体池中的压力和温度等情况，保证实验在相同环境下进行；待输入的气体充满整个气体池后，用红外光谱仪采集气体数据，通过计算机软件 OMNIC 软件得到光谱图并保存光谱数据。

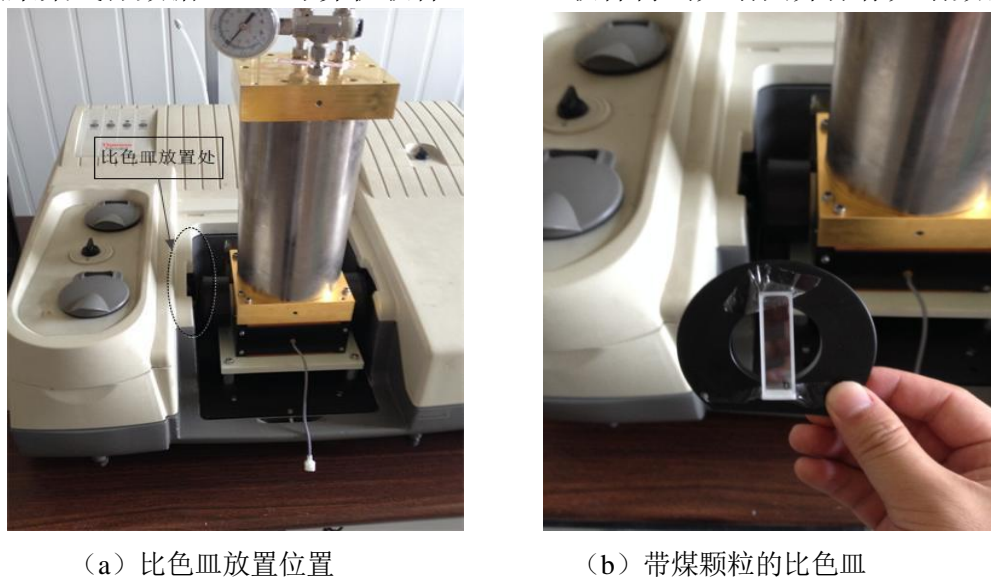


图 3.15 比色皿

本文实验系统所使用的傅里叶红外光谱仪为 Thermo Fisher 公司的型号为 Nicolet-Nexus870 的光谱仪，由于所检测气体的吸收波段都在中红外区域，所以选择中红外的 Globar 光源，设置其扫描范围为 $(1000\sim4000)\text{cm}^{-1}$ ，扫描次数为 32 次，扫描分辨率为 4cm^{-1} ，并选择光程长为 10m 的气体池。

实验步骤如下：

- ①气密性检测。在采集实验数据之前都要对实验系统进行气密性检测，即通过用真空泵将整个实验系统的气路中气体进行抽取，当真空压力表指针指在“0”刻度并维持稳定 5min 以上，即可确定系统的气密性完好。
- ②用除湿器对实验环境进行除湿，并保持室温在 25°C 。
- ③首先通过配气装置向气体池中充入氮气，当气体池上端的压力表显示为一个大气压后，将此时的光谱作为光谱实验的背景光谱，通常每隔 2 个小时采

一次背景光谱。

④重复步骤①将气路中的气体排光，抽真空。

⑤通过配气装置设置各组分气体的浓度比，向气体池中输入已知浓度的待测气体样本，其中氮气作为稀释气体最后输入，保证把之前输入的气体都充入气体池中。

⑥充气结束后，等 5~10 分钟，待充入气体充分混合均匀后，分别采集将比色皿放在光路前后的红外光谱，并保存。

⑦同一样品采三次数据，间隔为 3 分钟，观察光谱数据的重复性。

⑧一次数据采集完成后，利用氮气进行吹扫工作，消除前一次测量的残余气体。

⑨重复④~⑧进行下一个样本的数据采集。

如图 3.16 为采集到的甲烷光谱图，其中，红色为气固两相流中甲烷的光谱图，蓝色为光路中没有比色皿的甲烷光谱图。

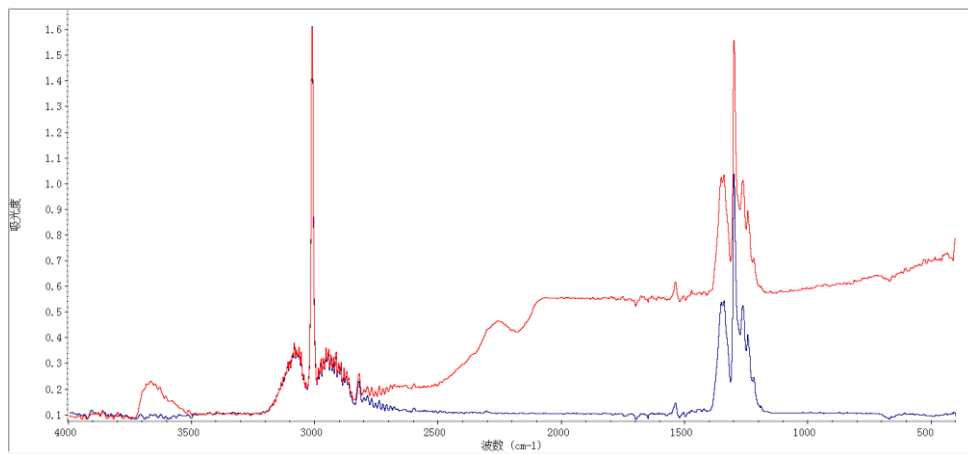


图 3.16 采集的气体光谱图

3.3 实验样本

根据实验需要，确定合理的配气方案，然后采集光谱数据作为测试集是光谱定量分析的基础和关键。本文根据矿井下的实际情况，确定了五种气体作为分析对象，即 CH_4 、 C_2H_6 、 C_3H_8 、 C_4H_{10} 、 CO_2 ， N_2 为稀释气体。在采集气体红外光谱数据之前，需要根据矿井中的实际情况确定五种气体各组份的浓度范围

及配置方案，如表 3.2 所示。

表 3.2 每种气体的浓度范围

序号	组份名称	浓度范围
1	CH ₄	0~20%
2	C ₂ H ₆	0~1%
3	C ₂ H ₄	0~1%
4	C ₄ H ₁₀	0~1%
5	CO ₂	0~20%
6	N ₂	0~100%

如图 3.17 为上述五种单组份气体浓度为 1% 时的红外光谱图。

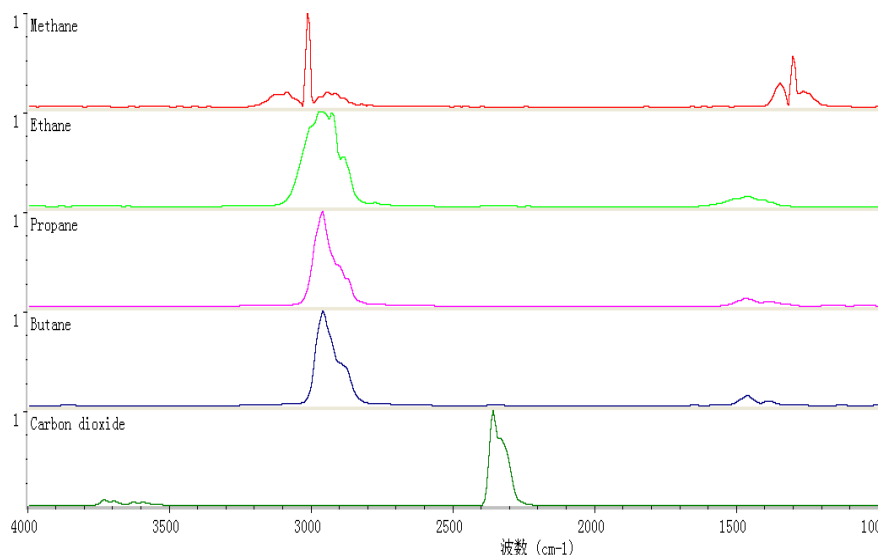


图 3.17 五种单组份气体的红外光谱图

从图 3.17 可以看出，5 种气体中的烃类气体 CH₄、C₂H₆、C₃H₈、C₄H₁₀ 的特征吸收峰存在交叉重叠，主要集中在 (2800~3200) cm⁻¹ 波段，可能给光谱特征提取带来困难，特别是在气固两相流环境中，存在固相煤颗粒的干扰，这也是气体定量分析的难点。CO₂ 的特征峰在 (2200~2400) cm⁻¹ 这个区域，独立于其他烃类气体，故理论上 CO₂ 的预测精度应该高于其他四种气体。各组分特征吸收峰的波段如表 3.3 所示。

表 3.3 各气体组份特征峰吸收范围/ cm^{-1}

名称	主要吸收范围/ cm^{-1}	
CH_4	1200~1400	2850~3200
C_2H_6	1280~1600	2720~3125
C_3H_8	1280~1580	2550~3380
C_4H_{10}	1200~1600	2500~3250
CO_2	2200~2400	3550~3750

如果五种气体按 0.1% 的间隔配置,则需要配置 4×10^7 个样本并采集其红外光谱数据,工作量太大。所以采用 Matlab 2010b 在每种气体浓度范围内随机产生单组分数据 200 组,多组分数据 1200 组(已去掉相同的数据),然后将 200 组单组分样本数据和 1200 组多组分样本数据分别按 CH_4 的浓度排序,再稍加改动即可作为本次实验的气配方案。如表 3.4 为部分多组分气体浓度配置样本。

表 3.4 部分多组分气体浓度配置

序号	CH_4	C_2H_6	C_3H_8	C_4H_{10}	CO_2
1	1%	0.1%			
2	1%	0.1%	0.1%		1%
3	5%	0.2%	0.1%		2.5%
4	10%	0.1%			5%
5	10%	0.1%	0.2%	0.1%	7.5%
6	20%	0.3%	0.3%	0.3%	10%
7	2.5%	0.5%	0.5%		12%
8		1%	0.4%	0.6%	15%
9	7.5%		0.6%	0.2%	
10	15%			1%	6%
...

按气配方案配置 1400 组样本并采集其红外光谱数据,由于采集光谱数据量较大,所以在采集过程中可能会出现光谱数据和标定浓度不匹配的情况,需要踢除这种异常的光谱数据。本文先通过观察气体光谱图,根据经验判断其是否为异常样本,如果遇到难以判断的样本则采用马氏距离法来识别是否为异常样本。根据马氏距离法去除 14 组异样样本,剩余 1386 组样本的光谱数据作为建模数据。

3.4 本章小结

本章详细介绍了气固两相流中气体组分红外光谱数据采集系统以及其中最为重要的配气装置的研发，并设计了合理的配气方案。根据本文的研究目标研发了简单易操作的配气装置，可以根据配气方案，精确地计量输入到气体池中气体的量，通过红外光谱仪得到每个样本的红外光谱图，分析每种气体的特征吸收峰、吸收波段等，为气体定量分析模型提供可靠的实验数据。

4 基于支持向量机的定量分析

由于矿井中气体组分多且部分气体浓度变化范围较大以及空气中煤粉颗粒等干扰,使得气体浓度和吸光度之间呈现非常复杂的非线性关系,加上烃类气体特征吸收波段严重重叠,这给建立气体定量分析模型增加了难度,使得一些常规的线性分析方法已经无法满足要求。支持向量机(SVM)通过引入核函数,可以将样本从低维度的映射到高纬度的特征空间,能够很好地解决非线性关系问题。

由于本文的研究对象是矿井下的气固两相流,其中的固相即煤粉等颗粒对气体的红外光谱干扰非常严重,所以在建立气体分析模型之前,光谱的预处理和特征提取显的尤为重要。本文采用小波分析对光谱进行降噪,然后对降噪后的光谱数据进行归一化处理,并通过提取特征峰值和主成分分析法对处理后的数据进行降维和特征提取,最后建立气体的浓度预测模型,并对预测结果进行分析和讨论,其流程图如 4.1 所示。

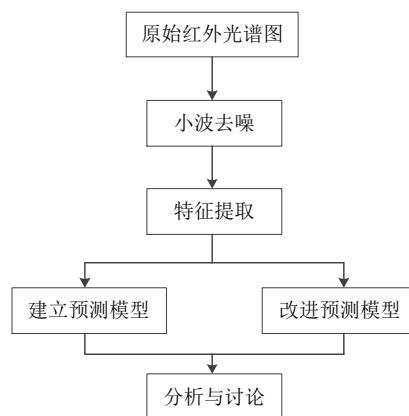


图 4.1 建模流程图

4.1 光谱数据处理

由于气固两相流中固相煤粉颗粒的干扰,使得红外光谱仪在采集气体光谱图时难免会掺杂一些噪声。因此,为了还原信号中的有效光谱信息,在分析光谱数据之前需要对采集到的光谱数据进行去噪处理。小波阈值去噪是一种目前使用较普遍的去噪方法,其理论依据为:由于小波变换去数据相关性很强,在小波域内,小波变换可以使光谱信号的能量在一些大的小波系数中被集中,然

而噪声的能量却散布于整个小波域内。可以认为噪声的小波系数较小，而信号的小波系数幅值一般较大，较平稳，因此能够把正常信号和噪声区分开，然后将有用的信号系数设定阈值的方法将有进行保留，而对于噪声信号则将其系数减小至零，达到滤波的目的。小波阈值去噪的基本步骤为：选择合适的小波对光谱数据进行分解，可以得到各尺度系数；再根据上述理论，设定阈值，对小波系数进行进一步的处理；最后进行小波逆变换重构信号，获得消噪后的信号。

将小波阈值去噪应用于采集到的原始光谱信号，得到生成信号即去噪后的光谱信号，并计算两者之间的信噪比 SNR。通过信噪比来考察在原始光谱信号信息去噪过程中的丢失情况，信噪比越高代表原始光谱信号信息丢失的信息越少，光谱信息保留的越好。由于光谱数据是离散的数据信号，所以采用离散小波分析方法，设置小波分解层数为 5 层^[66]，采用四种离散小波系对光谱信号进行分解即 Db 小波系、Symlet 小波系、Biorthogonal 小波系、Coiflet 小波系，并分别计算每种小波系生成信号与原始信号的信噪比，选择信噪比最大的小波系对光谱数据进行分解去噪。

