

## 摘 要

近红外光谱分析技术作为一种快速、安全、环保的分析技术，在石油化工领域有着广泛的应用与发展前景。由于柴油产品实验室分析测定手段耗时长、费用较高、需要更多的人力物力，满足不了现代工业实时控制的需求。本文利用近红外光谱分析技术，在分析油品组成成分的基础上，利用柴油产品组分的基团和表征结构的参数建立模型，来预测柴油的性质。

近红外光谱分析技术在测定石油石化产品的结构基团有着其独有的优越性，基于C-H基团的近红外光谱段的吸收，构建了表征结构的支链化和烯烃度两个变量，和对应特征结构基团的吸光度数据一并引入到所建模型中。建立柴油性质预测的线性和非线性模型，非线性模型采用L-M算法估计模型参数。所建模型有着清晰的组分含量表征和结构表征，比以往的化学计量学模型含有更多的信息，能为生产过程实时优化提供信息。所建模型的预测效果与以往的多元线性回归和偏最小二乘进行了比较，仿真实验结果表明，由于柴油的黏度、密度、闪点等性质均是由柴油的各组分共同作用的最后体现，与油品自身组成的结构有很大的关系，较之其他模型，引入体现结构特征变量的非线性机理模型有着更高的预测精度和适应性。

**关键词：**近红外；基团；柴油性质；L-M算法；机理模型

# **Mechanism Model Development for Diesel Oil Properties**

## **Based on Near-Infrared Spectroscopy**

Yang Xiaoyan (Control Science and Engineering)

Directed by Prof. Xia Bokai

### **Abstract**

As a fast, safe and environment-friendly analysis technology, near infrared spectrum analysis technology has been widely applied in petroleum and chemical industry with sound prospective. Since the former laboratory analysis and measurement method of diesel oil product is time-consuming , costly, and needs more input of labor and fund, it can not meet the demand of real-time control in modern industry. In this article, near-infrared spectrum analysis technology is utilized to predict nature of diesel oil. On the basis of analyzing constituent of diesel oil, models are created by selecting component groups and structural parameters of diesel oil products.

Near-infrared spectrum analysis technology is uniquely superior in measuring structure groups of petroleum and petrochemical products. Based on absorption of C-H groups of near-infrared spectrum coverage, two variables called attribute structure branched chain and alkenes are constructed and introduced into models together with absorbance data of structural groups corresponded to characteristics. Besides, linear and nonlinear models are created to predict natures of diesel oil, and L-M algorithm is used to predict parameters of nonlinear models. The models have clear characteristics of component groups and structural in diesel oil, and comparing to the former models of Chemical Measurement, more information is contained in the models, so it can provide real-time and optimal information for production process. The models this paper built is compared with the former MLR and PLS models in terms of predicting effect. Since natures of diesel oil including viscosity, density, flash point, etc. are all final manifestation of co-functions by various components of diesel oil, thus are closely related with components structure of oil product itself, the results of the simulation experiment show that the model manifesting structure characteristic variables has higher predicting precision and compatibility compared with other models.

**Key words:** Near infrared; group; Diesel oil products properties; L-M method;  
Mechanism model

## 关于学位论文的独创性声明

本人郑重声明：所提交的论文是本人在指导教师指导下独立进行研究工作所取得的成果，论文中有关资料和数据是实事求是的。尽我所知，除文中已经加以标注和致谢外，本论文不包含其他人已经发表或撰写过的研究成果，也不包含本人或他人为获得中国石油大学（华东）或其它教育机构的学位或学历证书而使用过的材料。与我一同工作的同志对研究所做的任何贡献均已在论文中作出了明确的说明。

若有不实之处，本人愿意承担相关法律责任。

学位论文作者签名： 杨晓燕

日期： 2010年 6 月 9 日

## 学位论文使用授权书

本人完全同意中国石油大学（华东）有权使用本学位论文（包括但不限于其印刷版和电子版），使用方式包括但不限于：保留学位论文，按规定向国家有关部门（机构）送交学位论文，以学术交流为目的赠送和交换学位论文，允许学位论文被查阅、借阅和复印，将学位论文的全部或部分内容编入有关数据库进行检索，采用影印、缩印或其他复制手段保存学位论文。

保密学位论文在解密后的使用授权同上。

学位论文作者签名： 杨晓燕

日期： 2010年 6 月 9 日

指导教师签名： 夏伯精

日期： 2010年 6 月 9 日

## 第 1 章 绪论

### 1.1 研究石油产品性质的重要意义

石油石化工业作为我国的支柱工业,在不断满足日益严格的环保要求,同时降低成本,增加经济收入的前提下,需要对生产过程进行精确的监控。生产过程的监控又需要间接的通过监控过程中的各类物质的多项性能指标或各种化学组成来进行。

石油和烃类的物性研究是进行石油加工和石油化工基础研究、工艺研究、设备放大、流程模拟和过程优化控制等工作不可缺少的基础研究<sup>[1]</sup>,石油产品都是复杂的混合物,各种物性的测量方法也与纯物质测量有很大的区别<sup>[2-3]</sup>。

### 1.2 石油产品的现行标准状况

石油产品的性质如密度、芳烃含量、馏程等是表征产品性能的重要参数<sup>[4-5]</sup>,是石油产品分类的重要质量指标,也是正确、合理使用石油产品的依据。

石油产品中密度的大小不仅反映石油及石油产品相关组分及质量,同时也关系到合同签订买卖双方的经济利益。

运动黏度是表征石油产品物理、储运性能的一项指标。某些石油产品的使用性能和石油产品的黏稠情况密切相关,故此类石油产品在出厂前及重要用户进货入库前,都要先测试产品的运动黏度,以保证安全。

馏程的意义在于可用沸点范围来区别不同的燃料,同时还用来表示燃料中轻重组分的相对含量。表征油品在使用时蒸发性能的好坏。油品馏程包括初馏点、10%馏出温度、50%馏出温度、90%馏出温度和干点这五个项目。初馏点表示燃料中最轻成分的沸点;油品的 50%馏出温度是表示它的平均蒸发性,它能影响发动机的加速性;汽油的 90%馏出温度和干点表示油品中不易蒸发和不能完全燃烧的重质馏分的含量;干点表示燃料中最轻成分的沸点。

十六烷值大小与柴油组分的性质有关,是表征柴油发火性能的指标。柴油的十六烷值低于工作条件要求,会使燃烧延迟和不完全,以致发生爆震,降低发动机功率,增加柴油消耗量。但十六烷值过高,也会使燃烧不完全而发生冒烟现象,并增加柴油消耗量。高速柴油机燃料的十六烷值约为 40-56。大多数的柴油机可采用的十六烷值 40-45。

芳烃含量的高低直接影响燃料的燃烧性能,并对大气污染产生不同程度的影响。炼

油厂通常采用加氢处理来降低催化裂化中的硫含量和芳烃含量，以提高柴油的十六烷值。柴油中芳烃特别是重芳烃的存在对环境危害很大，是形成大气中的微粒和发动机废弃中未燃烃类及氮氧化物的重要来源。一些国家和地区组织最近已出台了一些对石油产品中芳烃含量进行严格限制的新规定。世界燃油规范也对柴油的芳烃含量和多环芳烃含量作出了明确限值<sup>[6]</sup>。

研究表明，在十六烷值一定的情况下，燃料的芳香族含量增加，影响碳氢化合物、氮氧化物、柴油黑烟、粒状物质、可溶性有机成分及苯并芘的排出。随着环保法规要求的日益严格，机车排放标准日益严格，同时对机车燃料中与排放相关的有害物质含量也加以限制。许多国家都对柴油中的芳烃含量提出了更高的要求使燃料中限制的芳烃含量不断下降。

### 1.3 柴油的组成与性质测定

质谱法常用来测定柴油的饱和烃、芳烃和胶质，还可以测定柴油的详细组成。因为饱和烃的链烷烃和环烷烃，芳烃中不同环数的含量分布信息对研究柴油燃烧环境和炼油工艺和催化剂的开发有着重要的意义。但是质谱法有它的致命缺点，耗时较长，对环境和测试人员的身体健康有危害。近红外以其分析周期短、测试费用低、环境友好的优点在 20 世纪 90 年代以后快速发展起来，被广泛应用于很多领域。

柴油的黏度、密度、十六烷值、芳烃含量、凝点等性质已经被前人多采用化学计量的方法和柴油样本的近红外光谱做了不同的关联。比较常见的有多元线性回归和偏最小二乘等。

### 1.4 近红外光谱分析石化产品的理论基础

近红外光谱的产生主要是样品含 X-H(X=C、N、O、S)键基团的化合物在近红外区域基频振动的倍频及合频吸收，与中红外光谱相比，谱带较宽且强度较弱，因此在近红外光谱范围内样品不经稀释可直接在常规样品池中进行测定<sup>[4]</sup>。由于不同基团产生的光谱在吸收峰位和强度上有所不同，随着样品组成的变化其光谱特征也将发生变化，这就为近红外光谱的定性和定量分析奠定了理论基础。其诸多优点，决定了它应用领域的广阔，使其在国民经济发展的许多行业中都能发挥积极作用，并逐渐扮演着不可或缺的角色。主要的应用领域包括：石油及石油化工、基本有机化工、精细化工、冶金、生命科学、制药、医学临床、农业、食品、饮料、烟草、纺织、造纸、化妆品、质量监督、环境保

护、高校及科研院所等。在石化领域可测定石油产品的辛烷值、族组成、十六烷值、闪点、冰点、凝固点、馏程、甲基叔丁基醚（MTBE）含量等。

石油产品主要为烃类化合物,这些化合物含有各种不同的 C-H 基团,这些基团的变化会在石油产品的近红外光谱特征上有所体现<sup>[5]</sup>。石油产品的物理化学性质如十六烷值、馏程、密度等样品的性质,均由石油产品的组成与结构内因所决定。因此,石油产品近红外光谱和样品的物理化学性质之间存在内在函数关系。换句话说,近红外光谱包含了样品的物理化学性质的信息。只要在近红外光谱和物理化学性质之间架起一座桥梁,就可以通过很容易获得的近红外光谱图得到通常使用传统分析方法或标准方法非常费时费力才能得到的被测结果,即样品组成与其物理化学性质数据。方法是对样品的近红外光谱和样品的被测量性质和浓度数据进行数学关联,从而建立两者之间的联系。

由于近红外光谱区主要为 X-H 的泛频振动吸收,所对应的有机化合物组成、结构性性质较单一,在测样技术上又具有简便、快速的特点,使得近红外光谱分析技术可以作为一种研究组成与性质关系的媒介工具。根据我国的实际生产状况和柴油产品性质测量需求现状,本课题以近红外光谱技术为基础,对柴油产品通过近红外光谱检测,依据所测定光谱的具体情况对吸光度数据进行适当的预处理后,建立基于柴油产品组分近红外特征光谱的性质测量模型,以缩短分析周期、降低测试费用、实现柴油性质的准确测量,满足石油工业的计量分析和在线生产控制的需要。

## 1.5 本文研究的主要内容和意义

本文在参考学习过去近红外光谱分析技术和柴油性质测量现状的基础上研究,同时将理论研究与生产过程中的实际需求情况相结合,在近红外预测模型中引入烯烃度和支链化程度,从物质结构的机理上对柴油的性质进行预测,具体包括以下几个方面:

1. 第一章通过大量的中外文献阅读,介绍了石油产品的现行标准状况及柴油组成和性质测定方法,对近红外光谱分析技术及其在石油产品品质分析中的发展与应用研究进行了阐述。

2. 第二章介绍了近红外光谱定量分析技术的理论基础,包括近红外光谱技术的工作原理、数学原理及定量分析的基本定律;介绍了近红外光谱分析技术的常用光谱预处理方法,近红外分析技术中常用的多元校正方法,对非线性模型求解所用的 L-M 算法进行了详细介绍。

3. 第三章基团法建模的理论依据,详细说明了基团振动频率的影响因素及不同烃

类在近红外区段的不同的振动形式。分析了柴油组成及影响柴油性质的一些特征基团和分子结构，并点名本文所用到模型评价标准。

4. 第四至第六章是柴油样本的黏度性质进行定量预测模型的检验。通过一批柴油样品及其黏度、密度、闪点性质数据建立不同的模型，将三种基团法模型与多元线性回归 (MLR) 模型和偏最小二乘 (PLS) 模型进行了比较，非线性基团法预测模型取得了最好的预测效果。

5. 结论与展望部分对本课题的研究内容、研究方法和研究情况作了简要的总结，并阐述了当前研究所存在的问题及今后研究的发展方向。



## 第 2 章 近红外光谱定量分析技术

### 2.1 近红外光谱技术原理

近红外光谱主要是由于分子振动的非谐性使分子振动从基态向高能级跃迁时产生的。近红外光具有吸收作用的物质随着其浓度的改变其光谱特性也将改变。由于近红外光谱含有丰富的物质分子结构信息，所以可用来分析绝大多数种类的化合物及其混和物的成分浓度或者品质参数。随样品成份组成或者结构的变化，其特征光谱也将发生变化，这是近红外光谱分析技术的理论基础。表 2-1 是 C-H 键在近红外光谱区的特征吸收谱带<sup>[7]</sup>。

表 2-1 各种含氢基团在 NIR 区内的特征吸收谱带 (nm)

Table 2-1 The typical absorbed spectroscopy of hydrogenous group in NIR (nm)

	甲基C-H	亚甲基C-H	烯烃C-H	芳烃C-H
组合频	2250~2360	2290~2450	2120~2140	2150, 2460
一级倍频	1695, 1705	1725, 1765	1620~1640	1680
组合频	1360, 1435	1395, 1415	1340	1420~1450
二级倍频	1150, 1190	1210	1080~1140	1145
组合频	1015	1053	1040	——
三级倍频	913	934	——	875
四级倍频	745	762	——	——

就柴油而言，柴油主要为含有各种不同的 C-H 基团信息的烃类化合物，例如甲基-CH<sub>3</sub>(1150 和 1190 nm)、烯烃-C=C-H-(1638 nm)、亚甲基-CH<sub>2</sub>-(934 和 1210 nm)和芳烃 C-H(1436 nm)等基团信息。柴油中不同结构的烃类化合物含量变化都会导致近红外光谱的变化，尽管这种变化非常细微，但通过化学计量学方法对光谱数据的处理，便能得到样品组成变化的信息。因此可将柴油的近红外光谱和组成数据，通过合理的谱图预处理（如平滑、微分）和化学计量学校正方法建立起近红外光谱与油品性质间的预测模型，通过建立的预测模型便可快速得到柴油的性质。

#### 近红外光谱定量分析的依据

近红外谱区主要是 C-H、N-H 和 O-H 键的泛频振动吸收，在近红外光谱区能够获得烃类化合物的组成、结构信息。在近红外光谱的定量分析中，待测样品浓度与吸光度的关系符合分光光度定量分析的两个基本定律：朗伯-比尔定律和吸光度加和定律。

朗伯-比尔定律作为光谱定量分析的理论基础，体现了吸光度与溶液浓度之间的线

性关系，是人们在分光光度分析中应用物质体系中某组分对光的吸收来定量分析其含量的理论依据，该定律说明了溶液的吸光度  $A$  与溶液的浓度  $C$  和吸收层的厚度  $b$ （即光通过溶液的路程）的乘积成正比。朗伯-比尔定律的数学表达式为式（2-1）：

$$A = Kbc \quad (2-1)$$

式中： $A$ —吸光度；

$C$ —样品溶液浓度(单位： $\text{mol} \cdot \text{l}^{-1}$ )；

$b$ —吸收层的厚度，测定液体试样时就是光通过溶液的路程(单位： $\text{cm}$ )，当光程固定时它是常数；

$K$ —比例系数，也是吸光物质的摩尔吸光系数(单位： $\text{l} \cdot \text{mol}^{-1} \cdot \text{cm}^{-1}$ )，它与溶液的性质、温度以及入射光波长等因素有关。在近红外光谱分析中，由于处于不同化学环境的相同键结构基团受其相邻化学键和相邻基团的影响，其  $K$  值是不同的。

其中吸光度的大小由入射的单色光强度和透过的光强度的比值决定。通过（2-2）式计算得到：

$$A = \log I/I_0 \quad (2-2)$$

实际应用中，像原油和成品油这样的混合物样品往往含有很多种组分，并且很多组分对近红外光谱的吸收谱带是相互重叠的。即使对于同一类型样品也可能由于压强、温度等物理性质的不同而产生不同的近红外光谱吸收。近红外光谱仪器也会受到环境中一定的噪声、光源微弱基线漂移等因素的影响。因而，一个特定波长的光谱数据很难代表样品中某种组分对近红外光的真实吸收。

并且朗伯-比耳定律有其成立的前提：

- (1) 入射光为平行单色光且垂直照射。
- (2) 吸光物质为均匀非散射体系。
- (3) 吸光质点之间无相互作用。
- (4) 辐射与物质之间的作用仅限于光吸收，无荧光和光化学现象发生。

如果实际测定的吸光度曲线明显偏离比尔-朗伯定律，可以考虑是由下列物理、化学因素造成的。物理因素有：非单色光引起的偏离；非平行入射光引起的偏离；介质不均匀引起的偏离。化学因素有：溶液浓度过高引起的偏离；化学反应（如水解、解离）

引起的偏离；

### 吸光度加和定律

吸光度加和定律说明如果在某一波长 $\lambda$ ，样品溶液中含有多种对光产生吸收的物质，那么该溶液对该波长光的总吸光度应该等于溶液中每一成分的吸光度之线性加和，该定律称为吸光度加和定律。这一定律表示见式（2-3）：

$$A_{\text{总}} = A_1 + A_2 + A_3 + \cdots + A_n = \sum_{i=1}^n A_i \quad (2-3)$$

吸光度加和定律是多组分混合体系对光吸收的重要性质，是对多组分混合体系进行定量分析的理论依据。

由于近红外光谱重叠比较严重，一个特定波长处的光谱数据不仅仅包含了某个基团的信息，还会包含其他的基团信息。因此，对于物质中某组分的定量分析，须取多个波长进行分析。

## 2.2 近红外光谱技术的模型建立

### 2.2.1 数据预处理

近红外光谱往往包含一些与待测样品性质无关的因素带来的干扰，如样品的状态、光的散射、杂散光及一起响应等的影响，导致了近红外光谱的基线漂移和光谱的不重复。因此对原始光谱进行预处理是非常必要的，光谱预处理方法主要包括平滑、扣减、微分、归一化、标准化、多元散射校正、分段多元散射校正。

#### （1）光谱数据的平滑

数据或者信号的平滑是消除小方差、保留大方差的常用信号处理方法之一。大方差信号通常认为是一种有用分析信号，而小方差信号则常常认为是由于随机误差或仪器噪声等引起。目前，应用较多的平滑方法是 Savitzk 与 Golay 提出的多项式平滑方法，亦即卷积平滑，该方法基于最小二乘原理，能够保留分析信号中的有用信息，消除随机噪声。

#### （2）光谱数据的求导

对光谱图求导主要有两个目的，一是降低重叠谱带的干扰，二是消除基线漂移的干扰。在近红外光谱的谱图处理中，求光谱图的导数光谱的主要目的是消除光谱图基线漂移的影响，对于像石油产品这样的液体产品，基线的漂移主要出现在近红外光谱区短波段，也就是文中研究讨论的区段，所以文中采用二阶导数处理原始光谱。

### (3) 中心化变化

数据中心化变化也成均值中心化, 中心化变换就是将原始数据集的每一个元素减去该元素所在列的均值操作。如式(2-4)所示:

$$x'_{ij} = x_{ij} - \bar{x}_j \quad (i=1,2,3,\dots,n; j=1,2,3,\dots,m) \quad (2-4)$$

式(2-4)中  $x_{ij}$  是原始数据集的元素,  $x'_{ij}$  是均值化处理后的元素,  $\bar{x}_j$  是原始数据集  $X$  的第  $j$  列样本的平均值。

经过中心变化法处理后的数据具有每一列的均值都等于零的性质。当样本数据集中数据全都是由光谱数据构成时, 样本数据的变化类型或集中特征基本相同, 大小范围基本一致, 可以采用这个方法。

### (4) 数据正规化变换

数据正规化变化是将原始数据矩阵的各元素减去该元素所在列的最小值后再除以该元素的极差, 极差也就是该列元素最大值和最小值之差。

$$x'_{ij} = \frac{x_{ij} - x_{j,\min}}{x_{j,\max} - x_{j,\min}} \quad (2-5)$$

式(2-5)中  $x_{j,\max}$  和  $x_{j,\min}$  分别是原始数据矩阵中第  $j$  列的最大值和最小值。

这样可将量纲不一、范围不同的原始数据表达为数据值均在 0~1 范围内的数据。适用于同类型同范围大小的原始数据集, 也适用于不同数据类型和范围大小差别较大的数据集。

### (5) 数据标准化处理

数据标准化处理也称均值方差化或方差归一化, 数据标准化是将原始数据矩阵中各元素减去该列元素均值后再除以所在列元素的方差。

$$x'_{ij} = \frac{x_{ij} - \bar{x}_j}{S_j} \quad (2-6)$$

式(2-6)中  $S_j$  和  $\bar{x}_j$  分别是数据矩阵  $X$  的第  $j$  列元素的标准偏差和平均值。

$$S_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2} \quad \bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \quad (2-7)$$

经过标准化预处理的数据权重相同(一列元素), 均值为零, 方差和标准差都是 1。这种方法适用范围很广, 是光谱预处理经常采用的方法。

### (6) 对数变化法

当数据大于零时，对数据集中的每一个元素作对数运算的操作。近红外光谱数据一般不采用此法进行数据预处理，所以不多加介绍。

在近红外光谱数据预处理时，采用数据标准化处理、数据正规化变化和中心化变换预处理数据时，可以按行或列进行。

#### 2.2.2 建模方法介绍

##### 1. 多元线性回归

多元线性回归<sup>[8]</sup> (Multiple Linear Regression, MLR) 方法是早期近红外光谱定量分析常用的定量校正方法。在近红外光谱定量分析中，成分含量与吸光度的关系满足朗伯-比尔定律，也就是满足线性关系。由于朗伯-比尔定律，有式 (2-8)：

$$Y = XB + E \tag{2-8}$$

其中：

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix},$$

$$B = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}, \quad E = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

式中  $Y$  为校正集成分含量矩阵 ( $n \times m$ )，由  $n$  个样本、 $m$  个组分组成； $X$  为校正集光谱矩阵 ( $n \times k$ )，由  $n$  个样本、 $k$  个波长组成； $B$  为回归参数矩阵； $E$  为性质残差矩阵，服从正态分布  $N(0, \sigma^2)$ 。

式 (2-9) 通过矩阵运算得到  $B$  的解为：

$$B = (X^T X)^{-1} X^T Y \tag{2-9}$$

预测集预测性质即可通过预测方程得到。预测方程为式 (2-10)：

$$\hat{Y} = XB \tag{2-10}$$

在多元线性回归中只要知道样品中某些组分的含量，就能建立它的定量模型。但有一个要求就是要选择好对应于被测组分的特征光谱吸收。实际应用中可能有几百个甚至

上千个波长变量，筛选十几个变量的工作是巨大的。此法适用于线性关系特别好的简单体系。

多元线性回归有其局限性：变量的多重相关性会严重影响参数估计，增大模型误差，并破坏模型的稳健性<sup>[9]</sup>。如果把这种方法应用于成分复杂的柴油产品各质量指标的建模时，就需要设法消除多重性的不良影响。首先需要在预先了解成分化学结构和对应吸光度的基础上进行波长的快速有效的筛选，还要做到所选波长尽量不出现相关问题<sup>[10]</sup>。

## 2. 偏最小二乘法

偏最小二乘法<sup>[8]</sup> (Partial Least Squares, PLS)，作为一种多元统计数据分析方法，将因子分析和线性回归结合的方法，可以把超维空间数据向数据变动量大方向投影，即线性坐标系变换，使数据维数减少，但分析体系的信息被安全保留下来。降维后的数据再与分析体系的基础数据回归建立模型。因子分析则使与化学组成信息有关的因子参与回归，无关因子被分离。

偏最小二乘对光谱矩阵  $X$  进行分解，消除无用的噪声信息，同时对成分含量矩阵  $Y$  也做相应的处理，在分解光谱矩阵  $X$  时要考虑成分含量矩阵  $Y$  的影响。具体方法是，先对光谱矩阵  $X$  和成分含量矩阵  $Y$  进行分解，具体模型见式 (2-11)：

$$\begin{aligned} X &= TP + E \\ Y &= UQ + F \end{aligned} \quad (2-11)$$

式中  $T$  和  $U$  分别是  $X$  和  $Y$  矩阵的得分矩阵； $P$  和  $Q$  分别是  $X$  和  $Y$  矩阵的载荷矩阵； $E$  和  $F$  分别是  $X$  和  $Y$  矩阵的 PLS 拟合残差矩阵。

然后将  $T$  和  $U$  作线性回归，见式 (2-12)：

$$\begin{aligned} U &= TB \\ B &= (T^T T)^{-1} T^T Y \end{aligned} \quad (2-12)$$

在预测时，首先根据  $P$  求出位置样品光谱矩阵  $X_{\text{未知}}$  的得分  $T_{\text{未知}}$ ，即可由下式得到成分含量的预测值：

$$Y_{\text{未知}} = T_{\text{未知}} BQ \quad (2-13)$$

此法同时考虑了谱数据和分析体系的性质数据对因子的影响，因此，PLS 算法可获得比多元线性回归更优的结果<sup>[11]</sup>。

在利用偏最小二乘进行多元校正回归模型训练时，要求确定最优的主成分数或者主因子数，这需要有确定它们的方法，也需要有评价校正模型好坏的方法。目前采用交互验证法解决这个问题。

在采用交互验证法来确定最优的主成分数或主因子数时，采用不同数目的特征矢量来构造校正模型，当所采用的特征矢量个数少于或超过实际组分数时，所得的预测残差平方（PRESS, Predictive Residual Error Sum of Squares）较大。特征矢量数过少会导致构造的校正模型没有包含体系中实际存在的组分信息，造成信息的缺失；特征数超过实际存在的独立主分数，则会引入多余的误差，造成模型的预测精度下降。所以，只有在所取的特征矢量数等于体系中实际存在的独立组分数时，PRESS 才最小。因而也可以通过比较不同数目特征矢量的 PRESS，来确定最优的主成分数或者主因子数。

### 3. Levenberg—Marquardt 算法

Levenberg—Marquardt（简称 L-M）算法继承了梯度下降法的全局搜索特性，同时具有高斯-牛顿（Gauss-Newton, 简称 G-N）法的局部快速收敛的特性，并克服了牛顿法不能有效处理奇异和非正定矩阵的弱点<sup>[12]</sup>。

对于为了得到非线性数学模型中未知待定的参数的非线性函数拟合问题，常采用 Gauss-Newton（G-N）法和 Levenberg-Marquardt（L-M）法。

G-N 法和 L-M 法都可以用来确定非线性函数数学模型经验公式中的待定参数。

设  $y$  是自变量  $x_k$  和待定参数  $B$  的非线性函数，表达式形如式（2-14）：

$$y_k = f(x_k, B), k = 1, 2, \dots, n \quad (2-14)$$

式（2-14）中  $x_k$  可以是单一变量或  $l$  个变量； $B$  是单个参数或  $m$  个参数。

$$\left. \begin{aligned} x_k &= (x_{1k}, x_{2k}, \dots, x_{lk}) \\ B &= (b_1, b_2, \dots, b_m) \end{aligned} \right\}$$

在  $b_i^{(0)}$  点附近进行 Taylor 展开，忽略二次及二次以上的高级项。

$$y_k = f(x^k, b_1, b_2, \dots, b_m) = f_{k0} + \frac{\partial f_{k0}}{\partial b_1} \Delta_1 + \dots + \frac{\partial f_{k0}}{\partial b_m} \Delta_m$$

式中  $f_{k0} = f(x_k, b_1^{(0)}, b_2^{(0)}, \dots, b_m^{(0)})$ ， $\Delta_i = b_i - b_i^{(0)}$ ，上式转化为求  $\Delta_i$  的线性方程组。

$$\sum_{j=1}^m a_{ij} \Delta_j = a_{iy}, \quad i = 1, 2, \dots, m \quad (2-15)$$

其中

$$a_{ij} = \sum_{k=1}^n \frac{\partial f_{k0}}{\partial b_i} \frac{\partial f_{k0}}{\partial b_j}, \quad a_{iy} = \sum_{k=1}^n \frac{\partial f_{k0}}{\partial b_i} (y_k - f_{k0}), \quad i, j = 1, 2, \dots, m \quad (2-16)$$

用矩阵形式表示式(2-15)为

$$\begin{bmatrix} \sum_{k=1}^n \frac{\partial^2 f_{k0}}{\partial b_1^2} & \sum_{k=1}^n \frac{\partial^2 f_{k0}}{\partial b_1 \partial b_2} & \dots & \sum_{k=1}^n \frac{\partial^2 f_{k0}}{\partial b_1 \partial b_m} \\ \sum_{k=1}^n \frac{\partial^2 f_{k0}}{\partial b_2 \partial b_1} & \sum_{k=1}^n \frac{\partial^2 f_{k0}}{\partial b_2^2} & \dots & \sum_{k=1}^n \frac{\partial^2 f_{k0}}{\partial b_2 \partial b_m} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k=1}^n \frac{\partial^2 f_{k0}}{\partial b_m \partial b_1} & \sum_{k=1}^n \frac{\partial^2 f_{k0}}{\partial b_m \partial b_2} & \dots & \sum_{k=1}^n \frac{\partial^2 f_{k0}}{\partial b_m^2} \end{bmatrix} \begin{bmatrix} \Delta_1 \\ \Delta_2 \\ \vdots \\ \Delta_m \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^n \frac{\partial f_{k0}}{\partial b_1} (y_k - f_{k0}) \\ \sum_{k=1}^n \frac{\partial f_{k0}}{\partial b_2} (y_k - f_{k0}) \\ \vdots \\ \sum_{k=1}^n \frac{\partial f_{k0}}{\partial b_m} (y_k - f_{k0}) \end{bmatrix}$$

首先人为给定一组参数初值  $b_i^{(0)}$  ( $i=1, 2, \dots, m$ )，解(2-15)线性方程求解出  $\Delta_i$ ，

求出的  $\Delta_i$  用来作初值的修正。

$$b_i^{(1)} = b_i - b_i^{(0)}, \quad i = 1, 2, \dots, m$$

以  $b_i^{(1)}$  作新的初始值，重复以上两步，直到  $\Delta_i$  中最小者小于给定精度 ( $\varepsilon$ ) 为止。

为了防止迭代矩阵  $a_{ij}$  奇异或病态，若在 G-N 迭代格式中增加一个阻尼项  $\mu I$ 。

则式(2-16)变成

$$\left( \sum_{j=1}^m a_{ij} + \mu I \right) \Delta_j = a_{iy} \quad (2-17)$$

这就是 L-M 方法的迭代格式。

L-M 算法实际上是梯度下降法和高斯-牛顿法的结合。当阻尼系数  $\mu=0$  时，为高斯-牛顿法，适用于矩阵收敛的情况，可以减少迭代次数，但是对初值选取的要求比较严格。当不能保证相应的残差平方和比前次小时， $\mu$  应选取较大的值。当  $\mu$  取值很大，则 L-M 算法接近梯度下降法。当  $\mu$  足够大时，总可以保证  $[a_{ij} + \mu I]^{-1}$  是正定的，从而保证其可逆，因此 L-M 算法在求解病态的模型方面有着很好的稳健性。