表 4.1 四种小波去噪结果

Db 小波系		Symlet 小波系		Biorthogonal 小波系		Coiflet 小波系	
小波函数	SNR	小波函数	SNR	小波函数	SNR	小波函数	SNR
db1	64.6743	Sym1	64.6743	Bior1.1	64.6743	Coif1	66.7197
db2	66.1686	Sym2	66.1686	Bior1.3	65.6067	Coif2	67.2652
db3	66.5860	Sym3	66.5860	Bior1.5	65.0001	Coif3	67.5767
db4	66.7399	Sym4	67.0207	Bior2.2	66.7781	Coif4	66.9560
db5	66.6421	Sym5	67.0228	Bior2.4	66.9378	Coif5	66.6201
db6	66.6363	Sym6	67.0419	Bior2.6	67.6656		
db7	67.0016	Sym7	66.9665	Bior2.8	67.8408		
db8	66.2496	Sym8	67.0245	Bior3.1	64.7667		
db9	65.6139	Sym9	66.3347	Bior3.3	67.2735		
db10	65.6778	Sym10	67.0064	Bior3.5	67.3116		
db11	65.5172	Sym11	66.3531	Bior3.7	68.4333		
db12	65.6330	Sym12	66.9084	Bior3.9	67.1979		
db13	65.6622	Sym13	66.0611	Bior4.4	65.4266		
db14	65.4307	Sym14	66.7832	Bior5.5	65.4533		

如表 4.1 所示为四种小波去噪后的光谱信号和原始光谱信号的信噪比，可以发现当小波函数为 Bior3.7 时，信噪比最大为 68.4333，说明其在去噪方面保留的有效信息最多，同时兼顾去除大部分的干扰信号，所以选择 Bior3.7 为去噪的小波函数。如图 4.2 和 4.3 所示是 CH_4 、 C_2H_6 、 C_3H_8 、 C_4H_{10} 和 CO_2 五种气体浓度比为[6.5 0.8 0.1 0.1 3.4]的去噪前后的光谱图。

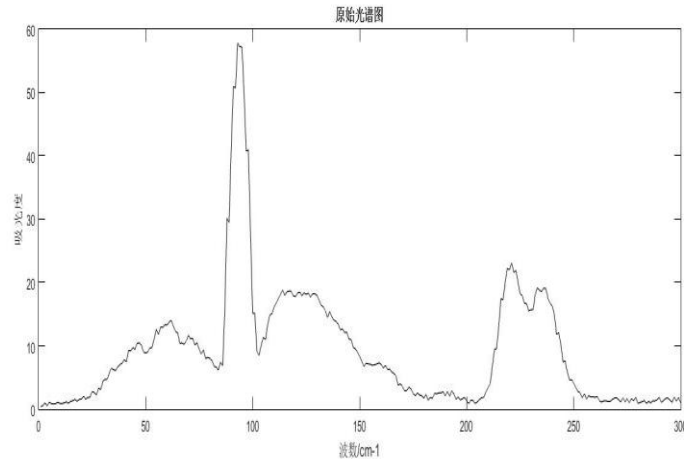


图 4.2 原始光谱图

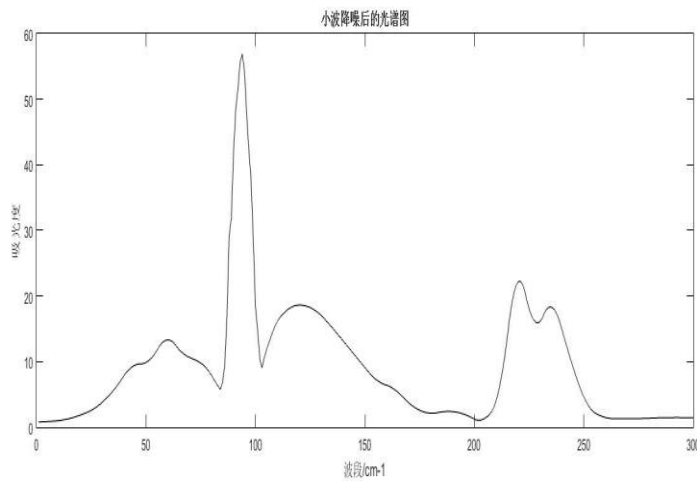


图 4.3 小波降噪后的光谱图

4.2 基于 SVM 气固两相流中气体组分的定量分析

4.2.1 基于 PEM-SVM 气固两相流中气体组分的定量分析

由于研究的五种矿井气体的红外吸收光谱信息主要集中在波段： $(2100\sim 2400)\cup(2500\sim 3400)\text{cm}^{-1}$ ，所以选择该区间的吸收光谱信息代替整个光

谱数据。红外光谱仪的分辨率为 4cm^{-1} ，所以一个样本将包含 300 个数据。将光谱数据用矩阵的形式表示，则光谱数据矩阵 X_1 的维度为 1386×300 ，浓度数据矩阵 C_1 的维度为 1386×5 。本文将五种矿井气体的吸收光谱峰值所在的波段附近作为其特征值进行提取即采用峰值提取法（Peak Extraction Method, PEM）对特征值进行提取。

表 4.2 特征提取波段及个数

名称	波段/ cm^{-1}	特征值个数
CH_4	2850~3200	16
C_2H_6	2720~3125	14
C_3H_8	2550~3380	24
C_4H_{10}	2500~3250	10
CO_2	2200~2400	11

如表 4.2 所示， CH_4 在其特征光谱区间 $2850\sim 3200\text{cm}^{-1}$ 内取一个主波峰和一个次波峰共 16 个特征值。 C_2H_6 在其特征光谱区间 $2720\sim 3125\text{cm}^{-1}$ 内取一个主波峰和两个次波峰共 14 个特征值。 C_3H_8 在其特征光谱区间 $2550\sim 3380\text{cm}^{-1}$ 内取一个主波峰和一个次波峰共 24 个特征值。 C_3H_8 在其特征光谱区间 $2550\sim 3380\text{cm}^{-1}$ 内取一个主波峰和一个次波峰共 10 个特征值。 $\text{n-C}_4\text{H}_{10}$ 在其特征光谱区间 $2500\sim 3250\text{cm}^{-1}$ 内取一个主波峰和一个次波峰共 11 个特征值。 CO_2 在其特征光谱区间 $2200\sim 2400\text{cm}^{-1}$ 内取一个主波峰和一个次波峰共 11 个特征值。提取完特征值的光谱数据矩阵 X_2 的维度为 1386×75 。浓度数据矩阵依然为 C_1 。

在建立模型之前，需要把光谱数据矩阵 X_2 和对应浓度数据矩阵 C_1 进行随机排序得到 X , C ，目的是把之前配气方案排序好的样本重新混合，使得样本分布的更加均匀。然后将光谱数据矩阵 X 的前 1000 个样本光谱数据作为训练集，后面的 386 组光谱数据作为测试集。PEM-SVM 的原理如图 4.4 所示。

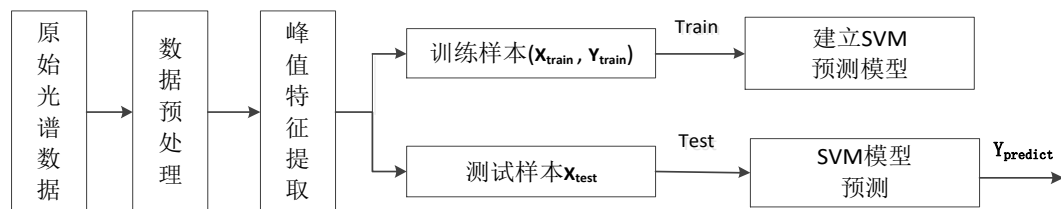


图 4.4 PEM-SVM 的原理

图 4.4 中， $(X_{\text{train}}, Y_{\text{train}})$ 为训练集样本， X_{train} 为经过峰值提取的特征值，

$\mathbf{Y}_{\text{train}}$ 代表训练样本的输出即浓度值, \mathbf{X}_{test} 代表测试集样本, $\mathbf{Y}_{\text{predict}}$ 代表 \mathbf{X}_{test} 的预测值。建立 PEM-SVM 的定量分析模型, 首先用训练样本 ($\mathbf{X}_{\text{train}}$, $\mathbf{Y}_{\text{train}}$) 建立一个反映 $\mathbf{X}_{\text{train}}$ 与 $\mathbf{Y}_{\text{train}}$ 之间的非线性关系的模型, 然后根据已建立的模型对测试集样本 \mathbf{X}_{test} 进行预测得到 \mathbf{Y}_{test} 。

4.2.2 基于 PCA-SVM 气固两相流中气体组分的定量分析

由于采集到的气体光谱数据样本比较大, 所以为了提高运算效率又要对光谱数据进行降维处理同时提取特征信息。采用主元分析法 (Principal Component Analysis, PCA) 的目的在于通过线性变化删减无用信息, 将经过小波降噪处理过的复杂光谱数据压缩降维, 降低数据之间的冗余信息以提高后续分类精度。PCA 不仅能对高维数据进行降维操作得到特征向量 \mathbf{Y} , 并且能够去除了冗余信息和噪声。

PCA 进行特征降维的实现过程为:

1) 用 x_1, x_2, \dots, x_n 代表 n 个样本的特征, 每个样本的特征维度为 m , 用矩阵 \mathbf{X} 表示为:

$$\mathbf{X}_{n \times m} = \begin{bmatrix} x_{11} & \dots & x_{1m} \\ \vdots & \vdots & \vdots \\ x_{n1} & \dots & x_{nm} \end{bmatrix} \quad (4-1)$$

2) 对 \mathbf{X} 的每一维进行零均值化, 即矩阵中的每一行的每个元素减去这一行的均值, 得到矩阵 \mathbf{X}' , 表示为:

$$x'_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (4-2)$$

式中, $i=1, 2, \dots, N$, $j=1, 2, \dots, n$, \bar{x}_j 代表 X_j 的均值, s_j 代表 X_j 的方差。

3) 计算 \mathbf{X}' 协方差矩阵 \mathbf{C} ;

4) 根据步骤 (3) 中得到的协方差矩阵 \mathbf{C} , 计算其特征向量并求出特征值;

5) 将求得特征向量按对应特征值的大小从小到大按行排列成矩阵, 计算特征向量的贡献率以确定最终降维的数目 k 。在贡献率大于 95%(不是定值, 一般取 95%) 的特征值中取前 k 个特征值, 并将其求对应的特征向量组成矩阵 \mathbf{P} 。

6) $\mathbf{Y} = \mathbf{P}\mathbf{X}$ 即为降维到 k 维后的数据。

如图 4.5 所示, ($\mathbf{X}_{\text{train}}$, $\mathbf{Y}_{\text{train}}$) 为训练集样本, \mathbf{X}_{test} 代表测试集样本, $\mathbf{T}_{\text{train}}$ 和 \mathbf{T}_{test} 分别表示由 $\mathbf{X}_{\text{train}}$ 和 \mathbf{X}_{test} 经过 PCA 特征提取后的得分矩阵, $\mathbf{Y}_{\text{train}}$ 代表训练样本的输出即浓度值, $\mathbf{Y}_{\text{predict}}$ 代表 \mathbf{T}_{test} 的预测值。PCA-SVM 建模过程和

PEM-SVM 建模过程类似，只是特征提取的方法不同。

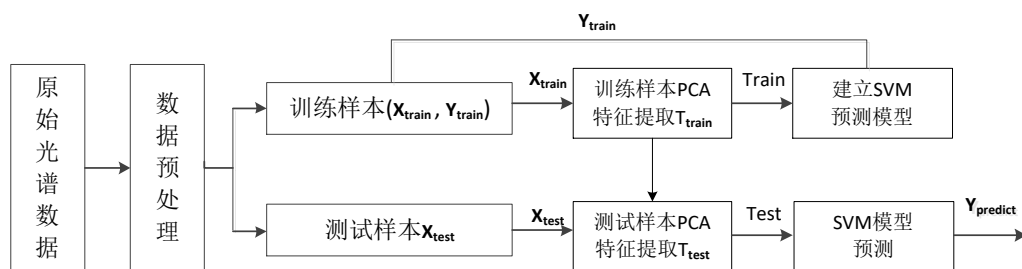


图 4.5 PCA-SVM 原理图

4.3 基于改进 SVM 气固两相流中气体组分的定量分析

4.3.1 SVM 参数寻优

本文在 SVM 核函数的选择中选用非线性拟合效果最好的 RBF 核函数，SVM 参数寻优主要是寻找惩罚参数 c 和核函数参数 g 的最优解使得目标函数最大化。目前，SVM 参数寻优方法主要有网格搜索法、粒子群算法 (PSO) 和遗传算法 (GA) 等。由于网格搜索法的寻优时间比较长，所以本文不采用这种寻优方法。PSO 算法具有算法结构简单、搜索能力强以及容易实现等优点，在众多参数寻优问题中表现出色，但有时会陷入局部最优解。GA 算法具有较好的全局搜索能力，但该算法需要确定初始解的值，初始值的好坏决定了能否快速的找到最优解。