### 2.3 算法应用范围讨论

近红外光谱预测模型算法应用范围是根据不同的场合来确定的。如果建模的研究目的是对预测集的某一类型性质进行准确预测，可以使用偏最小二乘法和人工神经网络等



进行建模。但是，这类化学计量学定量校正模型的物理意义不明确，所建模型不能提供表征样品成分、结构的各种 C-H 结构特征的贡献的信息。如果研究目的是为了给化工生产过程提供信息，就需要了解样品中不同组成、结构与所预测性质的关系，可以建立一种在多元线性回归或逐步多元线性回归模型的基础上的改进模型，确定对结果影响的固定波长作为模型的光谱矩阵，这样就能从所建立数学模型中表征不同波长下对应的结构基团的系数正负及数值大小，看出不同结构基团对预测性质的贡献的大小和正负影响作用。

## 第3章 柴油性质的基团选择与模型评价标准

### 3.1 理论依据

有机化合物的许多性质都是由其分子间力所决定的，它包括范德华力（色散力、诱导力和静电力）和氢键。有机化合物的凝点、黏度、沸点等都与分子间力的有关。对于由几十上百种单烃化合物组成的柴油来说，它的凝点就是与柴油分子间力的大小有关。

对于碳氢化合物分子间主要是色散力。首先，色散力与分子间距离有关，当分子间距离远时色散力小，从结构上讲，它与碳氢化合物的分子规整性有关。直链结构分子链规整，分子间距离近，色散力大；其次，色散力与分子大小有关，分子越大，色散力越大。

在碳氢化合物中，甲基、亚甲基是饱和烃分子链结构的主要结构基团。甲基基团的增加表明了两方面的结构信息：一是分子支化程度增加，也就是分子规整性降低；二是分子质量较小。这两个结构因素都是使得色散力减小。而亚甲基含量增加表征了与甲基相反的两点结构信息，也就是分子规整性增强；相对分子质量较大。它们都使得色散力增大。因而可以说饱和烃分子的甲基、亚甲基结构基团的相对含量多少表达了该饱和烃色散力大小的信息，也就说明了它们蕴含着柴油凝点高低等性质的信息。

#### 3.1.1 影响基团振动频率的因素

影响基团振动频率的因素有内因和外因两种<sup>[13]</sup>。内因是由分子内部导致变化产生的，是不可改变的。对于同一种基团或化学键，当它所连接的其它基团不同时，其振动频率吸收强度以及谱带的形态都将发生变化。外因是由化合物分子之外的原因引起的，是可以改变的。对于同一种化合物或是混合物，当测定条件（如样品的物理状态，溶剂、温度、浓度等）不同或使用不同档次的仪器时，它的近红外光谱也会发生变化。

本文涉及的是在尽量保证外因不变的情况下研究内因与物质外部性质变化的联系，为近红外光谱法研究产品性质应用于化工生产中提供支持。

#### 3.1.2 炔烃的近红外段振动形式类型

甲基伸缩振动和弯曲振动都分为两种形式<sup>[14]</sup>：反对称伸缩振动、对称伸缩振动（1192nm,4v）。甲基 C-H 键伸缩振动吸收带的分子吸收系数与烷烃的长短无关。弯曲振动吸收带的位置和与甲基相连的原子有关；甲基 C-H 键弯曲振动不会出现振动的偶合<sup>[15]</sup>。

亚甲基 C-H 键的伸缩振动<sup>[16]</sup>形式: 反对称伸缩振动(932nm<sub>4v</sub>, 1214nm<sub>3v</sub>)和对称伸缩振动。伸缩振动是由两个偶合的 C-H 键伸缩振动引起的。烷烃链的增长使亚甲基 C-H 键伸缩振动吸收带的分子吸收系数呈线性增加趋势。亚甲基 C-H 键弯曲振动存在四种形式: 对称(剪式)弯曲振动, 扭曲振动, 里外摇摆振动(偶合), 面内摇摆振动。

烷烃的 C-C 键伸缩振动和 C-C 键弯曲振动也存在弱的近红外吸收谱带<sup>[17]</sup>。

### 3.1.3 烯烃的近红外段振动形式类型

烯烃 C-H 键伸缩振动出现在末端烯基或内嵌烯基中, 但是振动吸收带位置有所区别。

近红外光谱区对烯烃结构分析很有作用, 双键在烯烃分子的位置可以通过近红外光谱的吸收带位置不同得以区分。

烯烃 C-H 弯曲振动形式分为: 面内和面外弯曲振动。C-H 面外弯曲振动吸收带对于确定烯键的存在非常有用, 但是在近红外谱带没有发现相应的吸收谱带。C-H 键面外弯曲振动吸收带是烯烃的特征吸收带, 并且不受烯烃分子中其他振动的共振或偶合的影响<sup>[18]</sup>。

源于芳烃的刚性分子结构, 芳香烃有强且锐的红外吸收谱带。它的振动形式有: C-H 键伸缩振动、面内弯曲振动、面外弯曲振动及环伸缩、弯曲振动。其中 1145nm 处的泛频吸收带位置稳定, 受苯环取代的影响较少。1436nm 是组频吸收带<sup>[19-20]</sup>。

## 3.2 基团的选取

本文实验所用的 NIR 光谱数据取自 Eigenvector Research 公司的网页<sup>[21]</sup>(<http://software.Eigenvector.com/Data/index.>), 分析仪扫描范围为 900nm-1700nm, 光谱间隔为 2nm, 实验数据为 786 个柴油样品的近红外光谱 900nm-1700nm 波段内吸光度数据及其对应的 50%馏程、十六烷值、密度、闪点、凝点、芳烃含量、黏度性质数据<sup>[22]</sup>, 其原始光谱数据图如图 3-1 所示。

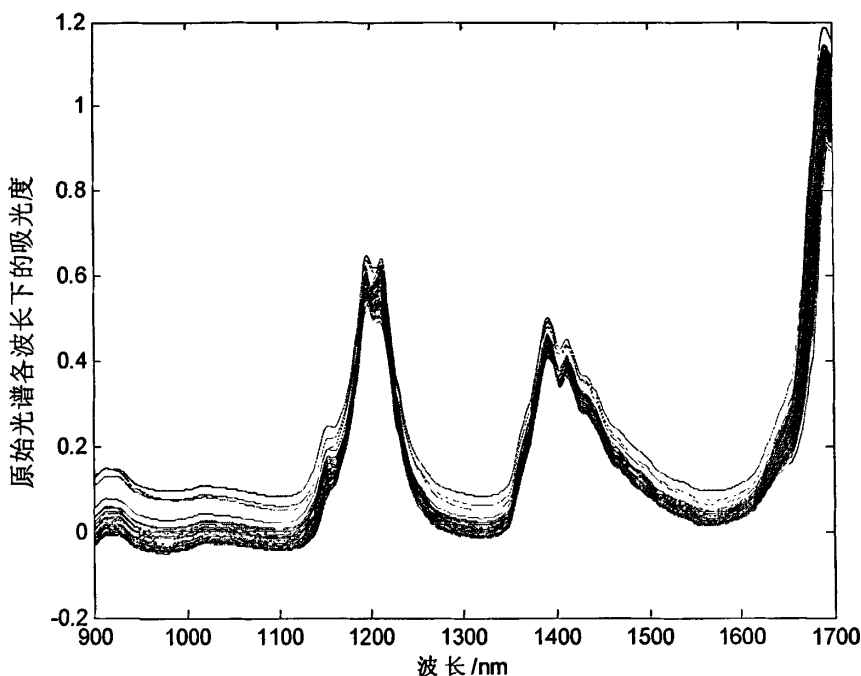


图 3-1 柴油样品的原始近红外光谱图

Fig 3-1 The original NIR spectroscopy of diesel fuel samples

从原始光谱可以看出，事实上柴油原始的近红外光谱各特征谱图相互重叠，对于能表征柴油成分结构的各基团构成，如甲基、亚甲基和烯基等并没有表现出明显的吸收峰来。因此，要想通过柴油原始近红外吸收光谱吸收图分析各基团对应含量或对柴油的构成、性质等进行定量分析，必须对光谱进行预处理。

在有限的光谱数据范围内选取甲基、亚甲基、烯基和芳香基的特征吸收光谱。其中包含甲基：1150nm (3v)、1192nm (3v)、1362nm (组合频)；亚甲基：932nm (4v)、1214nm (3v)、1392nm (组合频)、1414nm (组合频)；烯基：1638nm (2v)；芳香基：1436nm (组合频)

构建六个输入变量：甲基基团、亚甲基基团、烯烃基团、芳烃基团、烷烃支链化、烯烃度。其中支链化度为  $\text{CH}_2/\text{CH}_3$ ，表征了烷烃的支链化程度<sup>[23]</sup>，亦即异构化程度。由于柴油中会含有一部分不饱和烃类，以烯烃和芳香烃为主，引入烯烃度表征油品的不饱和度，也可以为化工生产过程提供一些操作信息。

### 3.2.1 光谱数据预处理

由于近红外光谱仪在采集光谱信号时不可避免的会有背景光的干扰，采用小波去噪对原始光谱进行第一步预处理。

考虑到采样过程中不可避免的会有本色光和光强的基线漂移，尽可能的降低重叠谱带的干扰，又采用九点四次二阶微分对光谱数据进行后续的处理，公式如式（3-1）所示：

$$\left(\frac{\partial^2 x_i}{\partial t^2}\right)_{i=0} = \frac{1}{462} [28(x_{-4} + x_4) + 7(x_{-3} + x_3) - 8(x_{-2} + x_2) - 17(x_{-1} + x_1) - 20x_0] \quad (3-1)$$

经过预处理后，各吸收谱带变得比较分散，方便后续分析。见图 3-2。

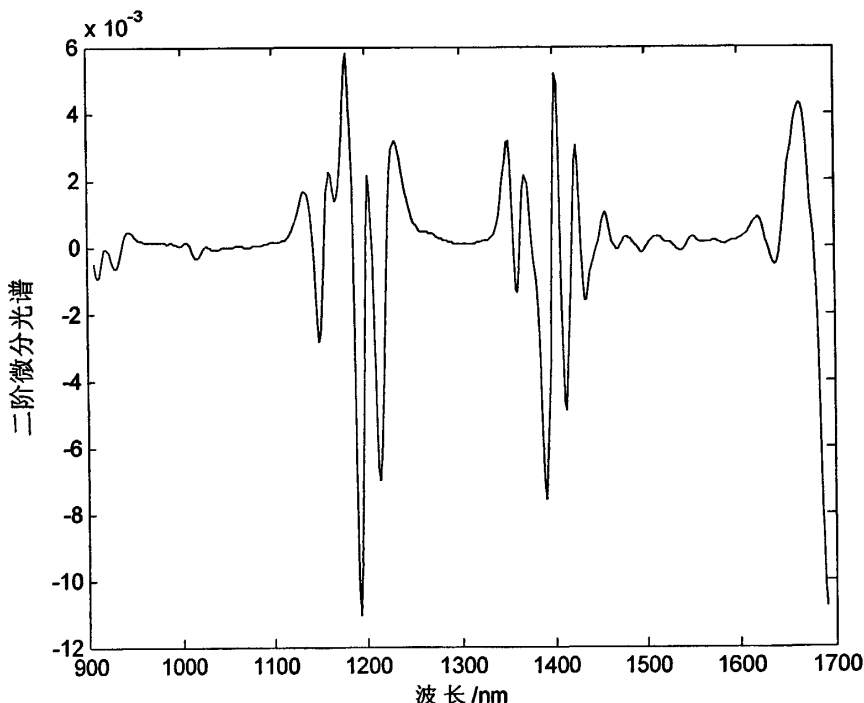


图 3-2 预处理后的柴油样品近红外光谱图

Fig 3-2 The pretreated NIR spectroscopy of diesel oil sample

### 3.2.2 结构基团的选择及构建

柴油成分的组成：（以胜利油田的催化裂化柴油为例）直链烷烃大约占到 55.97%，其中烷烃占 33.40%，烷基占 22.57%；环烷烃占到 9.98%，其中单环环烷烃占 3.19%，双环环烷烃占 1.17%，多环环烷烃占 0.19%，芳缩环烷烃占 5.43%；芳香烃占到 26.75%，其中单环芳烃占 9.17%，双环芳烃占 13.57%，三、四环芳烃占 4.01%；烯烃占到 6.70%，硫化物为 0.6%<sup>[24]</sup>。

同一类型键结构基团所处的分子键结构随着化学环境的不同，样品的组成情况不同，最大吸收峰的位置会发生偏移。柴油组成的特征基团有：甲基、亚甲基、芳基、烯基等，它们的特征吸收见表 2-1。

### 3.3 模型建立

对不同的柴油性质分别用不同的数学方法进行建模,分析比较各种建模的复杂度、计算精度及其对样本组成结构的展现性。文中采取多元线性回归、偏最小二乘、简单选取特征基团回归、组合特征基团及结构回归、L-M 基团法回归了柴油样本的光谱数据和样本理化性质之间的关系。

当光谱变量与所预测性质之间存在线性关系时,可用多元线性回归建立数学模型。简单基团法<sup>[25]</sup>采用自选的特征基团吸收点处的光谱参与模型建立;组合基团法加入了表征产品组成结构的一些变量参与建模。它们的一般形式见式(3-2):

$$y_i = b_0 + b_1 x_{i1} + \dots + b_m x_{im}, \quad i = 1, 2, \dots, n \quad (3-2)$$

但是由于光谱变量是由很多种结构基团共同贡献所得的一种外在显示性质,很在程度上与预测性质变量之间关系非线性的,所得数学表达式中的参数往往是非线性的。针对这种复杂情况,应该采用非线性回归方法回归模型中的待定参数。文中采用 L-M 算法回归拟合参数,所建模型的一般形式见式(3-3):

$$f(x_i) = (a_1 e^{x_1} + a_2 e^{x_2} + a_3 e^{x_3} + a_4 e^{x_4} + a_5 X_5 + a_6 X_6)^y \quad (3-3)$$

### 3.4 模型评价标准

我们如何判断所见模型的好坏和它的适应性,就需要对校正模型有一个客观的评价。在不考虑先期的数据采集和处理的耗资方面,校正模型好坏的唯一评价标准就是预测结果的误差大小,不同的评价表示方法可衍生不同的预测结果误差表达。常用的评价校正模型好坏的统计表达方式有:

#### 1. 残差 $e$ (Error)

$$e = y_i - \hat{y}_i \quad (3-4)$$

#### 2. 相对误差 $re$ (Relative Error)

$$re = (y_i - \hat{y}_i) / y_i \quad (3-5)$$

#### 3. 预测残差平方和 PRESS (Prediction Residual Error Sum of Square)

$$PRESS = \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (3-6)$$

#### 4. 采用标准偏差 SD (Standard Deviation)

$$SD = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-1}} \quad (3-7)$$

其中： $y_i$  是第  $i$  个样本的已知值；

$\hat{y}_i$  是第  $i$  个样本的预测值；

$n$  为预测集样本数目。

## 第4章 柴油黏度性质的预测模型建立

### 4.1 黏度测量方法及应用现状

黏度又称黏性系数、剪切黏度或动力黏度。是流体的一种物理属性，用以衡量流体的黏性。黏度指数是表征柴油黏温性能的重要参数，是正确、合理使用柴油的依据。

各种流体的黏度数据，仅在非平衡实验中才能测量得到。常用的黏度计有毛细管式、落球式、锥板式、转筒式等。在工业上有时用特定形式的黏度计来测定特定的条件黏度。在黏度测定时，实验室直接测得的是黏度和密度的比值，称为运动黏度，常直接用于化工计算。炼油工业中常用恩氏粘度（或恩格拉粘度）作为石油产品的一个指标，它表示某一温度下 200ml 油品与同体积 20℃ 纯水，从恩氏黏度计中流出所需时间之比。恩氏黏度与动力黏度的关系可按经验公式换算。

黏度测定有：动力黏度、运动黏度和条件黏度三种测定方法。

实验室测定黏度的原理一般大都是由斯托克斯公式和泊肃叶公式导出有关黏滞系数的表达式，求得黏滞系数。黏度的大小取决于液体的性质与温度，温度升高，黏度将迅速减小。因此，要测定黏度，必须准确地控制温度的变化才有意义。黏度参数的测定，对于预测产品生产过程的工艺控制、输送性以及产品在使用时的操作性，具有重要的指导价值，在印刷、医药、石油、汽车等诸多行业有着重要的意义。

石油产品的运动黏度和动力黏度是可以通过《GB/T 265 石油产品运动粘度测定法和动力粘度计算法》得到。

在温度  $t$  时，试样的运动黏度  $\nu_i$  ( $\text{mm}^2/\text{s}$ ) 按下式(4-1)计算：

$$\nu_i = c\tau_i \quad (4-1)$$

式中： $c$  为黏度计常数， $\text{mm}^2/\text{s}^2$ ； $\tau_i$  为式样的平均流动时间， $\text{s}$ 。

在温度  $t$  时，试样的动力黏度  $\eta_i$  ( $\text{mPa}\cdot\text{s}$ ) 按下式(4-2)计算：

$$\eta_i = \nu_i \cdot \rho_i \quad (4-2)$$

式中： $\nu_i$  为温度  $t$  时，试样的运动黏度， $\text{mm}^2/\text{s}$ ； $\rho_i$  为与测量运动黏度同温下的试样的密度， $\text{g}/\text{cm}^3$ 。

条件黏度可以通过图表进行换算，也可以通过运动黏度换算得到，我国一般不采用



条件黏度表征油品性质，在此不做详细介绍。

在缺少黏度实验数据时，可按理论公式或经验公式估算黏度。

对于固体悬浮液的粘度，爱因斯坦从流体力学理论上导出低浓度悬浮体的粘度  $\mu_s$  与体积分数  $\phi_B$  的关系，见式（4-3），即

$$\mu_s = \mu(1 + 2.5\phi_B) \quad (4-3)$$

爱因斯坦公式只适用于体积分数  $\phi_B < 0.02$  的低浓度悬浮体。

对于流体悬浮体的黏度，当两种互不相溶的流体混合时，构成液-气和液-液等两相体系，若连续相的黏度仍用  $\mu$  表示，非连续相的粘度用  $\mu_p$  表示，非连续相的体积分数用  $\phi_B$  表示，则混合物的黏度可以表示为

$$\mu_s = \mu \left( 1 + \frac{5\mu_p + \frac{2}{5}\mu}{2\mu_p + \mu} \phi_B \right) \quad (4-4)$$

当非连续相为气泡时，通常气体的黏度远远小于液体的黏度，上式可以简化为式（4-5）

$$\mu_s = \mu(1 + \phi_B) \quad (4-5)$$

温度升高，黏度下降，柴油的黏温关系符合 Arrhenius 方程<sup>[26]</sup>，用（4-6）表示：

$$\eta = Ae^{E_a/RT} \quad (4-6)$$

式中： $\eta$ ：表示表观黏度，Pa·s；

$E_a$ ：表征粘流活化能，J/mol；

$A$ ：指前因子或称频率因子，Pa·s；

$R$ ：指气体常数， $R=8.314\text{J}/(\text{mol}\cdot\text{K})$ ；

$T$ ：指绝对温度，K。

对于纯液体的非高温黏度估算<sup>[27]</sup>，有 Orrick-Erbar 法：

$$\ln \frac{\eta_L}{d^{20}M} = A + \frac{B}{T}$$

式中： $d^{20}$  为 20℃时的液体密度 ( $\text{g}\cdot\text{cm}^{-3}$ )，若在常压和 20℃下该物质处于气态，

是用加压下的 20℃ 时的液体密度，若熔点高于 20℃，使用正常熔点下的液体密度；A 和 B 为基团贡献值，需要查表得知。但本基团划分很粗，不能计算炔类或者含氮、硫的化合物。

还有 Saatri-Rao 方法<sup>[28]</sup>：

$$\eta = \eta_B (p_p^s)^{-N}$$

式中： $p_p^s$  是以大气压为单位的蒸汽压； $\eta_B$  是正常沸点  $T_b$  下的液体黏度，单位是 mPa·s。基团对  $\eta_B$  及 N 的贡献通常是累加的。通过查表获得基团对其贡献大小值。

此法需要对化合物的结构进行拆分，如甲苯基团可以拆分为：5 个=C—，一个=C<，一个—CH<sub>3</sub>。此法适合用于估算常压沸点前的黏度值。

对应状态法估计可用 Przewdziecki-Sridhar 法：

$$\eta_L = \frac{V_0}{E(V - V_0)}$$

$$E = -1.12 + \frac{V_c}{12.94 + 0.10M + 2.3p_c + 0.0424T_m - 1158(T_m/T_c)}$$

$$V_0 = 0.0085\omega T_c - 2.02 + \frac{V_m}{0.342(T_m/T_c) + 0.894}$$

式中： $T_m$  为熔点； $V_m$  是  $T_m$  下的液体摩尔体积 ( $\text{cm}^3 \cdot \text{mol}^{-1}$ )；

V 是所求温度下的摩尔体积；

$T_c$ 、 $p_c$ 、 $V_c$  的单位分别是 K、MPa、 $\text{cm}^3 \cdot \text{mol}^{-1}$ 。

高温带压下  $\eta_L$  的估算有 Saatri 法：

$$\ln \eta = \left[ \frac{\ln \eta_B}{\ln(\alpha \eta_B)} \right]^\Phi \ln(\alpha \eta_B)$$

式中： $\eta$  的单位是  $\text{mPa} \cdot \text{s}^{-1}$ ；

$\eta_B$  为温度为  $T_b$  时的黏度，单位是  $\text{mPa} \cdot \text{s}^{-1}$ ；

醇类对应  $\alpha=0.1175$ ，其他化合物  $\alpha=0.248$ 。

对于混合液黏度的计算，引入交互作用系数，计算公式为：

$$\ln \eta_m = \sum_i x_i \ln \eta_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n x_i x_j G_{ij}$$

## 4.2 近红外法测量油品黏度

除传统使用黏度计对黏度测量方法外,化学计量学也在油品黏度值的测量方面有很多应用。韦锐等双光路在近红外短波范围内采用 CCD 型多通道检测器采集光谱信号,利用专用化学计量学拟合软件,采用 PLS 方法回归确定喷气燃料的黏度,测定结果的 R2 为 0.8209,SEP 为 0.04<sup>[29]</sup>。王艳斌等在近红外短波范围内采用 CCD 近红外光谱仪采用 PLS 建模测定润滑油粘度指数,测定结果的 R2 为 0.9956 并且具有较好的重复性<sup>[30]</sup>。由于所测样品是比燃料油复杂的润滑油,不适应线性数据处理方法,冯新沪等在 1100~1500nm 波长范围内采用人工神经网络建模,比与 PLS 的模型进行比较,在实验的基础上证明了人工神经网络建模较偏最小二乘在精度方面的优越性<sup>[31-34]</sup>。

基于黏度作为柴油的一项重要质量指标。建立柴油黏度的快速分析方法对柴油机的研制、调配和使用等都有着重要的意义。

## 4.3 柴油黏度性质预测模型建立

实验所用的 NIR 光谱分析仪扫描范围为 900-1700 nm,采样间隔为 2 nm,样品池采样光程为 10cm,实验数据为随机选取的 80 个柴油样本的近红外光谱数据及其对应的黏度性质,取自 Eigenvector Research 公司的网页<sup>[21]</sup>,其原始光谱数据如图 3-1 所示。根据 3.2.1 章节提出的方法对样本原始光谱数据进行平滑求导处理,预处理后的光谱数据如图 3-2。

对于柴油黏度性质的样品,从预处理后的柴油近红外光谱数据集(共含 395 组样本)中随机选择 60 组作为校正集用来建立预测模型,从剩余样本中随机选择 20 组作为测试集采用预测集残差(e)、预测集相对误差(re)、预测残差平方和(PRESS)及预测样本的标准偏差(SEP)作为模型评价标准。

### 4.3.1 多元线性回归模型

将上述选定的 60 组柴油样本的近红外光谱数据值及黏度性质测定值进行多元线性回归(MLR)。MLR 建模的难度在于要求选择对应于被测性质的特征光谱吸收,对于吸光度与性质之间是线性关系的性质拟合的效果比较好一些。但是这也涉及到需要知道与性质对应的特征光谱区域,并且在选择拟合的方程维数方面,参与回归的变量数(波

长点数)不能超过校正集的样本数目,也就是说波长点数选取有限制,这样难免丢失部分有用的光谱信息,也会造成选取信息的不全面,存在无用信息的选取有用信息丢失的问题。还有就是光谱矩阵可能存在共线的问题,导致算法程序无法求解逆矩阵。如果再加上光谱信号中噪音和背景光的干扰没有处理好的话,往往会导致过度拟合的情况发生,从而在一定程度上减弱了模型的预测和适应能力。

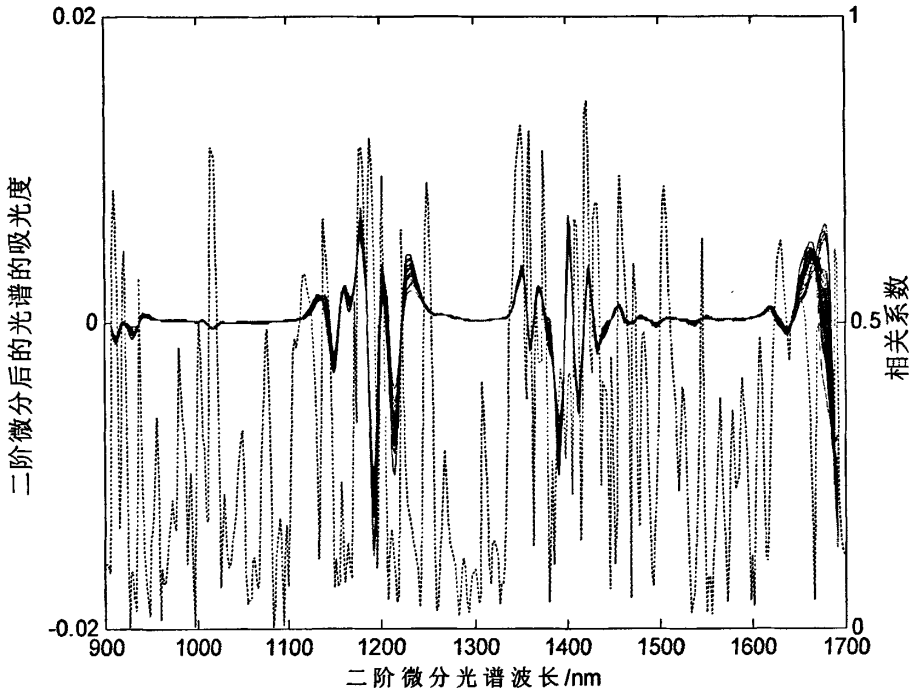


图 4-1 吸光度与黏度性质相关系数

Fig 4-1 Correlation coefficient(R) between viscosity and absorbency

为了解决这个参与回归变量数量以及共线性问题,首先要确定如何进行参与回归模型的波长变量的筛选,如果在体系组成成分不明确了解的前提下,MLR法只适用与线性关系特别好的简单体系,因为这样不需要考虑组分间的相互干扰的影响,筛选变量的工作量很难度也不会那么大。选择合适的相关系数也作为在对样本了解信息不全面的情况下选择光谱维数的不二选择。

可以通过统计学方法选择不同的相关系数拟合的模型的预测残差平方、校验集中最大误差和校验集最大的相对误差来评价所建模型的好坏。这些评价指标都是越小对应的模型预测能力和精度越好。公式如 4.3 中描述。

实验结果如表 4-1 所示:

表 4-1 选取不用波长点数的 MLR 模型的评价指标质量

Table 4-1 Table 4-2 The evaluating indicator at different components of MLR model

相关系数	波长点数	PRESS	SD	最大e	最大re
0.82	3	1.911	0.317	0.849	0.254
0.79	6	0.997	0.229	0.619	0.185
0.78	12	1.38	0.270	0.671	0.201
0.75	15	1.136	0.245	0.564	0.195
0.72	20	0.778	0.220	0.450	0.179
0.68	28	0.356	0.137	0.364	0.109
0.66	34	0.565	0.172	0.361	0.118
0.65	40	1.036	0.233	0.398	0.158

预测残差平方和 (PRESS) 和标准偏差 (SD) 是对模型预测能力的一个总体评价, 而预测集残差 (e)、预测集相对误差 (re) 的最大值则是对模型预测样本的一个可信度评价, 由它们可以确定模型的置信范围, 是对某一个样本的预测能力的评价。这三种评价标准各有其应用范围, 选择建模适合的评价标准进行要合理的模型评价。

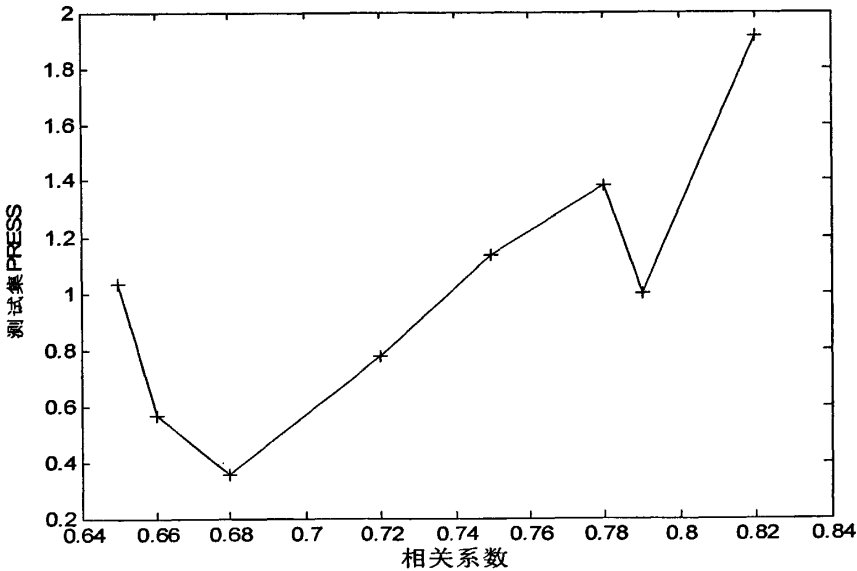


图 4-2 PRESS 值随 R 的变化

Fig 4-2 PRESS value at different R

分析上表和上图所得结论: 算法精度与所选的波长点数有很大的关系, 自变量应变量相关系数增大, 符合条件的波长点数就会减少, 反之亦然。如何选择合适的波长点数需要进行大量的实验, 记录结果对比, 比较费时费力。在本次实验中, 选择自变量应变量相关系数为 0.68, 也就是选择 28 个波长点数时的预测精度最好。