(1) 基于 PSO-SVM 的参数寻优

使用 PSO 算法对 SVM 的 c 和 g 进行寻优，其流程如图 4.6 所示。

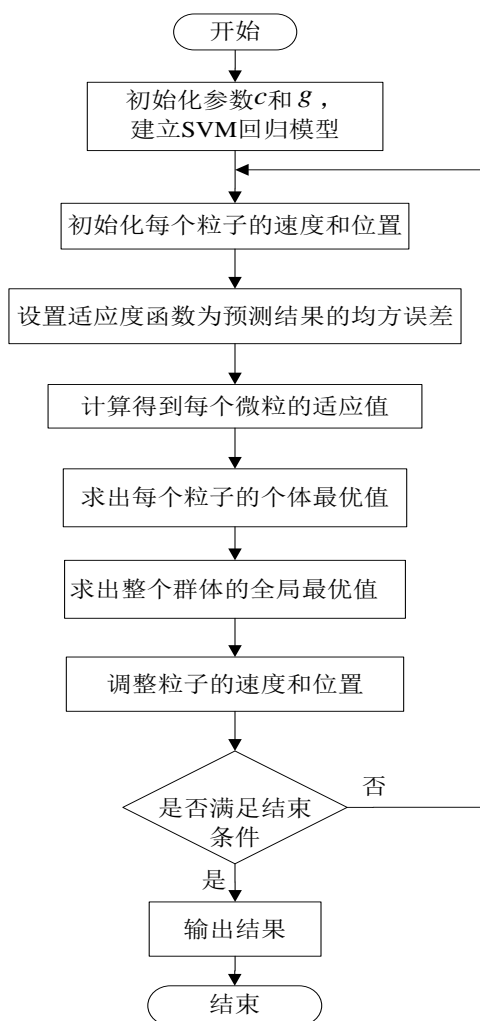


图 4.6 PSO 寻优流程图

PSO 优化算法的具体步骤如下：

- 1) 初始化，粒子群的个数为 2，种群个数为 20，迭代次数为 100，局部搜索参数 $c_1=1.5$ ，全局搜索参数 $c_2=1.7$ ；
- 2) 对种群进行评价，对每个粒子的适应度函数值进行计算；
- 3) 更新过程，比较粒子的适应度函数值和最优历史位置，若此时适应度函数值比 pbest 更优，则把当前适应度值更新为 pbest；
- 4) 比较粒子适应度值和全局最优位置 gbest，如果当前值比 gbest 更优，则将此粒子位置更新为 gbest；
- 5) 停止条件，若满足停止条件，结束寻优过程，输出寻优结果，若不满足，返回步骤 3)继续搜索，直到满足终止条件，搜索到全局最优解为止。

(2) 基于 GA-SVM 的参数寻优

如图 4.7 所示为 GA 算法寻优流程图，首先随机产生一组参数，并对其进行编码，生成初始种群。构造目标函数的适应度函数，通过选择算子、交叉算子以及变异算子来优化适应度函数，通过比较将适应度大的个体遗传给下一代，然后经过不断迭代，直到找到最优解为止。

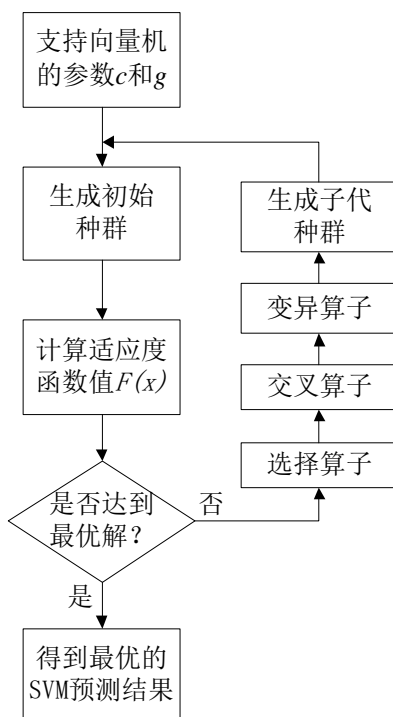


图 4.7 GA 寻优流程图

以甲烷（ CH_4 ）为例，使用经过峰值提取法提取到的特征数据的训练集作为模型输入，利用 PSO 和 GA 优化算法分别对 SVM 参数进行寻优，其结果如图 4.8 和 4.9 所示。

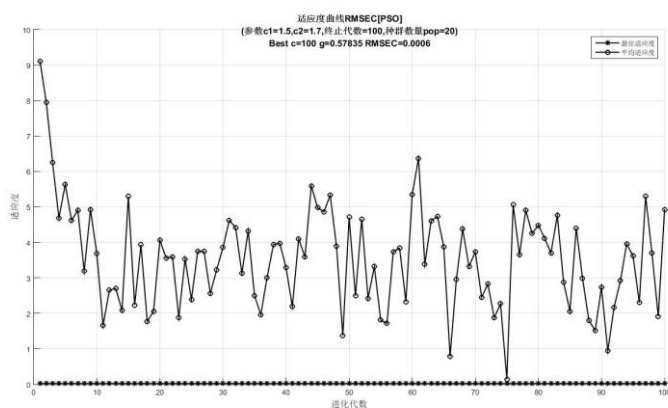


图 4.8 PSO 算法的参数优化结果

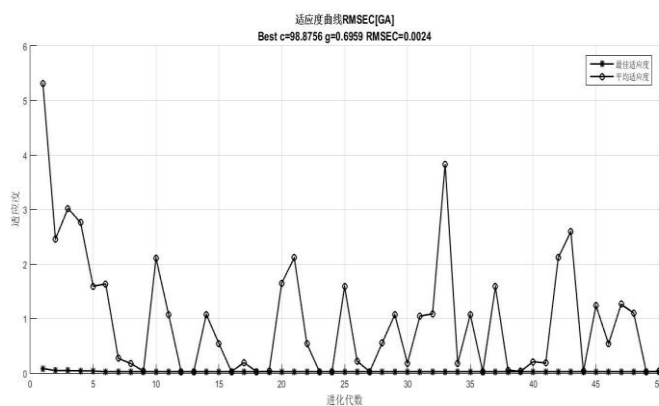


图 4.9 GA 算法的参数优化结果

PSO 和 GA 对 SVM 参数寻优结果如表 4.3 所示。通过比较可知，PSO 参数寻优建立的模型的均方差(RMSEC)比 GA 小，所以不管在预测的精度方面还是寻优所有时间方法，PSO 算法都占有优势。故本文采用 PSO 算法对 SVM 的惩罚参数 c 和核函数参数 g 进行寻优建立定量分析模型。

表 4.3 GA 和 PSO 寻优结果

寻优方法	c	g	RMSEC	寻优时间 (T/s)
PSO	100	0.5784	0.0006	2520
GA	98.8756	0.6959	0.0024	2880

4.3.2 基于改进 PEM-SVM 气固两相流中气体的定量分析

用小波分析法对气固两相流中气体的原始光谱进行降噪处理，并根据峰值提取法对降噪后的光谱进行特征提取，然后对其进行随机排序，使样本混合的更加均匀，得到特征光谱矩阵 X 维度为 1386×75 和对应的浓度矩阵 C 维度为 1386×5 ，将特征光谱矩阵 X 的前 1000 个数据作为训练集样本，后 386 个数据作为测试集样本。由于 SVM 所建立模型的输出是一维，所以预测五种气体浓度，需要利用 PSO 算法分别对五种气体的 SVM 模型参数进行寻优建立预测模型，表 4.4 为模型结果，建立模型所有时间在 3~4 小时左右，均方差 RMSEC 值比较理想，预测精度高。

表 4.4 改进 PEM-SVM 模型结果

气体种类	c	g	$RMSEC$	寻优时间 (T/s)
CH ₄	100	0.5783	0.0006	2520
C ₂ H ₆	74.5735	0.1	0.0002	1943
C ₃ H ₈	100	0.3776	0.0009	2990
C ₄ H ₁₀	100	0.5139	0.0008	2678
CO ₂	76.7923	0.1	0.0017	2789

为了体现改进 PEM-SVM 模型的优势,本文也对普通的 PEM-SVM 建立预测模型并与经过 PSO 寻优后的 PEM-SVM 模型结果进行比较如表 4.5 所示。

表 4.5 两种模型结果比较

气体种类	建模方法	$RMSEC$	寻优时间 (T/s)
CH ₄	改进 PEM-SVM	0.0006	2520
	PEM-SVM	0.0277	356
C ₂ H ₆	改进 PEM-SVM	0.0002	1943
	PEM-SVM	0.0022	279
C ₃ H ₈	改进 PEM-SVM	0.0009	2990
	PEM-SVM	0.0278	422
C ₄ H ₁₀	改进 PEM-SVM	0.0008	2678
	PEM-SVM	0.0016	375
CO ₂	改进 PEM-SVM	0.0017	2789
	PEM-SVM	0.0318	396

通过两种模型的结果比较可知,改进 PEM-SVM 模型的 $RMSEC$ 值明显比普通 PEM-SVM 小的多,即在预测精度上,改进 PEM-SVM 模型的优势十分明显。这是由于普通的 PEM-SVM 模型的惩罚参数 c 和核函数参数 g 是随机产生的,没有对其进行寻优,所以其在预测精度上很不理想,但是由于省略了参数寻优时间,故其建模速度比改进 PEM-SVM 要快很多。

为了验证已经建立的改进 PEM-SVM 气固两相流中气体组分的定量分析模型,需要用测试集样本对改进 PEM-SVM 模型进行验证。取特征光谱矩阵 \mathbf{X} 后面 386 个混合气体样本数据作为测试集,利用已建立好的改进 PEM-SVM 模型对其浓度进行预测,将测试样本的前 100 个预测结果如图 4.10~4.14 的 (a) 所