选择 28 个波长点数进行多元线性回归建模，预测集结果如下：

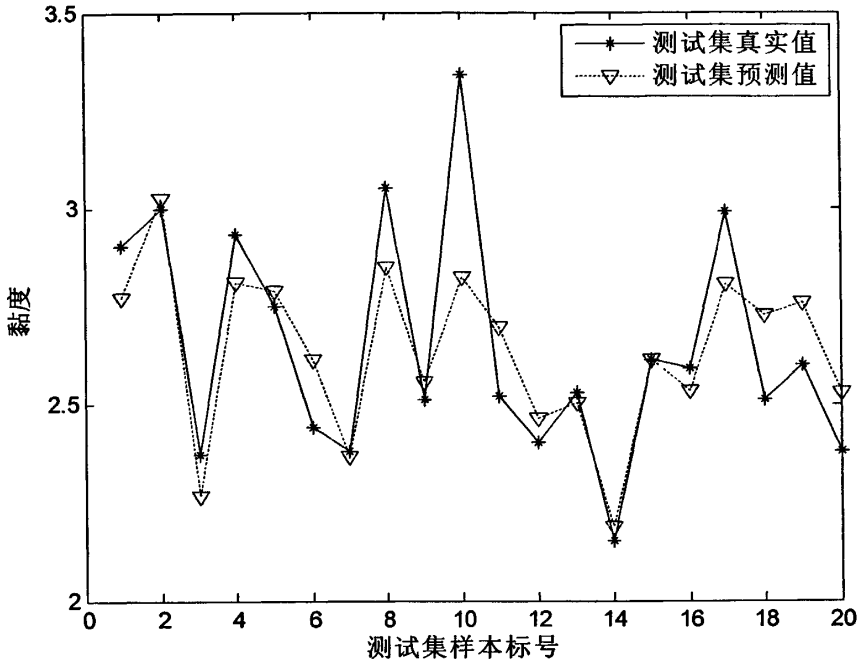


图 4-3 MLR 模型测试集真实值和预测值比较

Fig 4-3 The comparison between true and predictive of MLR method

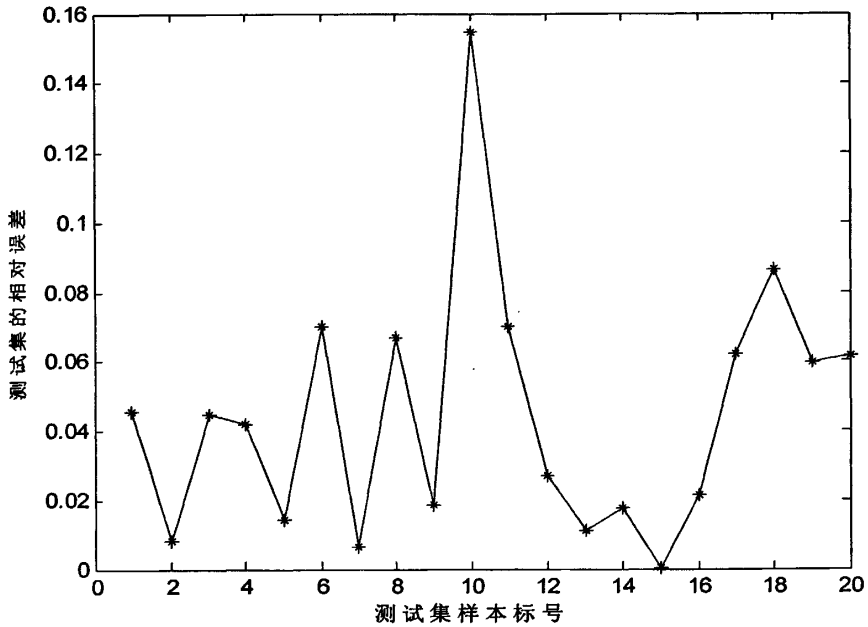


图 4-4 MLR 模型测试集的相对误差

Fig 4-4 The relative predictive error of MLR method

### 4.3.2 偏最小二乘模型

将上述选定的 60 组柴油样本的近红外光谱数据值及黏度测定值进行偏最小二乘法 (PLS) 回归。PLS 建模的最大难点在于建模所需主成份数目的确定。在参与计算的多个主成份中，第一主成份最为重要，对所拟合的性质最为重要，随着主成份数的增加，重要程度依次降低，一定数目之后的主成份则反映的是噪音信息，会对所拟合性质的预测造成不良影响。因此，前面的主成份在建立模型时比后面的主成份更有用。建模所用主成份数要进行合理选择，参与建模的主成分数目过多或过少，都会导致过拟合 (Overfit) 或不充分拟合 (Underfit) 情况发生，使所拟合的模型预测能力和精度受到影响。所以，合理确定参加建模的主成份数是充分利用光谱信息和滤除噪音的有效方法。

对于所建模型好坏的评价，也是用到前面提到的预测残差平方和、标准偏差、预测集残差、预测集相对误差来评价。

表 4-2 选取不同主成分的 PLS 模型的评价指标

Table 4-2 The evaluating indicator at different principal components of PLS model

主成分数	PRESS	SD	最大 e	最大 re
3	4.931	0.510	1.223	0.366
4	4.180	0.469	1.080	0.348
5	3.800	0.447	0.977	0.364
6	2.198	0.30	0.884	0.265
7	1.260	0.258	0.695	0.208
8	0.637	0.183	0.510	0.153
9	0.570	0.173	0.432	0.147
10	0.601	0.178	0.490	0.146
11	0.553	0.171	0.517	0.155
12	0.583	0.175	0.488	0.146
13	0.639	0.183	0.519	0.156
14	0.660	0.186	0.473	0.154
15	0.698	0.192	0.508	0.169

不同的主成分与 PRESS 的关系见图 4-5:

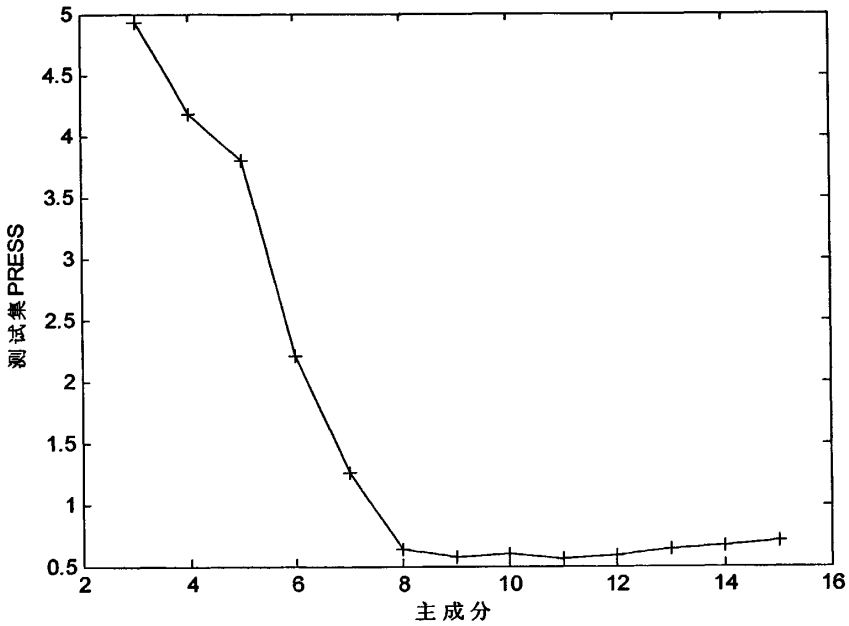


图 4-5 不同的主成分与 PRESS 的关系

Fig 4-5 PRESS value at different principal components

主成分为 11 时效果最好，预测结果如下：

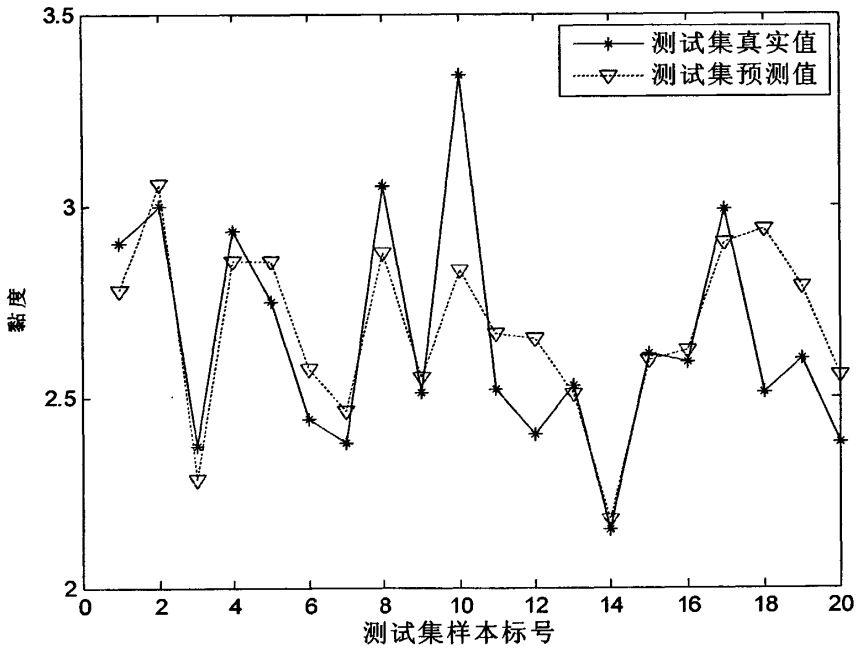


图 4-6 PLS 模型测试集真实值和预测值比较

Fig 4-6 The comparison between true and predictive of PLS method



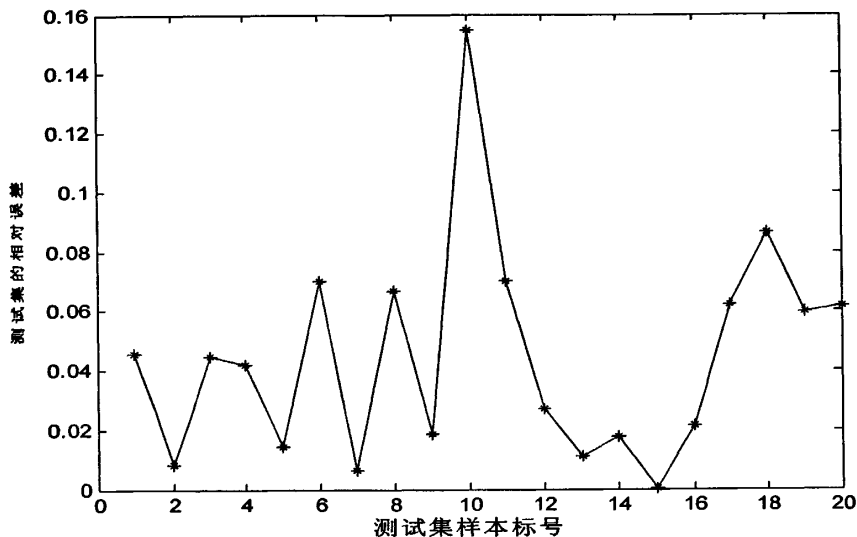


图 4-7 PLS 模型测试集的相对误差

Fig 4-7 The relative predictive error of PLS method

### 4.3.3 柴油黏度性质基团法预测模型

经过小波去噪和二阶微分处理后选择黏度最大的和黏度最小的四组样本，绘制他们的光谱处理图，标示有对应基团的吸光度位置，见图 4-8：

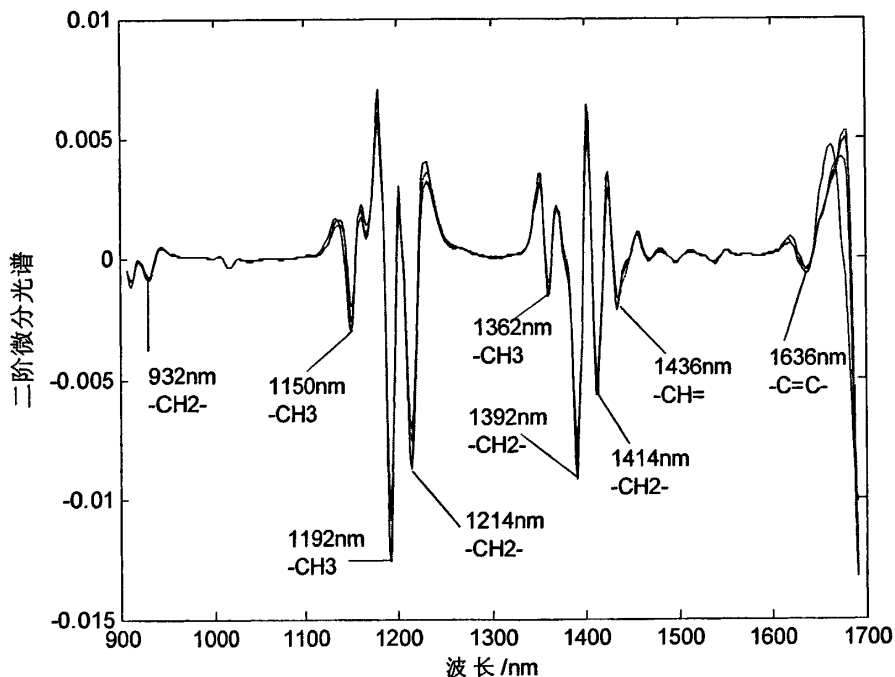


图 4-8 基团特征峰示意图

Fig 4-8 The general view of characteristic group

下面为了基团对应的波长表示得清晰,绘制各波长段的细节图,见图 4-9 至图 4-12,每幅表示 1000nm 的波长长度。

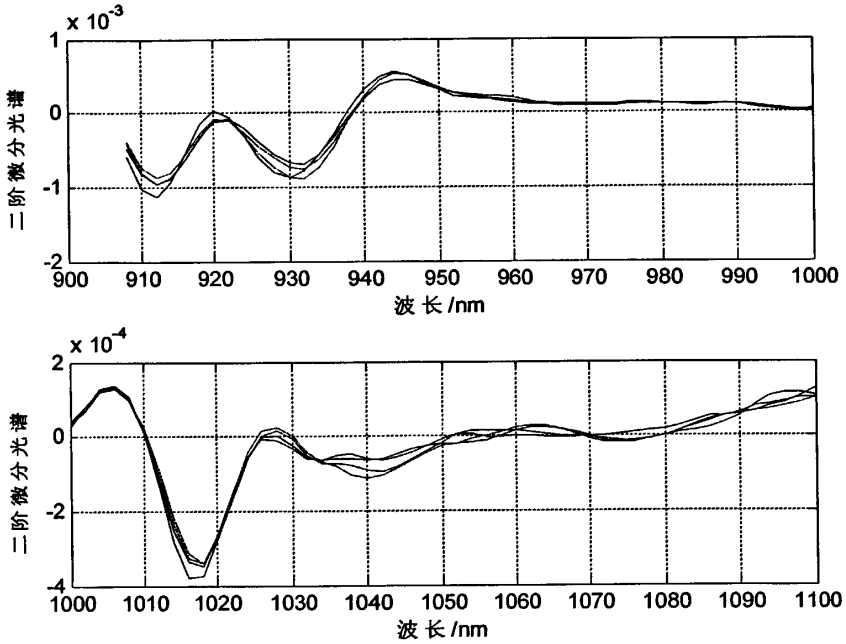


图 4-9 900-1100nm 基团特征峰细节图

Fig 4-9 The detail view of characteristic group in 900-1100nm

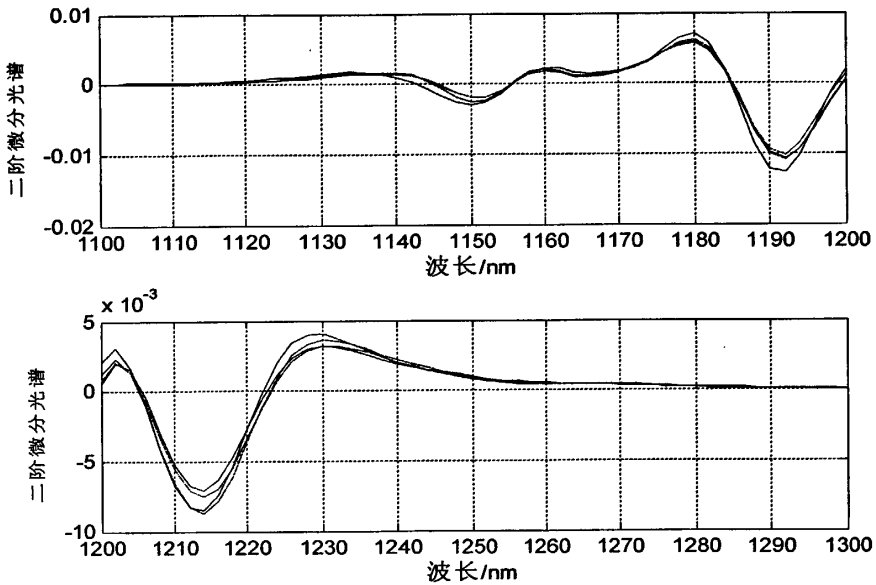


图 4-10 1100-1300nm 基团特征峰细节图

Fig 4-10 The detail view of characteristic group in 1100-1300nm

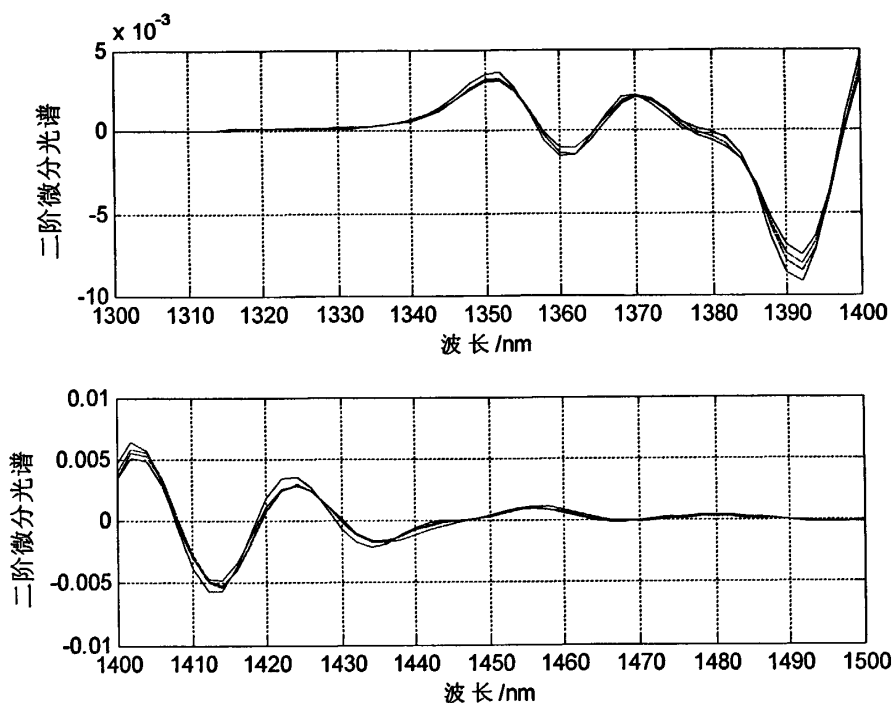


图 4-11 1300-1500nm 基团特征峰细节图

Fig 4-11 The detail view of characteristic group in 1300-1500nm

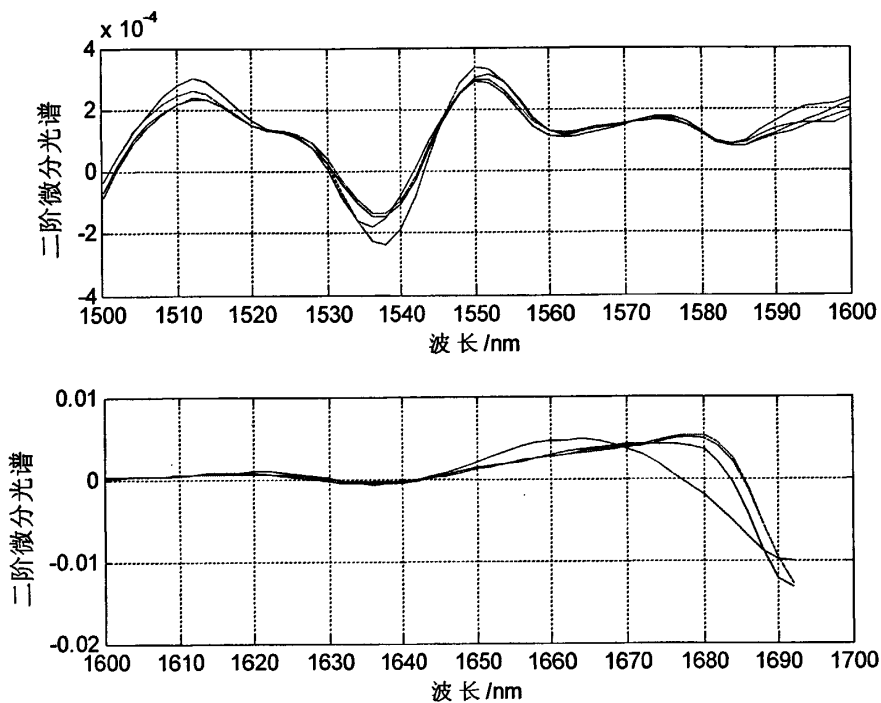


图 4-12 1500-1700nm 基团特征峰细节图

Fig 4-12 The detail view of characteristic group in 1500-1700nm

由图中能明显的表征特征基团的有 1150nm 和 1192nm 处甲基的二倍频、1362nm 处甲基的组合频；932nm 处亚甲基的三级倍频、1214nm 处亚甲基的二级倍频、1392nm 处亚甲基的组合频、1414nm 处亚甲基的组合频；1436nm 处芳香基的组合频；1638nm 处烯基的一级倍频。

(1) 基团法简单线性模型

选择 1150nm 和 1192nm 处的甲基的二级倍频作为  $X_{11}$  和  $X_{12}$ ，1362nm 处的甲基组合频吸收作为  $X_{13}$ ；932nm 的亚甲基三级倍频作为  $X_{21}$ ，1214nm 的亚甲基二级倍频作为  $X_{22}$ ，1392nm 和 1414nm 的亚甲基组合频作为  $X_{23}$ ；1436nm 处的芳烃组合频作为  $X_3$ ；1638nm 处的烯基一级倍频作为  $X_4$ 。这些特征基团吸收作为模型的因变量建立简单线性模型，模型如式 (4-7)：

$$y = -1537.05X_{11} - 904.54X_{12} + 7405.80X_{13} - 605.63X_{21} + 126.92X_{22} - 599.52X_{23} + 1035.08X_{24} - 1059.17X_3 + 2746.35X_4 \quad (4-7)$$

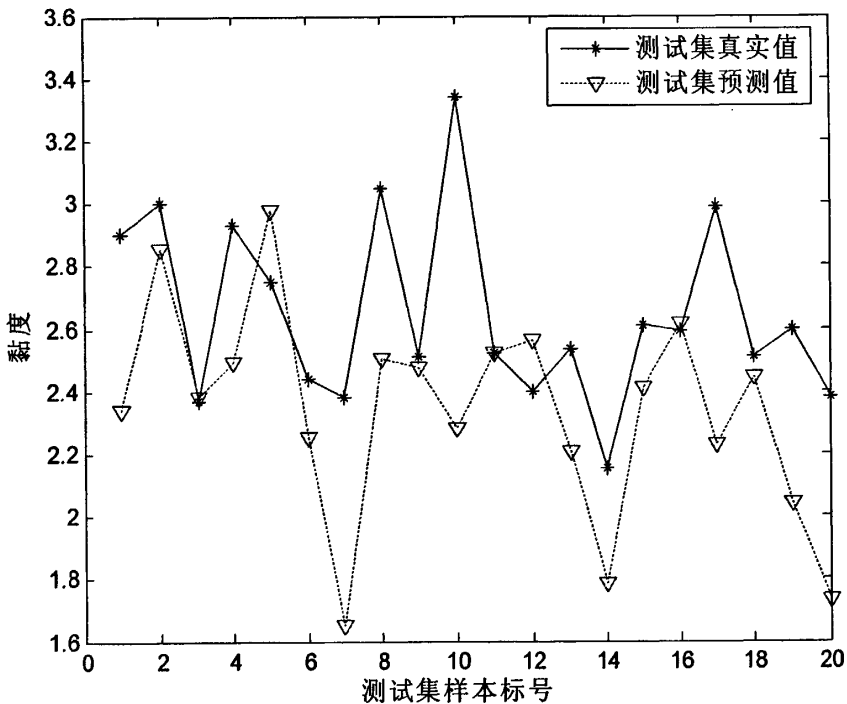


图 4-13 简单线性模型测试集真实值和预测值比较

Fig 4-13 The comparison between true and predictive of simple group method

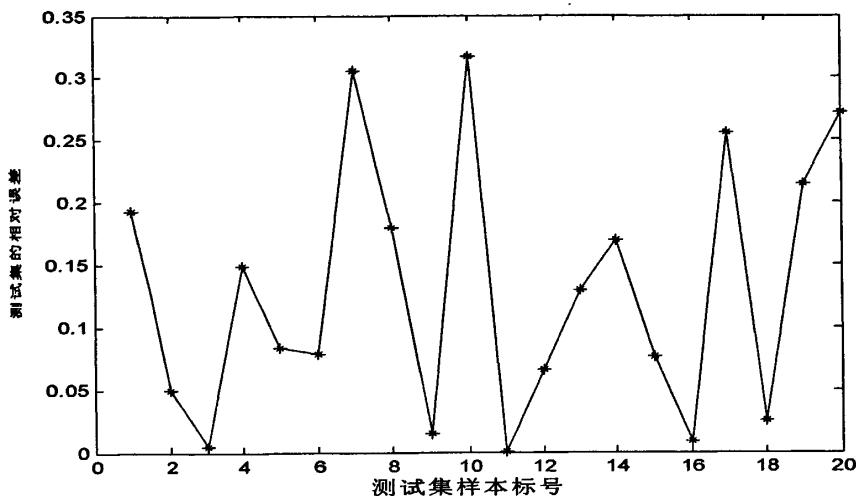


图 4-14 简单线性模型测试集的相对误差

Fig 4-14 The relative predictive error of simple group method

(2) 基团法加入支链化程度和烯烃度的线性模型

下面是对它进行简单改进的算法，添加了  $X_5$  和  $X_6$  表征支链化和烯烃度的项，具体拟合公式见式 (4-8)：

$$\begin{aligned}
 y = & 658.77X_{11} - 753.56X_{12} + 2591.68X_{13} \\
 & -1045.06X_{21} - 466.13X_{22} + 828.71X_{23} + 1076.11X_{24} \\
 & + 193.69X_3 + 10217.17X_4 + 3.14X_5 + 292.75X_6
 \end{aligned} \tag{4-8}$$

这里涉及到的九个波长点数都是按照柴油的特征基团按照图 4-8 对应选取的。

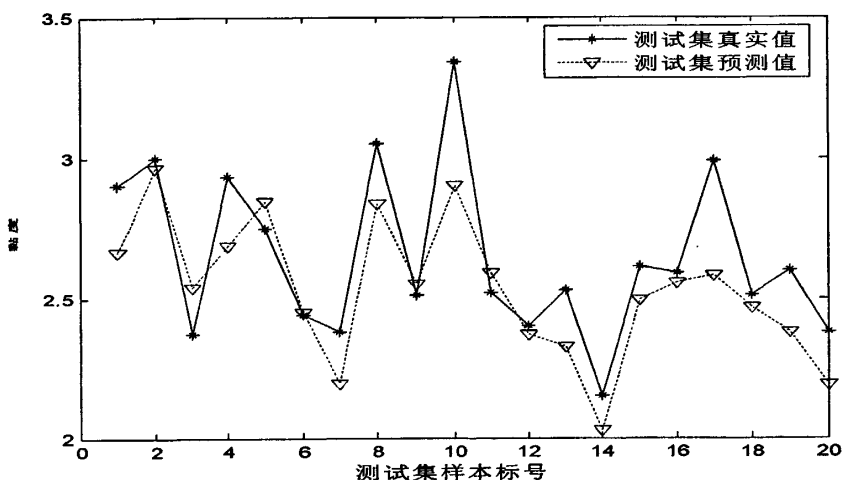


图 4-15 基团组合模型测试集真实值和预测值比较

Fig 4-15 The comparison between true and predictive of combined group method

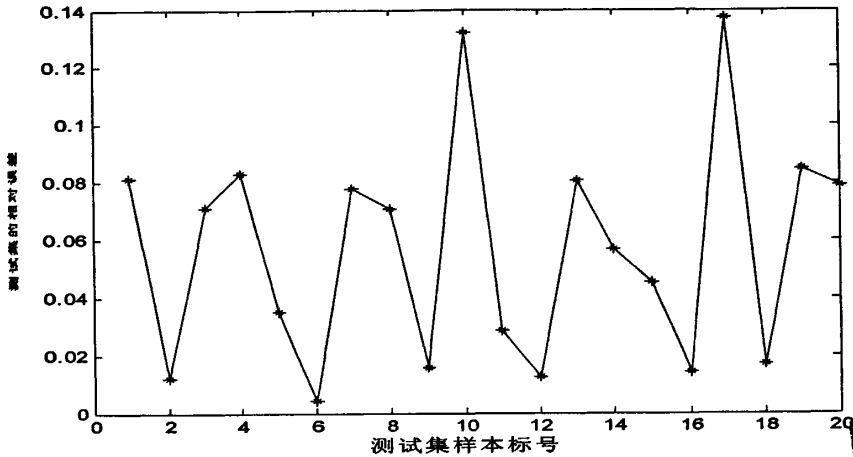


图 4-16 基团组合模型测试集的相对误差

Fig 4-16 The relative predictive error of combined group method

式 (4-7) 预测结果 PRESS 为 4.192, 最大的预测误差为 1.0596, 最大预测的相对误差为 0.317。式 (4-8) 预测结果 PRESS 为 0.761, 最大的预测误差为 0.440, 最大预测的相对误差为 0.137。可以看出优于上面的系统自选数据相关度模型。

其中  $X_{11}$ 、 $X_{12}$ 、 $X_{13}$  表示的是甲基在 1150nm、1192nm、1362nm 处的吸光度的值； $X_{21}$ 、 $X_{22}$ 、 $X_{23}$ 、 $X_{24}$  表示的是亚甲基在 932nm、1214nm、1392nm 和 1414nm 处的吸光度的值； $X_3$ 、 $X_4$  表示的是芳基和烯基分别在 1436nm 和 1636nm 处的吸光度的值。在式 (4-8) 中引入的  $X_5$  和  $X_6$  则分别表示了柴油成分的烷烃支链化和烯烃度。定义如下：

$$X_5 = \frac{X_{21} + X_{22} + X_{23} + X_{24}}{X_{11} + X_{12} + X_{13}} \quad (4-9)$$

$$X_6 = \frac{X_4}{X_{11} + X_{12} + X_{13} + X_{21} + X_{22} + X_{23} + X_{24} - X_3 - X_4} \quad (4-10)$$

$\text{CH}_2$  官能团表征黏度指数增大, 主要表征正构烷烃的组成。 $\text{CH}_3$  官能团表征黏度指数降低, 主要表征直链烷烃的组成<sup>[34]</sup>。碳氢化合物分子量增大会使得不同基团在近红外光谱中的响应差别变小; 柴油成分比较复杂, 还有烯烃和多环芳烃等成分; 柴油各组分间的相互作用会对所预测性质产生或正、或负或协同效应的贡献。上述三种原因都会使得近红外光谱的吸光度信息和很多性质之间呈现非线性关系。

总体而言, 分析式 (4-9) 和式 (4-10) 可知, 两式的甲基吸光度前面的系数为负, 表示了甲基对柴油黏度性质是负贡献的; 亚甲基前面的吸光度系数为正, 表征了亚甲基

对柴油黏度的性质是正贡献的；式（4-9）表征支链化程度，它前面的系数为正，符合理论分析的相对分子量越大，支链化程度越大，黏度性质越大的规律；式（4-10）表示了烯炔度，烯炔度越大，柴油黏度性质越大，也就是烯炔的含量增多也会使得油品黏度性质上升。

### （3）L-M 算法建模

考虑到  $X_{11}$ 、 $X_{12}$ 、 $X_{13}$ 、 $X_{21}$ 、 $X_{22}$ 、 $X_{23}$ 、 $X_{24}$ 、 $X_3$ 、 $X_4$  是光谱吸光度数据，而  $X_5$ 、 $X_6$  是表征组分结构的数据，对吸光度数据进行 e 的指数运算，采用 L-M 算法，建立形如式（4-11）的黏度预测模型。

$$f(X) = (153.696e^{X_{11}} + 65.5945e^{X_{12}} + 165.7182e^{X_{13}} - 258.74e^{X_{21}} - 127.8544e^{X_{22}} + 126.8544e^{X_{23}} + 164.3754e^{X_{24}} + 42.0153e^{X_3} - 328.7399e^{X_4} - 0.2997X_5 - 6.3328X_6)^{2.6124} \quad (4-11)$$