示，测试样本的相对误差如图 4.10~4.14 的 (b) 所示。

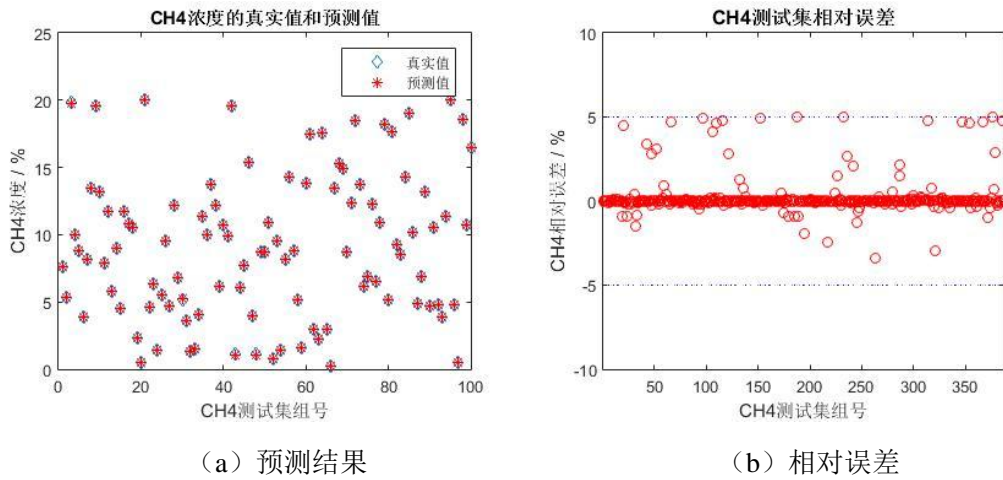


图 4.10 CH₄ 测试集改进 PEM-SVM 的预测结果

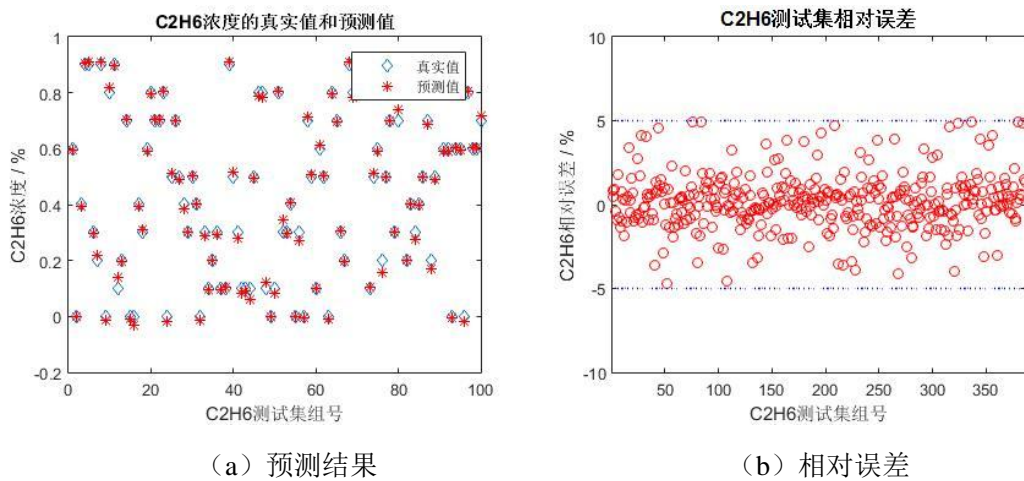


图 4.11 C₂H₆ 测试集改进 PEM-SVM 的预测结果

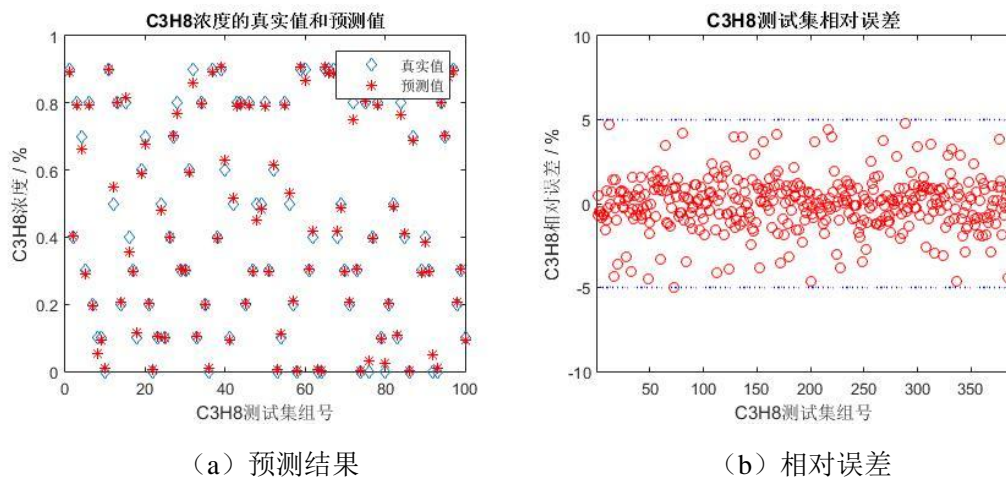
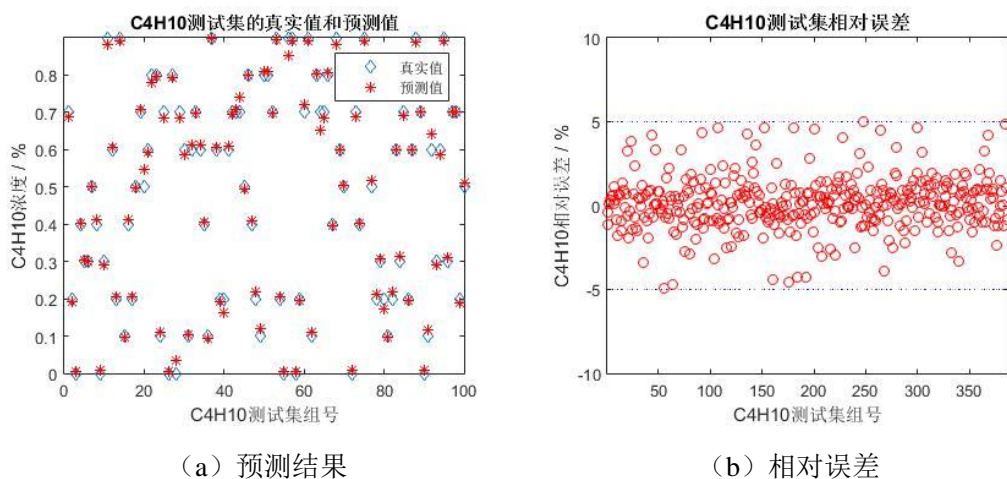
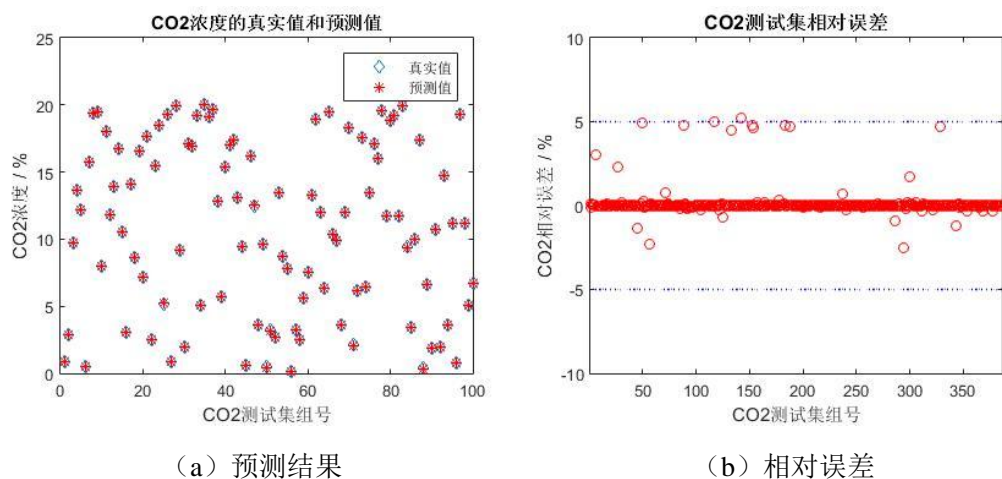


图 4.12 C₃H₈ 测试集改进 PEM-SVM 的预测结果

图 4.13 C_4H_{10} 测试集改进 PEM-SVM 的预测结果图 4.14 CO_2 测试集改进 PEM-SVM 的预测结果

根据改进 PEM-SVM 模型对 CH_4 、 C_2H_6 、 C_3H_8 、 C_4H_{10} 、 CO_2 五种气体测试集的预测结果，计算出每种气体预测浓度的均方差 ($RMSEP$)、相关系数 (R)、最大相对绝对误差、平均相对误差、相对误差超过 5% 的比例和预测一个样本的时间如表 4.6 所示。

表 4.6 改进 PEM-SVM 模型对五种混合气体的预测结果

气体 种类	<i>RMSEC</i>	<i>R</i>	最大相对 绝对误差	平均相 对误差	相对误差 超过 5% 的比例	预测时间 (T/s)
CH ₄	0.0441	0.9230	5.40%	1.37%	0.52%	1.03
C ₂ H ₆	0.0083	0.9964	4.66%	2.20%	0	1.43
C ₃ H ₈	0.0127	0.9325	4.98%	3.94%	0	0.99
C ₄ H ₁₀	0.0091	0.9699	4.57%	3.16%	0	1.34
CO ₂	0.0307	1.0000	5.23%	1.20%	0.78%	0.87

由表 4.6 分析可知, 总体的预测效果比较理想, 其中甲烷的预测结果的最大相对误差和平均相对误差都比较大, 但是最大相对绝对误差都不超过 5.40%, 各种气体的相关系数较高, 预测时间较短, 可以实现矿井中气固两相流中气体组分的在线快速预测分析。

在表 4.6 中, CH₄ 和 CO₂ 的相对误差分布比较集中在 -1%~1% 之间, 只是有少数几个样本的预测结果的相对误差比较大, 分析其原因可能有: 1) CH₄ 和 CO₂ 的浓度范围比较大; 2) 煤粉颗粒的干扰波段可能在 CH₄ 和 CO₂ 波段中, 煤粉颗粒的干扰没能彻底除去, 导致对个别样本的预测结果影响比较大。总体上整个模型对这两种气体的预测结果精度和稳定性都比较好, 只是个别样本不理想, 可预测精度的提高空间比较大。

C₂H₆、C₃H₈ 和 C₄H₁₀ 的相对误差分布比较分散在 -5%~5% 之间, 虽然平均相对误差比较大, 但是相对最大绝对误差比较理想, 都没有超过 5%, 总体来说, 改进的 PEM-SVM 模型对这三种气体的预测精度符合实际的要求。

4.3.3 基于改进 PCA-SVM 气固两相流中气体的定量分析

利用改进 PEM-SVM 建立对气固两相流中气体组分定量分析模型并用测试集对其进行验证。结果表明, 该模型具有较好的预测效果, 但是由于矿井下有煤粉颗粒的干扰, 使得峰值特征提取法在气体红外光谱特征提取方面效果欠佳, 导致 CH₄ 和 CO₂ 的预测精度并不能符合实际的要求, 所以需要引入主成分分析法对光谱特征进行提取, 来解决煤粉颗粒的干扰问题, 同时可以大幅度降低气体红外光谱数据的维度, 加快建模时间。

将经过小波降噪后的气体光谱数据进行 PCA 降维处理并提取其特征光谱。降噪光谱数据矩阵的维度为 1386×300, 对应的浓度矩阵 C 维度为 1386×5, 其

经过 PCA 处理后的结果如图 4.15 所示。

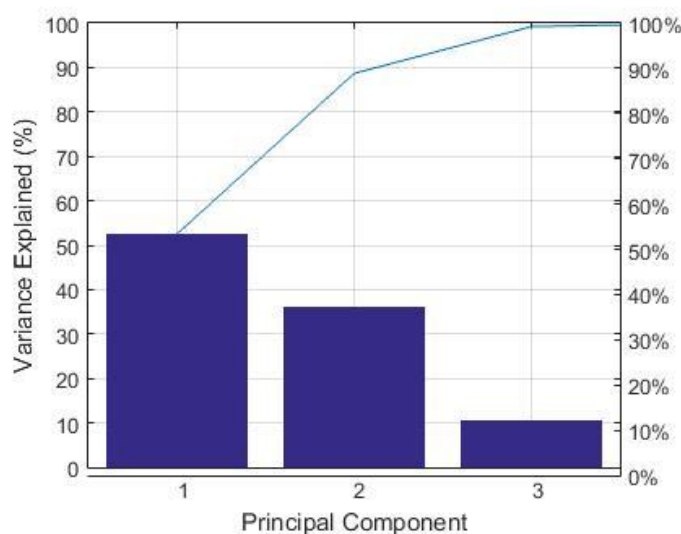


图 4.15 PCA 的降维结果

由图 4.15 可以看出，累加到第三个主成分，贡献度就达到了 99%，用其代替原有的光谱矩阵作为模型的输入，经过 PCA 特征提取后得到的得分矩阵 \mathbf{T} 的维度为 1386×3 ，浓度矩阵 \mathbf{C} 依旧为 1386×5 。利用得分矩阵 \mathbf{T} 的前一千个样本作为训练集 $\mathbf{T}_{\text{train}}$ ，后 386 作为测试集 \mathbf{T}_{test} 。 $\mathbf{T}_{\text{train}}$ 作为训练模型的输入，利用 PSO 算法对 SVM 惩罚参数 c 和核函数参数 g 进行寻优，其结果如表 4.7 所示。

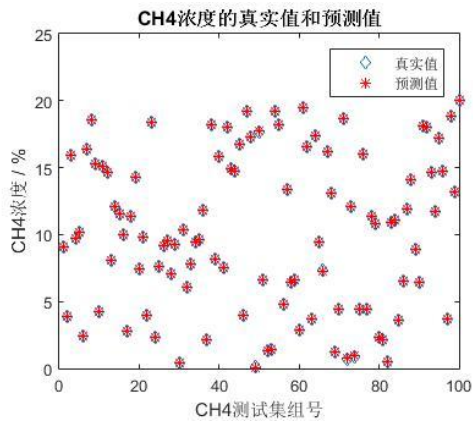
表 4.7 改进 PCA-SVM 模型结果

气体种类	c	g	$RMSEC$	寻优时间 (T/s)
CH_4	10.6818	0.1	0.0001	343
C_2H_6	1.4823	0.1	0.0009	289
C_3H_8	2.6773	0.2678	0.0014	325
C_4H_{10}	1.6919	0.5677	0.0006	348
CO_2	11.2728	0.1	0.0007	279

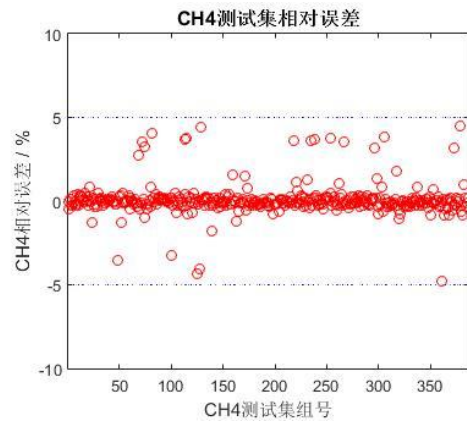
由表 4.7 可知，PCA-SVM 寻优建模时间比 PEM-SVM 明显减少，这是由于光谱数据经 PCA 处理后维度大幅减小，计算量大幅减少，故建模建模速度快。

为了验证已经建立的改进 PCA-SVM 气固两相流中气体组分的定量分析模型，需要用测试集样本对改进 PCA-SVM 模型进行验证。利用已建立好的改进 PCA-SVM 模型对测试集 \mathbf{T}_{test} 浓度进行预测，验证模型的可靠性。将测试样本的前 100 个预测结果如图 4.16~4.20 的 (a) 所示，测试样本的相对误差如图