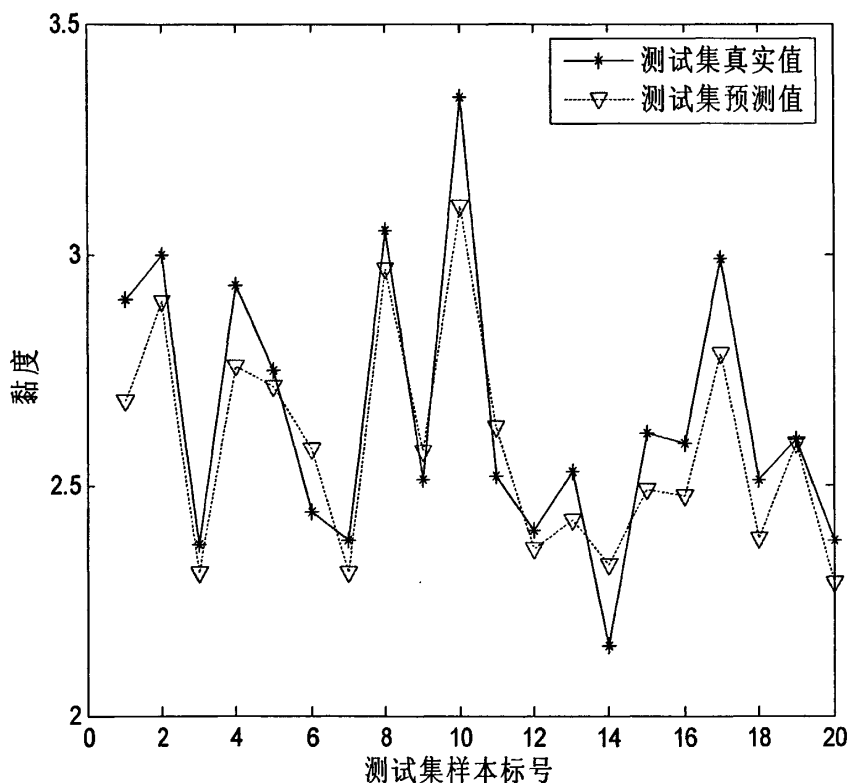


图 4-17 L-M 模型测试集的真实值和预测值比较

Fig 4-17 The comparison between true and predictive of L-M group method

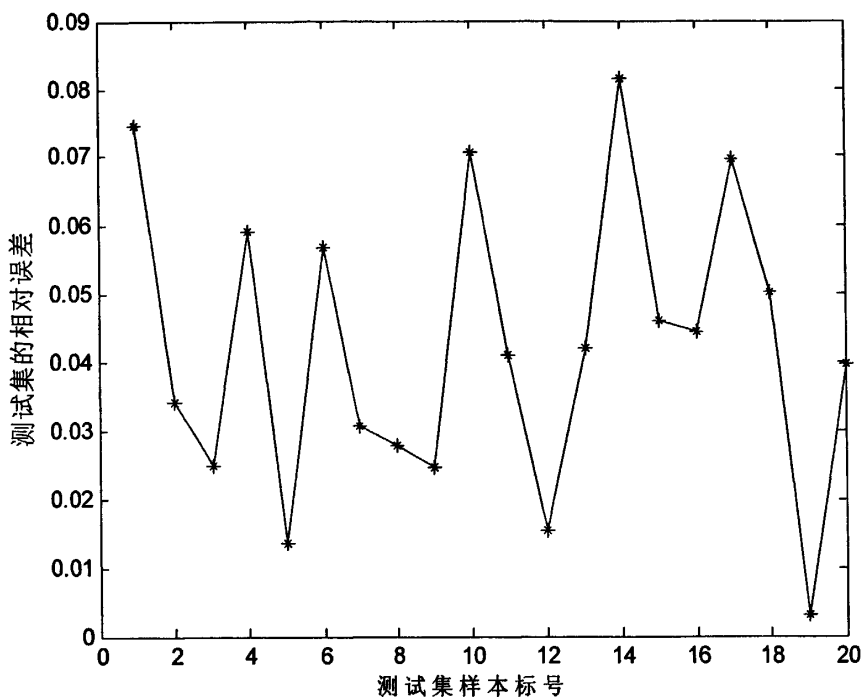


图 4-18 L-M 模型测试集的相对误差

Fig 4-18 The relative predictive error of L-M group method

#### 4.4 结论比较

通过进行多元线性回归、偏最小二乘的拟合实验和利用基团选择的线性非线性实验，结果比较见表4-3：

表 4-3 各模型测试集预测结果比较

Table 4-3 The comparison among evaluating indicators at different model

评价指标	MLR	PLS	线性基团法	组合基团法	L-M基团法
PRESS	0.356	0.553	4.192	0.761	0.332
SD	0.137	0.171	0.470	0.200	0.132
最大e	0.364	0.517	1.059	0.440	0.236
最大re	0.109	0.155	0.317	0.137	0.0815

通过对比表中数据，可知非线性的L-M基团法不论从对单个样本还是群体样本的预测性方面都优于化学计量学的模型的预测效果。



## 第 5 章 柴油密度性质的预测模型建立

### 5.1 柴油密度测试方法及现状

油品密度是油品检验过程中一项基本而且重要的指标，在测量中常受多种条件的影响。对油品密度由影响的主要有油品温度、油品的挥发性和不均匀性三个方面因素。温度对密度测定的影响比较显著，在测量密度时，温度选择方面应遵守的原则有二：保证油品具有足够的流动性；尽可能接近运输条件。这样测出的油品密度更具有实际意义。

传统的柴油密度都是通过实验室测定的。依据 GB/T 4756 取样，按照 GB/T 1884-2000 《原油和液体石油产品密度实验室测定法（密度计法）》进行测定。

测定原理是：使试样处于规定温度，将试样倒入温度大致相同的密度计量筒中，为了避免测定期间温度变化太大，需将密度计量筒及内装的试样一起放在恒温浴中，将合适的密度计放入已调好温度的试样中，让密度计静止。当温度达到平衡后，读取密度计刻度读数和试样温度。然后按照不同的实验油品，用 GB/T 1885-1998 石油计量表的表数据把观察到的密度计读数换算成 20℃ 下标准密度。

在实验温度下，玻璃密度计在液体试样中的读数为视在密度，单位为  $\text{kg/m}^3$  或  $\text{g/cm}^3$ 。标准密度则是在标准温度 20℃ 下的密度，单位为  $\text{kg/m}^3$ 。

由于密度与分子结构密切相关，王克强<sup>[35]</sup>将烷烃分子的基团键划分为十个种类分别用  $GB_i$  表示，采用密度函数，见式（5-1）计算烷烃的密度。

$$\rho = \frac{\sum_i c_i n(GB_i)}{\sum_i w_i n(GB_i)} \quad (5-1)$$

式中： $n(GB_i)$  为基团键  $GB_i$  的数目；

$c_i$  和  $w_i$  是与基团键特性相关的常数，利用拟合计算机拟合得到。

对 658 种烷烃的密度预测计算结果十分接近实验值。

冯瑞英等<sup>[36]</sup>利用分子结构，采用拓扑指数法预测芳烃的密度，采用式（5-2）对 212 种芳烃密度进行了预测，结果表明，大部分芳烃预测值的计算误差在 1% 之内，对芳烃同分异构体也具有良好的适应性。

$$d = B_1 + B_2 e^{-4W_s} + B_3 th(\lambda_2^3 P_s) + B_4 S_1 + B_5 S_2 \quad (5-2)$$

其中： $W_s$ 、 $P_s$ 、 $S_1$ 、 $S_2$  是拓扑指数，表征分子结构；

$B_1$ 、 $B_2$ 、 $B_3$ 、 $B_4$ 、 $B_5$ 、 $\lambda_1$ 、 $\lambda_2$  为常数, 用计算机采用非线性拟合方法得到。

为了生产计量和产品交接的需求, 对油品密度的在线测量方法有: 浮子法、静压法、射线法、超声波法、振动法等。这些方法都或多或少的有一些缺陷, 不利于大范围的推广使用。

## 5.2 近红外法测量柴油密度

近红外光谱法应用于柴油密度的检测已经有十几年的历史了。随着近红外光谱仪的研制和开发, 徐广通等<sup>[37-38]</sup>利用近红外光谱数据在 C-H 对应基团的吸光度与油品性质之间的相关度, 建立了密度等性质的相关模型, 为快速便利的研究油品组成和性质关系的方法上做了尝试。

徐广通等<sup>[39]</sup>利用近红外光谱建立偏最小二乘回归模型, 建立了快速测定柴油密度等性质的方法, 在一定的近红外波长范围内, 该法的测定结果达到了标准方法重复性的测试要求。

方利民等利用新兴的独立变量分析分析柴油的密度等与近红外光谱之间的关系, 提出柴油组分含量测定和基于独立变量分析-神经网络回归 (ICA-NNR) 的近红外光谱分析方法, 建立和柴油组分含量和密度性质的关联模型, 实验证明具有很好的可行性<sup>[40]</sup>。

王志芳等提出用基团法建模, 采用高斯-牛顿法确定模型参数, 对柴油密度进行软测量。在预测精度和在线测量方面都取得了很好的效果<sup>[41]</sup>。

## 5.3 柴油密度性质预测模型

对于柴油密度性质的非空样品, 将对应样本的原始光谱数据进行小波去噪、平滑求导处理后, 从预处理后的柴油近红外光谱数据集(共含 395 组样本)中随机选择 60 组作为校正集用来建立预测模型, 从剩余样本中随机选择 20 组作为测试集采用预测集残差(e)、预测集相对误差(re)、预测残差平方和(PRESS)及预测样本的标准偏差(SEP)作为模型评价标准。

### 5.3.1 多元线性回归模型

将上述选定的 60 组柴油样本的近红外光谱数据值及密度性质测定值进行多元线性回归(MLR)。但是这涉及到需要知道与性质对应的特征光谱区域, 并且在选择拟合的方程维数方面, 考虑到参与回归的变量数(波长点数)不能超过校正集的样本数目, 也

就是说波长点数选取有限制，这样难免丢失部分有用的光谱信息，也会造成选取信息的不全面，存在无用信息的选取有用信息丢失的问题。还应该考虑到的就是光谱矩阵可能存在共线的问题，导致算法程序无法求解逆矩阵。

为了解决这个参与回归变量数量以及共线性问题，利用选择合适的相关系数作为在对样本了解信息不全面的情况下确定参与建模光谱维数的方法。

表 5-1 选取不同波长点数的 MLR 模型的评价指标

Table 5-1 The evaluating indicator at different components of MLR model

相关系数	波长点数	PRESS	SD	最大e	最大re
0.88	3	0.0058	0.0175	0.0358	0.042
0.85	8	0.0056	0.0172	0.0278	0.0331
0.83	14	0.0019	0.0099	0.0245	0.0293
0.80	26	0.0098	0.0072	0.019	0.0228
0.78	33	0.0005	0.0052	0.0162	0.0193
0.77	39	0.0022	0.0107	0.0190	0.0224
0.76	49	0.0039	0.0144	0.0252	0.0295
0.75	59	0.0032	0.0130	0.0248	0.0291

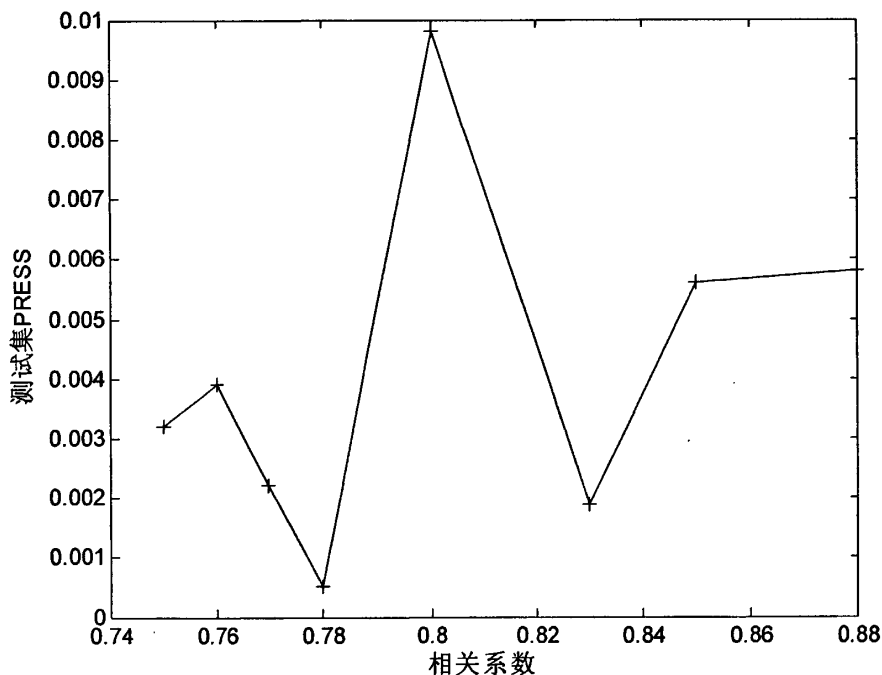


图 5-1 不同的 R 与 PRESS 的关系

Fig 5-1 PRESS value at different R

分析上表和上图所得结论：算法的预测精度与所选的波长点数有很大的关系，自变量应变量相关系数减小，参与建模的波长点数就会增加，反之亦然。在本次实验中，选择自变量应变量相关系数为 0.78，也就是选择 33 个波长点数时的预测精度最好。

选择 33 个波长点数进行多元线性回归建模，预测的真实值与预测值结果及相对误差如图 (5-2) 和 (5-3)：

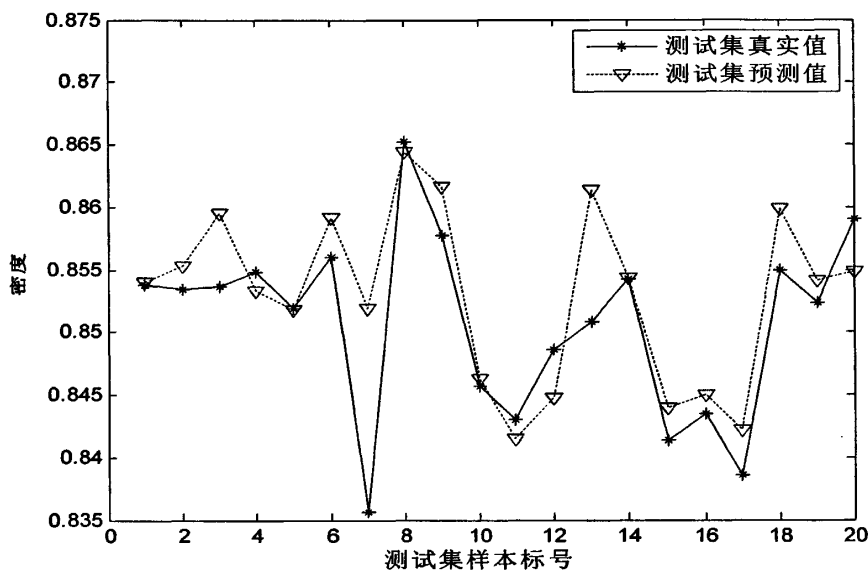


图 5-2 MLR 模型测试集真实值和预测值比较

Fig 5-2 The comparison between true and predictive of MLR method

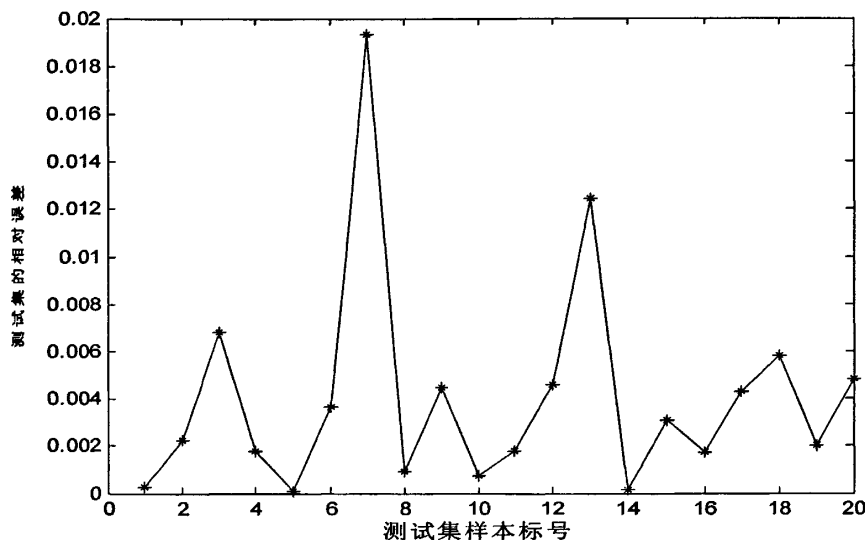


图 5-3 MLR 模型测试集的相对误差

Fig 5-3 The relative predictive error of MLR method

可以看出 MLR 建模需要进行多次的试验，选择合适的建模数据与所预测性质的相关系数，才选取合适数量的波长点数参与模型的建立，这是导致建模精度大小变化的很重要原因。