4.16~4.20 的 (b) 所示。

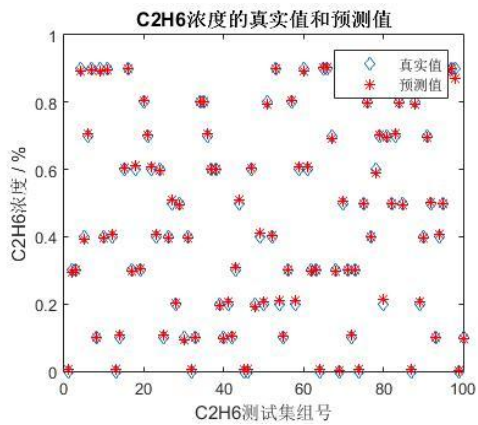


(a) 预测结果

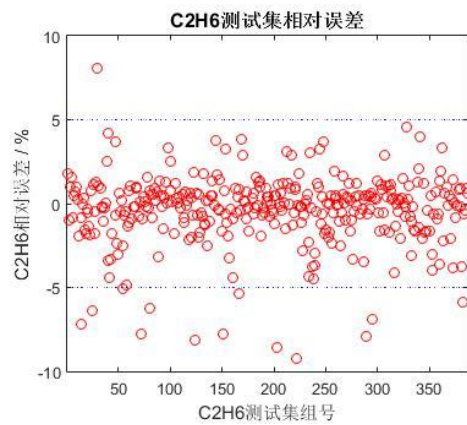


(b) 相对误差

图 4.16 CH₄ 测试集改进 PCA-SVM 的预测结果

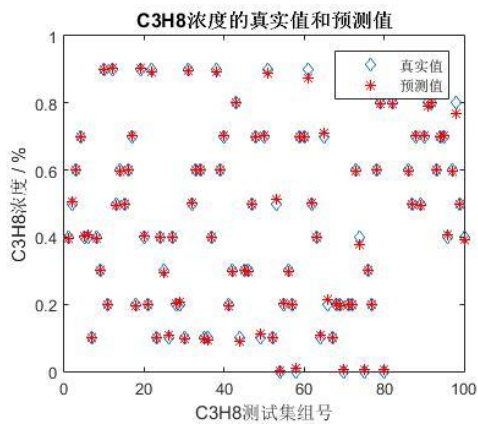


(a) 预测结果

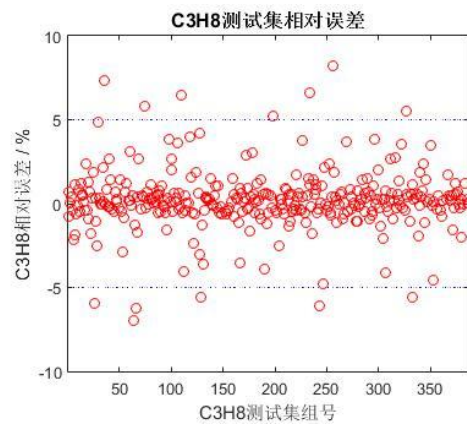


(b) 相对误差

图 4.17 C₂H₆ 测试集改进 PCA-SVM 的预测结果

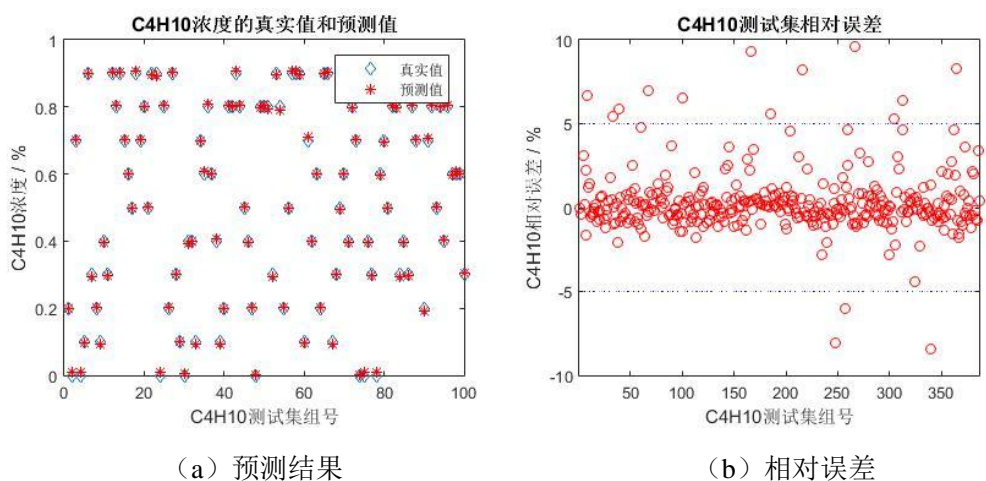
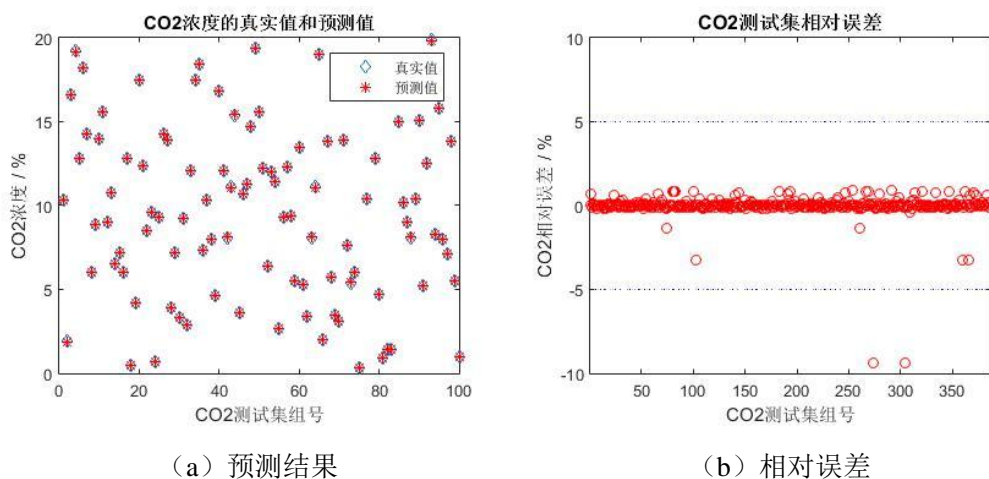


(a) 预测结果



(b) 相对误差

图 4.18 C₃H₈ 测试集改进 PCA-SVM 的预测结果

图 4.19 C₄H₁₀ 测试集改进 PCA-SVM 的预测结果图 4.20 CO₂ 测试集改进 PCA-SVM 的预测结果

根据改进 PCA-SVM 模型对 CH₄、C₂H₆、C₃H₈、C₄H₁₀、CO₂ 五种气体测试集的预测结果, 计算出每种气体预测浓度的均方差 (*RMSEP*)、相关系数 (*R*)、最大相对绝对误差、平均相对误差、相对误差超过 5% 的比例和预测一个样本的时间如表 4.8 所示。

表 4.8 改进 PCA-SVM 模型对五种混合气体的预测结果

气体种类	<i>RMSEC</i>	<i>R</i>	最大相对绝对误差	平均相对误差	相对误差超过 5% 的比例	预测时间 (T/s)
CH ₄	0.0241	1.0000	4.81%	1.25%	0	0.98
C ₂ H ₆	0.0253	0.9466	7.27%	3.49%	3.89%	1.41
C ₃ H ₈	0.0285	0.9032	8.20%	4.24%	3.63%	1.27
C ₄ H ₁₀	0.0324	0.8573	9.57%	4.39%	3.89%	1.54
CO ₂	0.0287	1.0000	9.37%	0.89%	0.52%	0.93

由表 4.8 分析可知, 改进 PCA-SVM 模型总体的预测效果并不是太理想, 主要是 C₂H₆、C₃H₈、C₄H₁₀ 三种低浓度烃类气体的预测结果的最大相对绝对误差和平均相对误差都比较大, 相对误差超过 5% 的比例分别为 3.89%, 3.63% 和 3.89%。CH₄ 模型的预测精度和稳定性和改进 PEM-SVM 模型相比都有大幅提升。CO₂ 模型的预测结果的平均相对误差有所减少, 说明预测的稳定性有所提高, 但是根据图 4.20 可以看出, 有两个样本的预测结果的相对误差很大, 原因可能是在前期的样本异常处理没有将这两个异常样本检测出来, 除去这两个异常点, CO₂ 模型预测结果的最大相对误差只有 3.28%, 预测精度很高。

C₂H₆、C₃H₈ 和 C₄H₁₀ 的相对误差分布比较分散在 -10%~10% 之间, 预测结果比较差, 分析其原因可能有: 1) 这三种气体的配气浓度范围都是在 0~1% 之间, 属于低浓度气体, 在配置低浓度气体时, 由于配气装置中的气体质量流量计的精度问题或者在配气时这三种气体没有被完全充入到气体池中导致浓度偏差; 2) C₂H₆、C₃H₈ 和 C₄H₁₀ 这三种气体的光谱严重交叉重叠, 经 PCA 特征处理后的数据可能加剧了交叉重叠区域的复杂度, 使得这三种气体的预测精度下降。

4.4 结果分析与讨论

由表 4.3 可知, 经过 PSO 算法对 PEM-SVM 模型参数寻优所建立的模型, 在预测精度上要比普通的 PEM-SVM 模型的预测结果要精确很多, 所以本章主要采用 PSO 算法对 SVM 模型的参数进行寻优, 建立了改进的 PEM-SVM 模型和改进的 PCA-SVM 模型。两个模型的预测结果分别如表 4.6 和表 4.8 所示, 可以发现改进 PEM-SVM 模型在预测 C₂H₆、C₃H₈ 和 C₄H₁₀ 三种低浓度气体效果比

较好,改进 PCA-SVM 模型在预测 CH_4 和 CO_2 两种高浓度气体的结果的精确度比较高,所以可以集成两种建模方法的优点,即整合两个模型构成 PEM-PCA-SVM 模型,在预测 C_2H_6 、 C_3H_8 和 C_4H_{10} 三种低浓度气体的时候使用改进 PEM-SVM 模型,在预测 CH_4 和 CO_2 两种高浓度气体的时候用改进 PCA-SVM 模型,则其预测结果如表 4.9 所示。

表 4.9 PEM-PCA-SVM 模型对五种混合气体的预测结果

气体 种类	<i>RMSEC</i>	<i>R</i>	最大相对 绝对误差	平均相 对误差	相对误差 超过 5% 的比例	预测时间 (T/s)
CH_4	0.0241	1.0000	4.81%	1.25%	0	0.98
C_2H_6	0.0083	0.9964	4.66%	2.20%	0	1.43
C_3H_8	0.0127	0.9325	4.98%	3.94%	0	0.99
C_4H_{10}	0.0091	0.9699	4.57%	3.16%	0	1.34
CO_2	0.0287	1.0000	3.28%	0.89%	0	0.93

由表 4.9 可知,PEM-PCA-SVM 模型对五种混合气体的综合预测效果比改进 PEM-SVM 模型和改进 PCA-SVM 模型都要好,最大相对绝对误差都不超过 5%,预测精度符合实际的要求。

4.5 本章小结

本章利用小波变换对原始光谱数据进行减噪处理,然后采用两种特征提取方法即峰值提取法(PEM)和主成分分析法(PCA)对减噪后的数据进行特征提取和降维,最后建立基于 SVM 为基础的矿井下气固两相流中气体组分的定量分析模型。通过以甲烷(CH_4)为例,采用不同参数寻优方法,对 SVM 模型的预测精度进行分析,确定了 PSO 算法对 SVM 的惩罚参数 c 和核函数参数 g 进行寻优,建立了基于 PSO 算法的改进 PEM-SVM 模型和改进 PCA-SVM 模型。结果表明,在预测 C_2H_6 、 C_3H_8 和 C_4H_{10} 三种低浓度气体时,改进 PEM-SVM 模型的预测精度较高,在预测 CH_4 和 CO_2 两种高浓度气体时,改进 PCA-SVM 模型预测的精度较高,最后整合两个模型的优点,建立了 PEM-PCA-SVM 集成模型。

5 基于极限学习机的定量分析

上一章主要介绍了基于 SVM 气固两相流中气体组分的定量分析研究，SVM 是一种小样本分析方法，当分析样本较大时，如果要保证预测精度，就要对参数进行寻优进行训练，同时由于 SVM 模型输出为一维，所以建模时间较长。本章引入了极限学习机 (ELM) 对矿井下气固两相流中气体组分进行定量分析，并对其进行了优化改进，建立了基于 D-ELM 的定量分析模型，使用测试集的样本对已建立的模型进行验证，并对其预测结果进行分析。

5.1 基于 ELM 气固两相流中气体组分的定量分析

经 4.2.1 中 PEM 处理得到的特征光谱矩阵 X 维度为 1386×75 和对应的浓度矩阵 C 维度为 1386×5 ，将特征光谱矩阵 X 的前 1000 个数据作为训练集样本 X_{train} ，后 386 个数据作为测试集样本 X_{test} 。将训练集样本 X_{train} 的 75 个特征值作为 ELM 的输入，对应的 CH_4 、 C_2H_6 、 C_3H_8 、 C_4H_{10} 和 CO_2 的预测浓度作为 ELM 的输出，即 5 维输出。不失一般性，本文设 ELM 的隐含层节点个数为 2000，即 ELM 的网络结构为 $75-2000-5$ 。表 5.1 为模型的预测结果。

表 5.1 ELM 模型对五种混合气体的预测结果

气体 种类	$RMSEC$	R	最大相对 绝对误差	平均相 对误差	预测时间 (T/s)
CH_4	0.0541	0.8731	9.98%	5.26%	0.82
C_2H_6	0.0253	0.7364	11.96%	6.19%	1.33
C_3H_8	0.0169	0.8025	9.68%	5.76%	0.91
C_4H_{10}	0.0226	0.7558	10.98%	6.22%	1.21
CO_2	0.0407	0.9023	9.04%	4.28%	0.79

由表 5.1 可知，ELM 模型的预测效果很差，其中平均相对误差都在 5% 以上，最大相对绝对误差达到了 11.96%，所以这种建模结果不可取。

5.2 基于 D-ELM 气固两相流中气体组分的定量分析

5.2.1 基于 PEM-D-ELM 气固两相流中气体组分的定量分析

PEM-D-ELM 模型的输入输出和 PEM-ELM 模型相同,不同是只是 D-ELM 的隐含层神经元节点是可以变化的,将特征光谱矩阵 X 的前 1000 个数据作为训练集样本 X_{train} ,后 386 个数据作为测试集样本 X_{test} 。为了提高训练速度,本文设 D-ELM 的初始隐含层节点个数为 1000,最大隐含层节点 L_{max} 为 8000,每次增加的隐节点 δL 为 2,期望误差 ξ 为 0.005。最终模型的最佳隐含节点数为 5486,即 D-ELM 网络结构为 75—5486—5,建模总时间约为 20.37 秒。

为了验证已经建立 PEM-D-ELM 气固两相流中气体组分的定量分析模型,需要用测试集样本 X_{test} 。对 PEM-D-ELM 模型进行验证。利用已建立好的 PEM-D-ELM 模型对测试集样本 X_{test} 浓度进行预测,将测试样本的前 100 个预测结果如图 5.1~5.5 的 (a) 所示,测试样本的相对误差如图 5.1~5.5 的 (b) 所示。

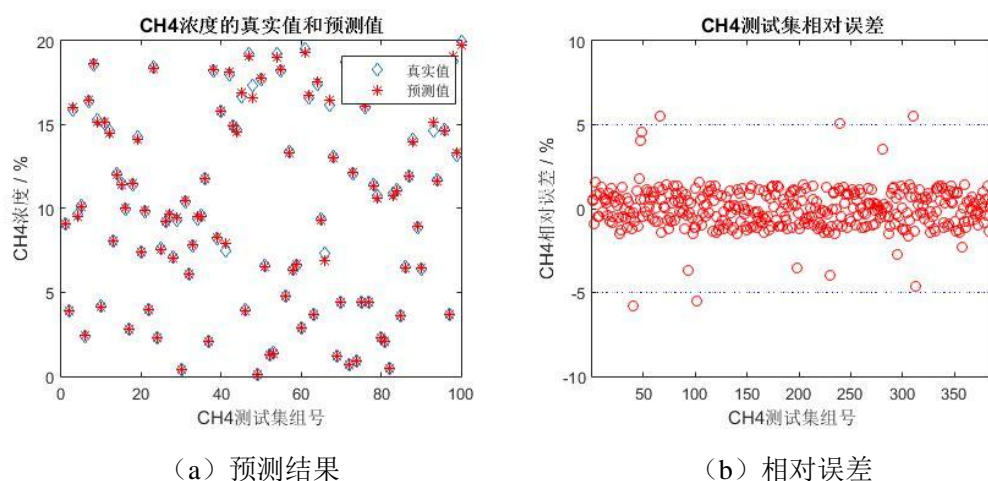
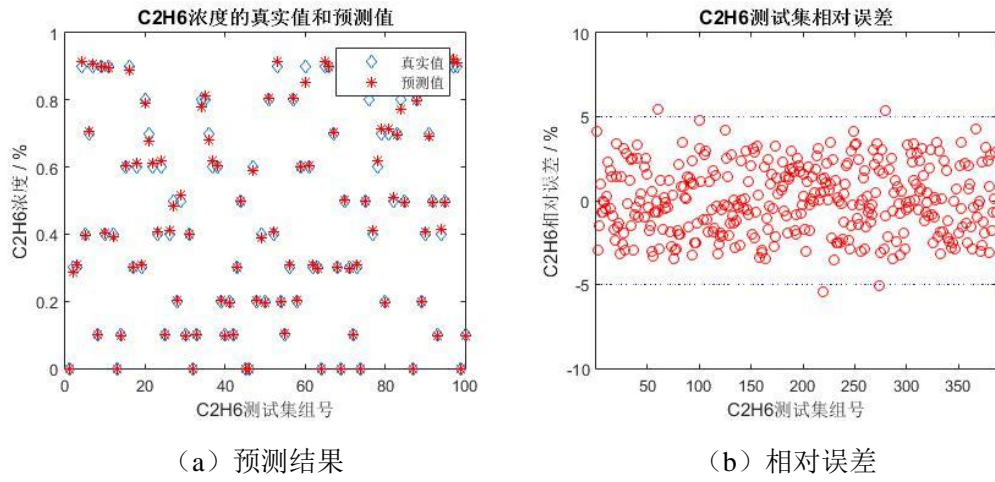
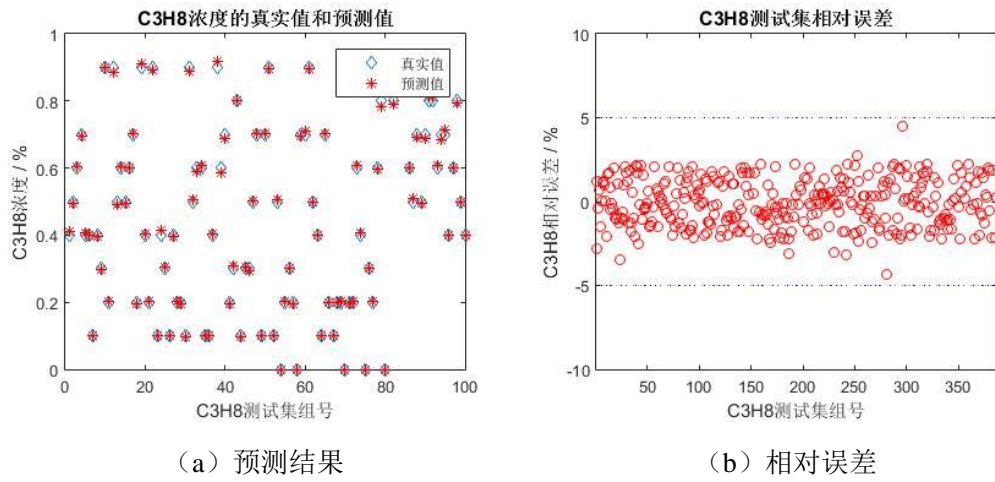
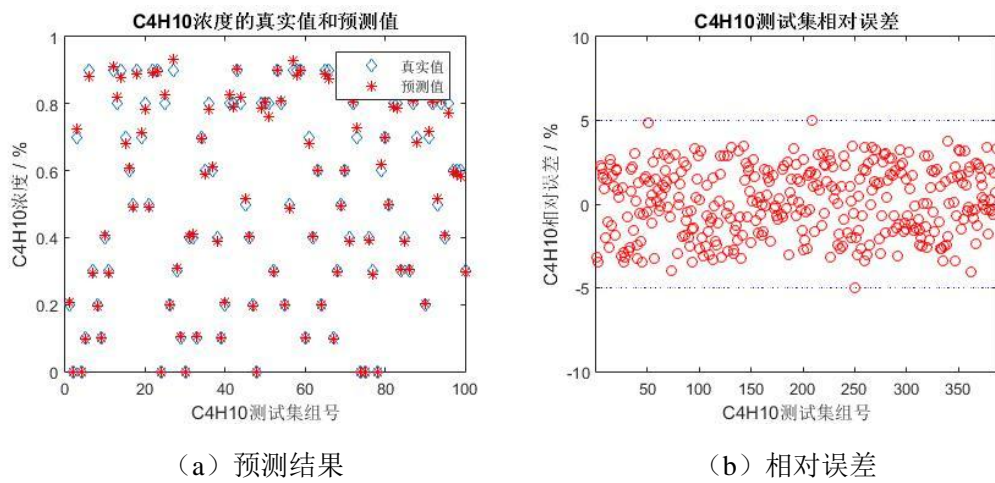
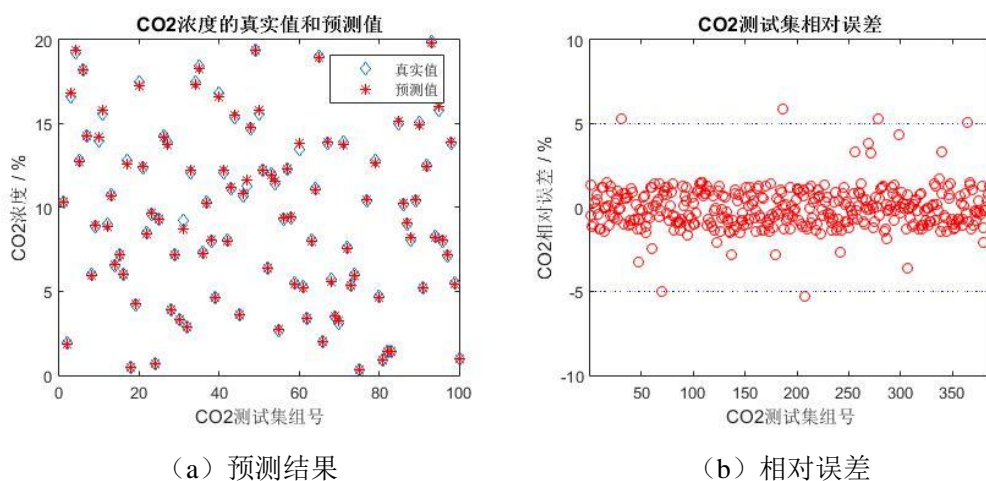


图 5.1 CH_4 测试集 PEM-D-ELM 的预测结果

图 5.2 C_2H_6 测试集 PEM-D-ELM 的预测结果图 5.3 C_3H_8 测试集 PEM-D-ELM 的预测结果图 5.4 C_4H_{10} 测试集 PEM-D-ELM 的预测结果

图 5.5 CO₂ 测试集 PEM-D-ELM 的预测结果

根据 PEM-D-ELM 模型对 CH₄、C₂H₆、C₃H₈、C₄H₁₀、CO₂ 五种气体测试集的预测结果，计算出每种气体预测浓度的均方差 (*RMSEP*)、相关系数 (*R*)、最大相对绝对误差、平均相对误差、相对误差超过 5% 的比例和预测一个样本的时间如表 5.2 所示。

表 5.2 PEM-D-ELM 模型对五种混合气体的预测结果

气体 种类	<i>RMSEC</i>	<i>R</i>	最大相对 绝对误差	平均相 对误差	相对误差 超过 5% 的比例	预测时间 (<i>T/s</i>)
CH ₄	0.0472	0.8876	5.79%	1.87%	1.30%	0.96
C ₂ H ₆	0.0106	0.9146	5.46%	2.73%	1.04%	1.32
C ₃ H ₈	0.0079	0.9549	4.52%	2.84%	0	0.87
C ₄ H ₁₀	0.0124	0.9267	5.01%	3.36%	0.52%	1.31
CO ₂	0.0107	0.9873	5.84%	0.91%	1.55%	0.93

由表 5.2 分析可知，PEM-D-ELM 模型总体的预测效果比较理想，其中 CH₄ 和 CO₂ 的预测结果的最大相对误差和平均相对误差都比较大，最大相对绝对误差都不超过 6%，相对误差超过 5% 的比例占总样本的 1.30% 和 1.55%，但是其误差主要分布主要是在 -3%~3% 之间，故平均相对误差不超过 4%，C₃H₈ 的相对误差也是主要分布在这个区间。

5.2.2 基于 PCA-D-ELM 气固两相流中气体组分的定量分析

经 4.3.3 章中 PCA 处理得到的特征光谱矩阵 \mathbf{T} 维度为 1386×3 和对应的浓度矩阵 \mathbf{C} 维度为 1386×5 ，将特征光谱矩阵 \mathbf{T} 的前 1000 个数据作为训练集样本 $\mathbf{T}_{\text{train}}$ ，后 386 个数据作为测试集样本 \mathbf{T}_{test} 。将训练集样本 $\mathbf{X}_{\text{train}}$ 的 3 个特征值作为 PCA-D-ELM 的输入，对应的 CH_4 、 C_2H_6 、 C_3H_8 、 C_4H_{10} 和 CO_2 的预测浓度作为 ELM 的输出，即 5 维输出。不失一般性，设 PCA-D-ELM 的初始隐含层节点个数为 1000，最大隐含层节点 L_{\max} 为 8000，每次增加的隐节点 δL 为 2，期望误差 ξ 为 0.005。最终模型的最佳隐含节点数为 4658，即 D-ELM 网络结构为 3—4658—5，建模总时间约为 10.86 秒。

为了验证已经建立 PCA-D-ELM 气固两相流中气体组分的定量分析模型，需要用测试集样本 \mathbf{T}_{test} 。对 PCA-D-ELM 模型进行验证。利用已建立好的 PCA-D-ELM 模型对测试集样本 \mathbf{T}_{test} 浓度进行预测，将测试样本的前 100 个预测结果如图 5.6~5.10 的 (a) 所示，测试样本的相对误差如图 5.6~5.10 的 (b) 所示。

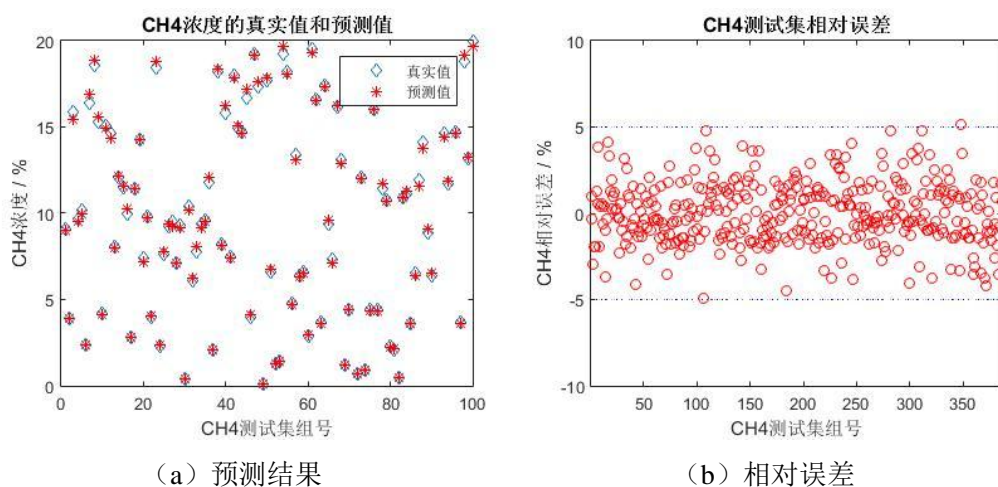
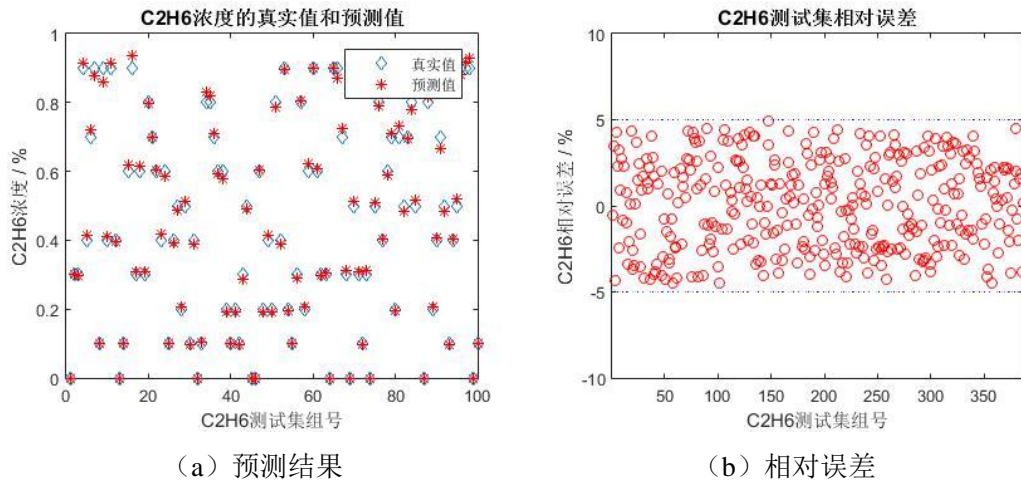
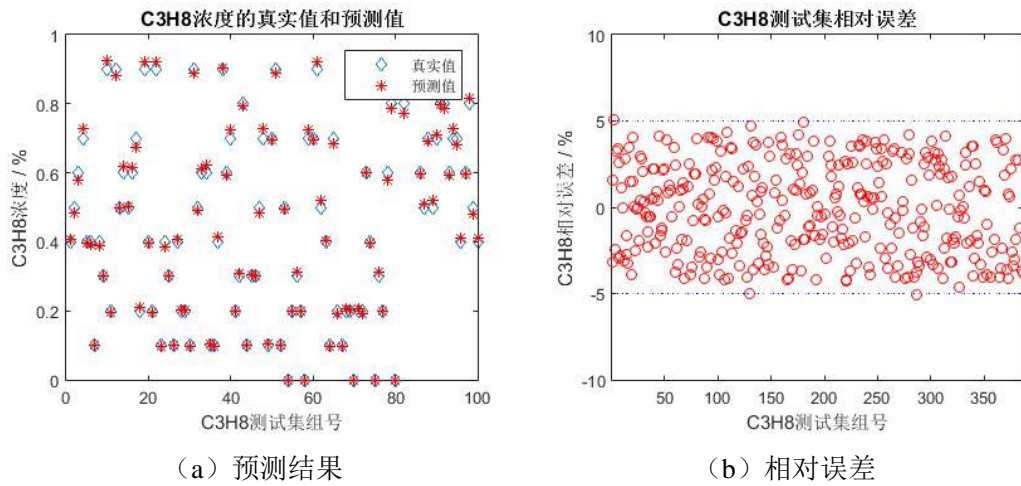
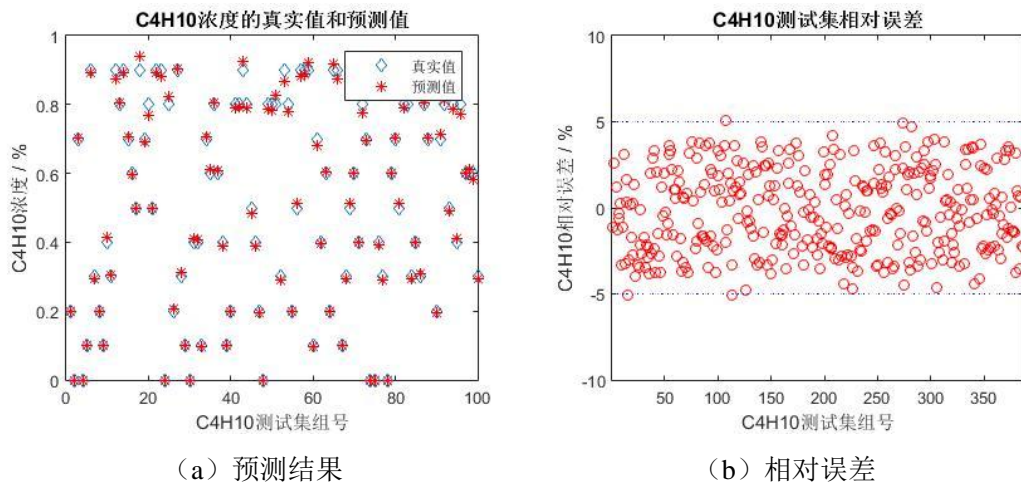
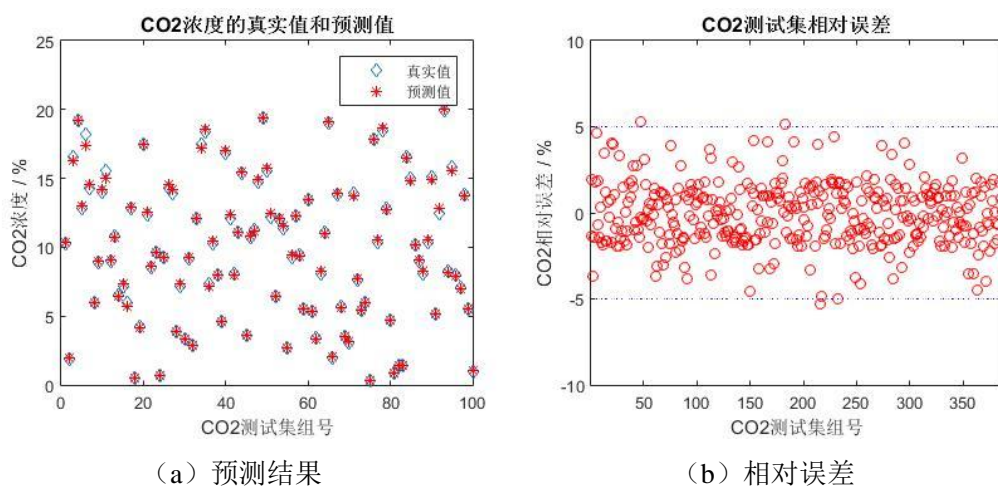


图 5.6 CH_4 测试集 PCA-D-ELM 的预测结果

图 5.7 C_2H_6 测试集 PCA-D-ELM 的预测结果图 5.8 C_3H_8 测试集 PCA-D-ELM 的预测结果图 5.9 C_4H_{10} 测试集 PCA-D-ELM 的预测结果

图 5.10 CO₂ 测试集 PCA-D-ELM 的预测结果

根据 PCA-D-ELM 模型对 CH₄、C₂H₆、C₃H₈、C₄H₁₀、CO₂ 五种气体测试集的预测结果，计算出每种气体预测浓度的均方差 (*RMSEP*)、相关系数 (*R*)、最大相对绝对误差、平均相对误差、相对误差超过 5% 的比例和预测一个样本的时间如表 5.3 所示。

表 5.3 PCA-D-ELM 模型对五种混合气体的预测结果

气体种类	<i>RMSEC</i>	<i>R</i>	最大相对绝对误差	平均相对误差	相对误差超过 5% 的比例	预测时间 (<i>T/s</i>)
CH ₄	0.0122	0.9876	5.11%	1.44%	0.26%	1.16
C ₂ H ₆	0.0086	0.9886	4.93%	2.26%	0	1.04
C ₃ H ₈	0.0113	0.9327	5.08%	3.78%	0.26%	0.92
C ₄ H ₁₀	0.0132	0.9183	5.09%	3.85%	0.78%	1.24
CO ₂	0.0131	0.9756	5.27%	1.45%	0.78%	0.87

由表 5.3 分析可知，PCA-D-ELM 模型对测试集样本预测的结果综合来看，要比 PEM-D-ELM 模型好一些。CH₄ 和 CO₂ 的预测结果的最大相对绝对误差依旧超过 5%，但是比例有所下降，C₂H₆ 的预测效果提升显著。

5.4 结果分析与讨论

由表 5.1、表 5.2 和表 5.3 对比可知，ELM 建立的气固两相流中气体组分的定量分析模型要比基于 D-ELM 建立的模型在预测结果的精度上要差很多，所

以本章主要采用基于 D-ELM 模型建立了 PEM-D-ELM 和 PCA-D-ELM 模型。两个模型的预测结果分别如表 5.2 和表 5.3 所示，可以发现 PEM-D-ELM 模型 C_3H_8 和 C_4H_{10} 两种气体的预测效果比较好，PCA-D-ELM 模型的 CH_4 、 C_2H_6 和 CO_2 三种气体的预测结果的精度比较高，所以也可以整合两个模型的优势，将 PEM-D-ELM 和 PCA-D-ELM 模型集成 PEM-PCA-ELM 模型，在预测 C_3H_8 和 C_4H_{10} 两种气体的时候调用 PEM-D-ELM 模型，在预测 CH_4 、 C_2H_6 和 CO_2 三种气体的时候调用 PCA-D-ELM 模型，集成后的模型的预测结果如表 5.4 所示。

表 5.4 PEM-PCA-ELM 模型对五种混合气体的预测结果

气体 种类	<i>RMSEC</i>	<i>R</i>	最大相对 绝对误差	平均相 对误差	相对误差 超过 5% 的比例	预测时间 (T/s)
CH_4	0.0122	0.9876	5.11%	1.44%	0.26%	1.16
C_2H_6	0.0086	0.9886	4.93%	2.26%	0	1.04
C_3H_8	0.0079	0.9549	4.52%	2.84%	0	0.87
C_4H_{10}	0.0124	0.9267	5.01%	3.36%	0.52%	1.31
CO_2	0.0131	0.9756	5.27%	1.45%	0.78%	0.87

由表 5.4 所示，PEM-PCA-ELM 集成模型整合了 PEM-D-ELM 和 PCA-D-ELM 模型的优势，使得预测的综合效果比单个模型的预测效果要好，但是由于 CH_4 和 CO_2 的浓度范围比较大，其最大相对绝对误差都超过了 5%，超过的比例分别为 0.26% 和 0.78%。

将第四章中 PEM-PCA-SVM 气固两相流中气体组分的定量分析模型的预测结果和 PEM-PCA-ELM 模型的预测结果相比，如表 5.5 所示。由表可知 PEM-PCA-SVM 模型的综合预测效果比 PEM-PCA-ELM 的综合效果要好，PEM-PCA-SVM 模型的对五种气体的预测结果的最大相对误差都不超过 5%，其中 CH_4 、 C_2H_6 、 C_4H_{10} 和 CO_2 的最大相对误差及其平均相对误差都小于 PEM-PCA-ELM 模型的预测结果，即 PEM-PCA-SVM 对这四种气体的预测精度要高于 PEM-PCA-ELM 模型的预测精度。PEM-PCA-ELM 模型预测 CH_4 、 C_4H_{10} 和 CO_2 三种气体的结果的最大相对误差分别为 5.11%、5.01% 和 5.27%，相对误差超过 5% 所占测试集样本比例为 0.26%、0.52% 和 0.78%。在建模时间方面 PEM-PCA-ELM 明显占有优势，所用总时间约为 14.66 秒，PEM-PCA-ELM 建模时间需要约 6271 秒。根据表 4.9 和表 5.4 可知，建立好的模型预测一个样本

的时间在 1 秒左右。

表 5.5 PEM-PCA-SVM 和 PEM-PCA-ELM 模型结果比较

气体 种类	建模 方法	$RMSEC$	R	最大 相对 绝对 误差	平均 相对 误差	相对误 差超过 5% 的比例	建模 时间 (T/s)
CH ₄	PEM-PCA-SVM	0.0241	1.0000	4.81%	1.25%	0	2520
	PEM-PCA-ELM	0.0122	0.9876	5.11%	1.44%	0.26%	2.93
C ₂ H ₆	PEM-PCA-SVM	0.0083	0.9964	4.66%	2.20%	0	289
	PEM-PCA-ELM	0.0086	0.9886	4.93%	2.26%	0	2.93
C ₃ H ₈	PEM-PCA-SVM	0.0127	0.9325	4.98%	3.94%	0	325
	PEM-PCA-ELM	0.0079	0.9549	4.52%	2.84%	0	2.93
C ₄ H ₁₀	PEM-PCA-SVM	0.0091	0.9699	4.57%	3.16%	0	348
	PEM-PCA-ELM	0.0124	0.9267	5.01%	3.36%	0.52%	2.93
CO ₂	PEM-PCA-SVM	0.0287	1.0000	3.28%	0.89%	0	2789
	PEM-PCA-ELM	0.0131	0.9756	5.27%	1.45%	0.78%	2.93

5.5 本章小结

本章利用极限学习机对矿井下气固两相流中气体组分进行定量分析，采用两种特征提取方法即峰值提取法（PEM）和主成分分析法（PCA）对小波降噪后的数据进行特征提取和降维处理，提高了建模效率，并对极限学习机算法做出了改进，建立了基于 D-ELM 的矿井下气固两相流中气体组分的定量分析模型，即 PEM-D-ELM 模型和 PCA-D-ELM 模型，并整合了两个模型的优势，建立了集成 PEM-PCA-ELM 模型。使用测试集样本对 PEM-PCA-ELM 模型进行测试验证并和 PEM-PCA-SVM 模型进行比较，发现 PEM-PCA-ELM 在建模速度虽然快，但是预测精度方面不及 PEM-PCA-SVM 模型。

6 总结与展望

6.1 总结全文

矿井下的两相流中的两相即为易燃易爆气体和煤粉颗粒，给井下工作人员的生命健康和安全带来极大威胁。因此，排除井下空气中煤粉颗粒的干扰实现高效、快速以及实时的矿井气体在线检测，对于井下作业人员的生命安全和防止瓦斯爆炸事故的发生都有重要意义。

本文基于红外光谱法对矿井下气固两相流中气体组分浓度的在线测量方法进行了研究，研发了高精度配气装置，模拟井下真实环境，搭建了采集气体红外光谱数据的实验系统，设计合理的配气方案和实验方案，进行实验并采集气体光谱数据，采用峰值提取法（PEM）和主成分分析法（PCA）对小波降噪后的气体光谱数据进行特征提取和降维处理，建立了基于改进 SVM 和基于 D-ELM 的矿井下气固两相流中气体组分的定量分析模型，并利用测试集样本对所建立的模型进行测试验证，通过比较各个模型的预测结果确定最佳的气固两相流中气体的在线测量方案。本文主要工作总结如下：

1) 研发了准确计量各种气体的量的高精度配气装置，搭建了气固两相流中气体组分红外光谱数据采集系统，并设计了合理的各种气体浓度比的配气方案，采集了 1386 组气固两相流中气体组分的有效光谱数据，其中 1000 组为训练集样本，386 组为测试集样本。实验表明：本实验系统安全可靠，采集气体光谱数据快，稳定性好。

2) 利用小波变换对原始光谱数据进行减噪处理，然后采用两种特征提取方法即峰值提取法（PEM）和主成分分析法（PCA）对减噪后的数据进行特征提取和降维，最后建立基于 SVM 为基础的矿井下气固两相流中气体组分的定量分析模型。通过以甲烷（ CH_4 ）为例，采用不同参数寻优方法，对 SVM 模型的预测精度进行分析，确定了 PSO 算法对 SVM 的惩罚参数 c 和核函数参数 g 进行寻优，建立了基于 PSO 算法的改进 PEM-SVM 模型和改进 PCA-SVM 模型。结果表明，在预测 C_2H_6 、 C_3H_8 和 C_4H_{10} 三种低浓度气体时，改进 PEM-SVM 模型的预测精度较高，在预测 CH_4 和 CO_2 两种高浓度气体时，改进 PCA-SVM 模型预测的精度较高，最后整合两个模型的优点，建立了 PEM-PCA-SVM 集成模型，集成模型的预测结果的最大相对绝对误差都不超过 5%，预测一次样本的时间约 1 秒，这表明基于 PEM-PCA-SVM 的气固两相流中气体组分的定量分析

模型能够满足矿井下的气体在线测量的精度和时间要求。

3) 建立了基于极限学习机对矿井下气固两相流中气体组分进行定量分析, 采用两种特征提取方法即峰值提取法 (PEM) 和主成分分析法 (PCA) 对小波降噪后的数据进行特征提取和降维处理, 提高了建模效率, 并对极限学习机算法做出了改进, 建立了基于 D-ELM 的矿井下气固两相流中气体组分的定量分析模型, 即 PEM-D-ELM 模型和 PCA-D-ELM 模型, 并整合了两个模型的优势, 建立了集成 PEM-PCA-ELM 模型。使用测试集样本对 PEM-PCA-ELM 模型进行测试验证并和 PEM-PCA-SVM 模型进行比较, 发现 PEM-PCA-ELM 在建模速度虽然快, 但是预测精度方面不及 PEM-PCA-SVM 模型。

6.2 工作展望

本文通过研发高精度气体配置装置搭建了可靠的实验采集系统, 设计了合理的实验方案, 模拟矿井下环境, 准确的得到了气固两相流中气体组分的红外光谱数据, 并建立了基于改进 SVM 和基于 D-ELM 的矿井下气固两相流中气体组分的定量分析模型, 通过测试集样本对所建立模型进行测试验证, 并对其预测的结果进行比较和分析。但在研究的工作中依然有很多问题和不足之处, 后续值得研究和注意的问题有:

1) 配气装置中的气路连接需要进一步优化, 建立更加完善的实验系统。进一步完善实验方案, 加强实验过程中的每个步骤的控制, 通过对红外光谱仪的气体池中的温度和压强控制, 保证采集得到的数据的重复性。

2) 研究如何更好的模拟矿井下的真实环境, 研究真实矿井下的煤粉颗粒粒径分布和浓度分布。

3) 本文对于异常样本的研究不够深入, 导致有异常样本点没有被准确判断出来, 在后续的工作中要加强对这方面的研究, 以提高数据的准确性和模型的稳健性。

4) 继续优化实验系统中的软件和硬件设备, 将各个模块集成在一块, 并研发方便、快捷、可实时显示的系统界面, 实现气固两相流中气体组分的在线测量。

参考文献

- [1] 袁亮. 我国深部煤与瓦斯共采战略思考[J]. 煤炭学报, 2016, 41(01): 1-6.
- [2] 陈晓坤, 蔡灿凡, 肖旻. 2005-2014 年我国煤矿瓦斯事故统计分析[J]. 煤矿安全, 2016, 47(2): 224-226.
- [3] Pan Z, Connell L D. Modelling permeability for coal reservoirs: a review of analytical models and testing data[J]. International Journal of Coal Geology, 2012, 92: 1-44.
- [4] Liu Q, Cheng Y. Measurement of pressure drop in drainage boreholes and its effects on the performance of coal seam gas extraction: a case study in the Jiulishan Mine with strong coal and gas outburst dangers[J]. Natural hazards, 2014, 71(3): 1475-1493.
- [5] 洪林, 王继仁, 邓存宝等. 煤炭自燃生成标志气体的红外光谱分析[J]. 辽宁工程技术大学学报, 2006, (05): 645-648.
- [6] 张国枢. 通风安全学[M]. 徐州: 中国矿业大学出版社, 2007: 143-167.
- [7] SOMOV A, BARANOV A, APIRJAKIN D. 2014 Sensors Actuators 210 157
- [8] SOMOV A, BARANOV A, APIRJAKIN D, SPIRJAKIN A, et al. Sleptsov V and Passerone R [J]. Sensors Actuators, 2013, 202-217.
- [9] NAGAI D, NISHIBORI M, ITOH T, et al. 2015 Sensors Actuators 206 488.
- [10] Zhang X, Wang S. Design a wireless sensor network for methane monitoring system[C]. 2008 6th IEEE International Conference on Industrial Informatics. IEEE, 2008: 614-618.
- [11] KUIJUN WU, FAQUAN LI, XUEWU CHENG, et al. Sensitive detection of CO₂ concentration and temperature for hot gases using quantum-cascade laser absorption spectroscopy near 4.2 μm [J]. Applied Physics B, 2014, 117:659-666.
- [12] R. M. SPEQRIN, C. S. GOLDENSTEIN, I. A. SCHULTZ, et al. Simultaneous sensing of temperature, CO, and CO₂ in a scramjet combustor using quantum cascade laser absorption spectroscopy[J]. Applied Physics B, 2014, 117:689-698.
- [13] CHARLES M. WYNN, STEPHEN PALMACCI, MICHELLE L. CLARK, et al. High-sensitivity detection of trace gases using dynamic photoacoustic spectroscopy[J]. Optical Engineering, 2014, 53(2):021103.
- [14] GE HAN, WEI GONG, HONG LIN, et al. On-line wavelength calibration of pulsed laser for CO₂ DIAL sensing [J]. Applied Physics B, 2014, 117:1041-1053.
- [15] WEI REN, WENZHE JIANG, NANCY P. SANCHEZ, et al. Hydrogen peroxide detection with quartz-enhanced photo acoustic spectroscopy using a distributed-feedback quantum cascade laser [J]. Applied Physics Letters, 2014, 104:041117.
- [16] 谭松. 应用于痕量气体检测的量子级联激光器研究[D]. 北京: 清华大学, 2014.
- [17] PRZEMYSŁAW STEFANSKI, RAFAL LEWICKI, NANCY P. SANCHEZ, et al. Measurements of carbon monoxide mixing ratios in Houston using a compact high-power CW DFB-QCL-based QEPAS sensor[J]. Applied Physics B, 2014, 117:519-526.
- [18] L. J. LAN, Y. J. DING, Z. M. PENG, et al. Multi-harmonic measurements of line shape under low absorption conditions [J]. Applied Physics B, 2014, 117:543-547.

- [19] D. MONDELAIN, T.SALA, S.KASSI, et al. Broadband and highly sensitive comb-assisted cavity ring down spectroscopy of CO near 1.57 μm with sub-MHz frequency accuracy [J]. *Journal of Quantitative Spectroscopy & Radiative Transfer*, 2015, 154: 35-43.
- [20] JESSICA BARRIENTOS BARRIA, ALEXANDRE DOBROC, HELENE COUDERT-ALTEIRAC, et al. Simultaneous remote monitoring of atmospheric methane and water vapor using an integrated path DIAL instrument based on a widely tunable optical parametric source[J]. *Applied Physics B*, 2014, 117:509-518.
- [21] F. Eigenmann, M. Maciejewski, A. Baiker. Quantitative calibration of spectroscopic signals in combined TG-FTIR system[J]. *Thermochimica Acta*, 2006, (6): 81-92.
- [22] E. Granada, P. Egu á, J. A. Vilan, et al. FTIR quantitative analysis technique for gases[J]. *Application in a biomass thermochemical process. Renewable Energy*, 2012, (41): 416-421.
- [23] K.CHAN, H.ITO, H.INABA. Optical remote monitoring of CH₄ gas using low-loss optical fiberlink and InGaAsp lighting-emitting diode in 1.33 μm region [J]. *Applied Physics Letters*, 1983, 43(7):634-637.
- [24] Makhoukhi N, Pere E, Creff R, et al. Determination of the composition of a mixture of gases by infrared analysis and chemometric methods[J]. *Journal of Molecular Structure*, 2005, 47(3): 855-859.
- [25] Zakia Sebih, Serge Bourbigot, René Delobel, et al. Quantitative analysis of automotive exhaust gas by Fourier Transform Infrared spectroscopy[J]. 2010, 13(2): 188-205.
- [26] Anna A.Stec, Peter Fardell, Per Blomqvist, et al. Quantification of fire gases by FTIR: Experimental characterisation of calibration systems[J]. *Fire Safety Journal*, 2011, 46(5): 225-233.
- [27] RITOBRATA SUR, KAI SUN, JAY B. JEFFRIES, et al. TDLAS-based sensors for in situ measurement of syngas composition in a pressurized, oxygen-blown, entrained flow coal gasifier [J]. *Applied Physics B*, 2014, 116:33-42.
- [28] 王汝琳, 王咏涛. 红外检测技术[M]. 北京: 化学工业出版社, 2006.
- [29] Burrus C S, Gopinath R A, Guo H. Introduction to wavelets and wavelet transforms: a primer[J]. 1997.
- [30] Mallat S G. A theory for multiresolution signal decomposition: the wavelet representation[J]. *IEEE transactions on pattern analysis and machine intelligence*, 1989, 11(7): 674-693.
- [31] Daubechies I. The wavelet transform, time-frequency localization and signal analysis[J]. *IEEE transactions on information theory*, 1990, 36(5): 961-1005.
- [32] 孙桂玲, 张翠兰, 方勇华, 等. 小波变换在光谱特征提取方面的应用[J]. *量子电子学报*, 2006, 23(1): 22-26.
- [33] 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.
- [34] Wang Yuanbin. Infrared Spectrum Analysis of the Gas in Coal Mine Based on SVM[J]. *Intelligent Computing and Intelligent System*, 2009, (1): 608-611.
- [35] 白鹏, 谢文俊, 刘君华. 混合气体红外光谱支持向量机分析的新方法[J]. *光谱学与光谱分析*, 2007, 27(7): 1323-1327.
- [36] 白鹏, 冀捐灶, 张发启, 等. 基于 SVM 的混合气体分布模式红外光谱在线识别方法

- [J]. 光谱学与光谱分析, 2008, 28(10): 2278-2281.
- [37] 杜振辉, 齐汝宾, 张慧敏等. 近红外光谱定量检测丙烷和异丁烷[J]. 天津大学学报, 2008, (05): 589-592.
- [38] 鲁波, 余桂英, 徐立恒等. 中红外光谱法测定有毒气体苯和甲苯的含量[J]. 中国计量学院学报, 2004, (03): 23-26.
- [39] 郝惠敏, 乔聪明, 汤晓君, 等. 核偏最小二乘特征提取在混合气体 FTIR 光谱定量分析中的应用[J]. 红外与毫米波学报, 2009, 28(2): 115-118.
- [40] Patterson D W. Artificial neural networks: theory and applications[M]. Prentice Hall PTR, 1998.
- [41] Yao X. A review of evolutionary artificial neural networks[J]. International journal of intelligent systems, 1993, 8(4): 539-567.
- [42] 孙秀云, 李燕, 王俊德. 人工神经网络法对多组分大气污染物的同时监测[J]. 光谱学与光谱分析, 2003, (04): 739-741.
- [43] 高明亮. 基于傅里叶变换红外光谱技术的多组分气体定量分析研究[D]. 安徽: 中国科学技术大学, 2010.
- [44] 杜涛. 矿井极性混合气体红外光谱定量分析研究 [D]辽宁工程技术大学 , 2011.
- [45] Shi Y, Eberhart R C. Empirical study of particle swarm optimization[C]. Evolutionary Computation, 1999. CEC 99. Proceedings of the 1999 Congress on. IEEE, 1999, 3.
- [46] 张利彪, 周春光, 马铭, 等. 基于粒子群算法求解多目标优化问题[J]. 计算机研究与发展, 2004, 41(7): 1286-1291.
- [47] 震, 纪, 惠连, 等. 粒子群算法及应用[M]. 北京: 科学出版社, 2009.
- [48] 李玉军, 汤晓君, 刘君华. 直接正交信号校正算法在烷烃类多组分气体定量分析中的应用[J]. 光谱学与光谱分析, 2012, 4: 045.
- [49] 张爱菊. 基于数据挖掘技术的瓦斯气体红外光谱定量分析方法的研究[J]. 光谱学与光谱分析, 2013, 33(10): 2646-2650.
- [50] Hartigan J A, Wong M A. Algorithm AS 136: A k-means clustering algorithm[J]. Journal of the Royal Statistical Society. Series C (Applied Statistics), 1979, 28(1): 100-108.
- [51] Krishna K, Murty M N. Genetic K-means algorithm[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 1999, 29(3): 433-439.
- [52] 张建辉. K-means 聚类算法研究及应用[J]. 武汉理工大学, 2007, 4.
- [53] 赵安新, 汤晓君, 张钟华, 等. 利用 2DCOS 进行多组分混合气体傅里叶变换红外光谱分析中同分异构体谱峰的辨别[J]. 光谱学与光谱分析, 2014 (10): 2623-2626.
- [54] 陈媛媛, 王志斌, 王召巴. 思维进化蝙蝠算法及其在混合气体红外光谱特征选择中的应用[J]. 红外与激光工程, 2015, 44(3): 845-851.
- [55] Daubechies I, Bates B J. Ten lectures on wavelets[J]. The Journal of the Acoustical Society of America, 1993, 93(3): 1671-1671.
- [56] Mallat S G. A theory for multiresolution signal decomposition: the wavelet representation[J]. IEEE transactions on pattern analysis and machine intelligence, 1989, 11(7): 674-693.
- [57] 刘美娟, 冯巍巍, 史丰荣, 等. 污染气体红外光谱特征的快速提取与识别[J]. 光谱学与光谱分析, 2006, 26(10): 1854-1857.
- [58] Vapnik V N. The nature of statistical learning theory[M]. New York: Springer Verlag, 1995.

- [59] Eker O F, Camci F, Kumar U. SVM Based Diagnostics on Railway Turnouts[J]. International Journal of Performability Engineering, 2012, 8(8): 289-298.
- [60] HUANG G-B, ZHU Q-Y, SIEW C-K. Extreme learning machine: theory and applications[J]. Neurocomputing, 2006, 70: 489 -501.
- [61] Huang G-B and Chen L. Enhanced random search based incremental learning machine[J]. Neurocomputing, 2008,71(16-18):3460-3468.
- [62] ZHANG R, LAN Y, HUANG G-B, et al. Extreme learning machine with adaptive growth of hidden nodes and incremental updating of output weight[J]. International Conference on Autonomous and Intelligence Systemms, Burnaby Canada, 2011, 6752: 253—262.
- [63] Zhang R., Lan Y.,Huang G. B.,Xu Z. B.. Dynamic extreme learning machine and its approximation capability[J]. IEEE Transactions on System, Man and Cybernetics-Part B, 2012, 99: 1-12.
- [64] WANG W, ZHANG R. Improved convex incremental extreme learning machine based on enhanced random search[M]. Lecture Notes in Electrical Engineering, 2012.
- [65] 张贤达.矩阵分析与应用[M].北京:清华大学出版社,2004.
- [66] Messer S R, Agzarian J, Abbott D. Optimal wavelet denoising for phonocardiograms[J]. Microelectronics journal, 2001, 32(12): 931-941.

作者简介

姓名：沈永，性别：男，民族：汉，出生年月：1990-06，籍贯：江苏省宿城区。

2010-09 至 2014-06 江西理工大学测控技术与仪器系学士；

2014-09 至 2017-06 中国计量大学精密仪器及机械系硕士。

参加项目：

- 1、“矿井瓦斯与矿尘气固两相流多参数在线测量方法研究”，国家自然科学基金，基金编号：51476154。

攻读硕士学位期间发表的学术论文：

沈永, 郭天太, 孔明, 等. 基于 D-ELM 的矿井气体 FTIR 光谱定量分析[J]. 应用光学, 2016, 37(5): 725-729.