### 5.3.2 偏最小二乘模型

将上述选定的 60 组柴油样本的近红外光谱数据值及密度测定值进行偏最小二乘法 (PLS) 回归。

选取不同的主成分数进行建模，用预测残差平方和、标准偏差、预测集残差、预测集相对误差来评价所建模型的好坏。

表 5-2 选取不同主成分的 PLS 模型的评价指标

Table 5-2 The evaluating indicator at different principal components of PLS model

主成分数	PRESS	SD	最大 e	最大 re
5	0.003	0.0126	0.0268	0.312
6	0.0028	0.0122	0.0296	0.0344
7	0.0014	0.0086	0.0231	0.0269
8	0.0008	0.0064	0.0131	0.0152
9	0.0004	0.0047	0.0118	0.0138
10	0.00032	0.0041	0.0094	0.0109
11	0.00027	0.0037	0.008	0.0093
12	0.0003	0.004	0.0099	0.0115
13	0.00032	0.0041	0.0083	0.0097

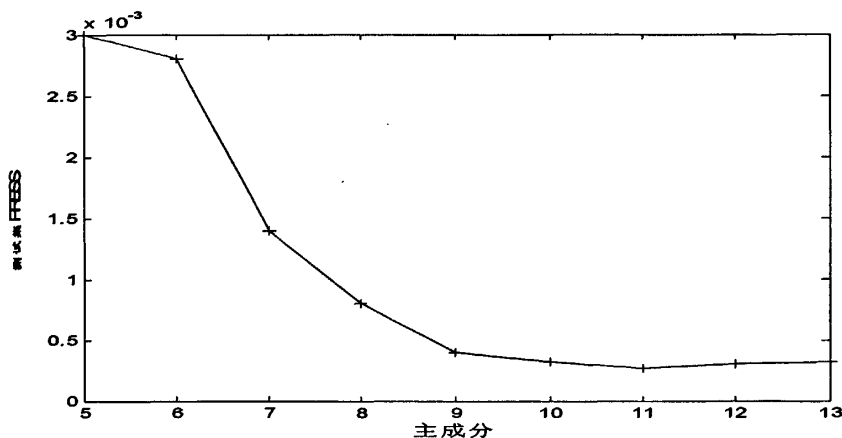


图 5-4 不同的主成分与 PRESS 的关系

Fig 5-4 PRESS value at different principal components

由上表可知选择11个主成分时所见到的密度预测模型精度会比较好，预测集的真实值和预测值及相对误差图见下表：

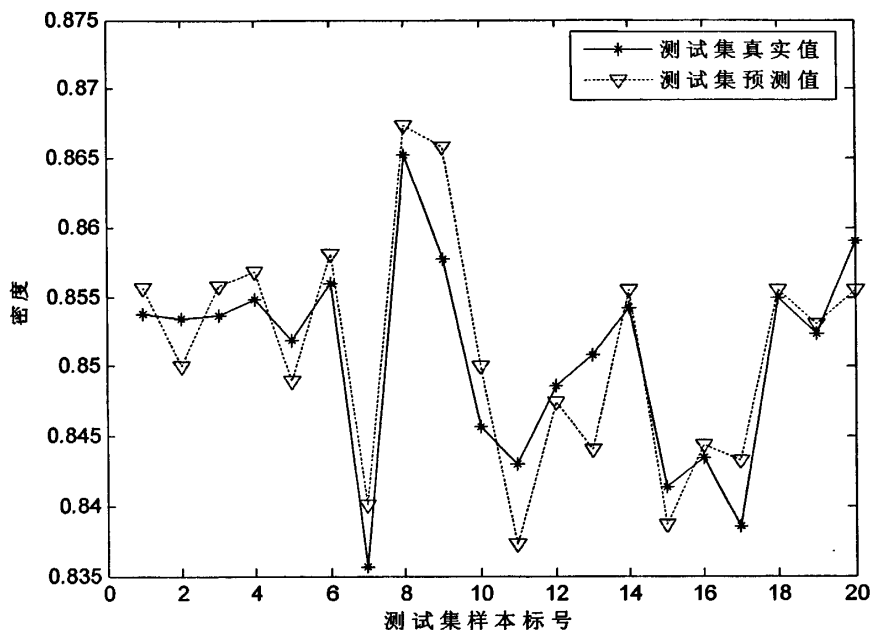


图 5-5 PLS 模型测试集真实值和预测值比较

Fig 5-5 The comparison between true and predictive of PLS method

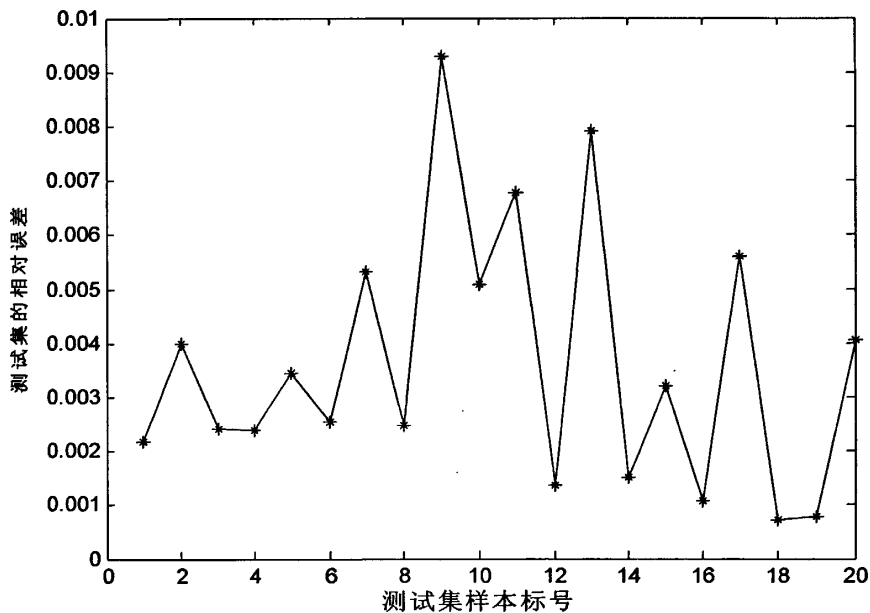


图 5-6 PLS 模型测试集的相对误差

Fig 5-6 The relative predictive error of PLS method

### 5.3.3 柴油密度性质的基团法预测模型

经过小波去噪和二阶微分处理后的光谱数据图与前章类似。

在进行波长自选的前提下建立简单线性模型，模型见式（5-3）：

$$y = -157.6974X_{11} - 61.33X_{12} + 571.2783X_{13} - 140.4409X_{21} + 64.1137X_{22} - 27.1411X_{23} - 149.4999X_{24} - 52.309X_3 + 123.5038X_4 \quad (5-3)$$

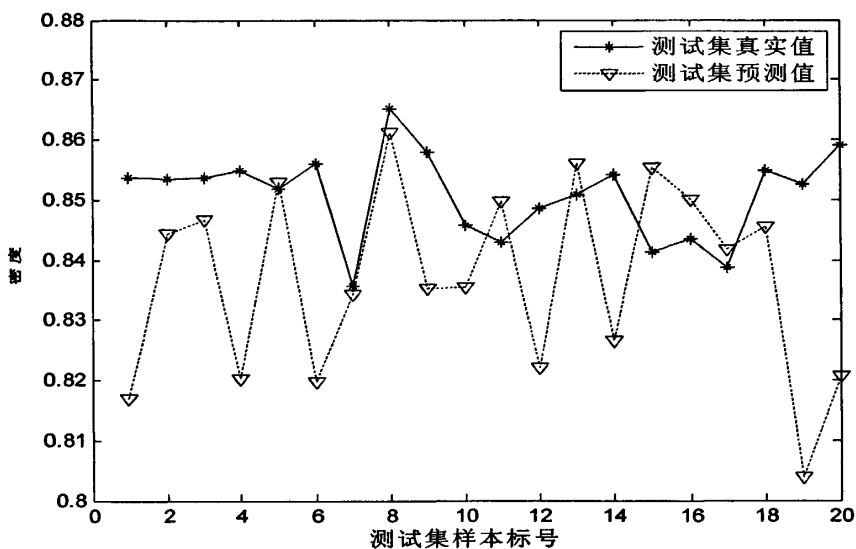


图 5-7 简单线性模型测试集真实值和预测值比较

Fig 5-7 The comparison between true and predictive of simple group method

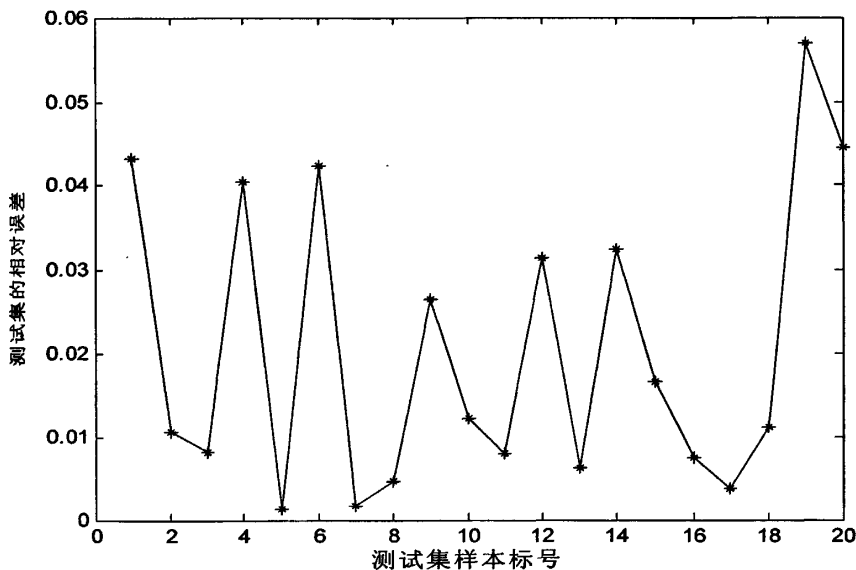


图 5-8 简单线性模型测试集的相对误差

Fig 5-8 The relative predictive error of simple group method

下面是对它进行简单改进的算法，添加了  $X_5$  和  $X_6$  表征支链化和烯烃度的项，具体拟合公式见式 (5-4)

$$\begin{aligned}
 y = & -72.6933X_{11} - 44.9816X_{12} + 60.157X_{13} \\
 & + 43.6884X_{21} - 0.5801X_{22} - 6.7617X_{23} - 14.9777X_{24} \\
 & + 23.0393X_3 + 1207.0207X_4 + 0.1963X_5 + 34.9827X_6
 \end{aligned}
 \tag{5-4}$$

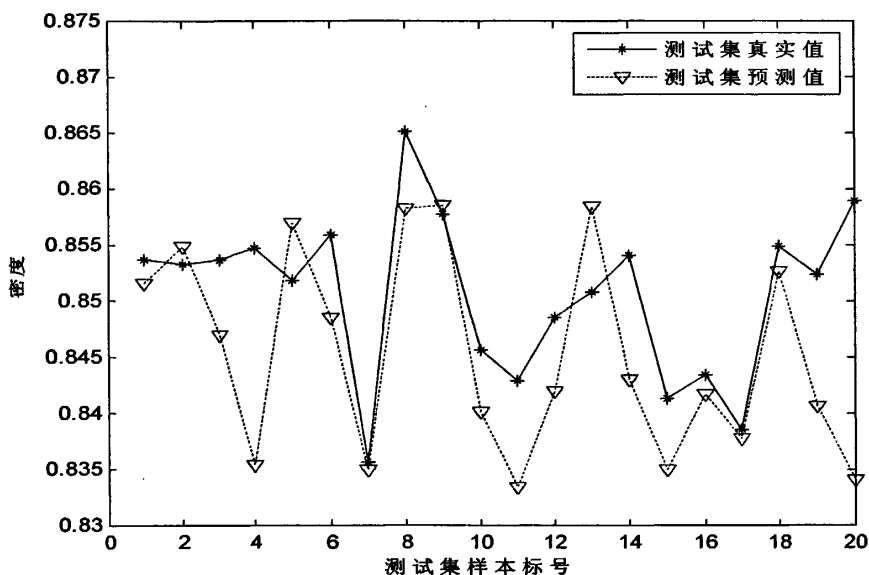


图 5-9 基团组合模型测试集真实值和预测值比较

Fig 5-9 The comparison between true and predictive of combined group method

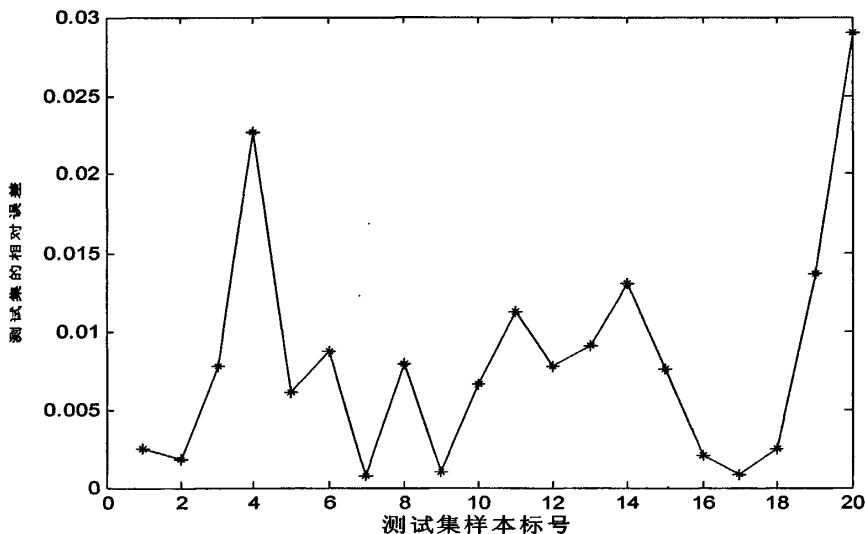


图 5-10 基团组合模型测试集的相对误差

Fig 5-10 The relative predictive error of combined group method



式 (5-3) 预测结果 PRESS 为 0.0103, 最大的预测误差为 0.0486, 最大预测的相对误差为 0.057。式 (5-4) 预测结果 PRESS 为 0.0017, 最大的预测误差为 0.0249, 最大预测的相对误差为 0.029。可以组合的基团法所见模型优于只是选择特征波长的基团法模型。

考虑到  $X_{11}$ 、 $X_{12}$ 、 $X_{13}$ 、 $X_{21}$ 、 $X_{22}$ 、 $X_{23}$ 、 $X_{24}$ 、 $X_3$ 、 $X_4$  是光谱吸光度数据, 而  $X_5$ 、 $X_6$  是表征组分结构的数据, 对吸光度数据进行 e 的指数运算, 采用 L-M 算法, 建立形如式 (5-5) 的密度预测模型。

$$f(X) = (4.925e^{X_{11}} + 5.9867e^{X_{12}} + 17.2237e^{X_{13}} - 12.22581e^{X_{21}} - 5.2651e^{X_{22}} + 6.1201e^{X_{23}} + 6.3112e^{X_{24}} - 0.4755e^{X_3} - 21.474e^{X_4} - 0.0471X_5 - 0.5943X_6)^{2.7026} \quad (5-5)$$

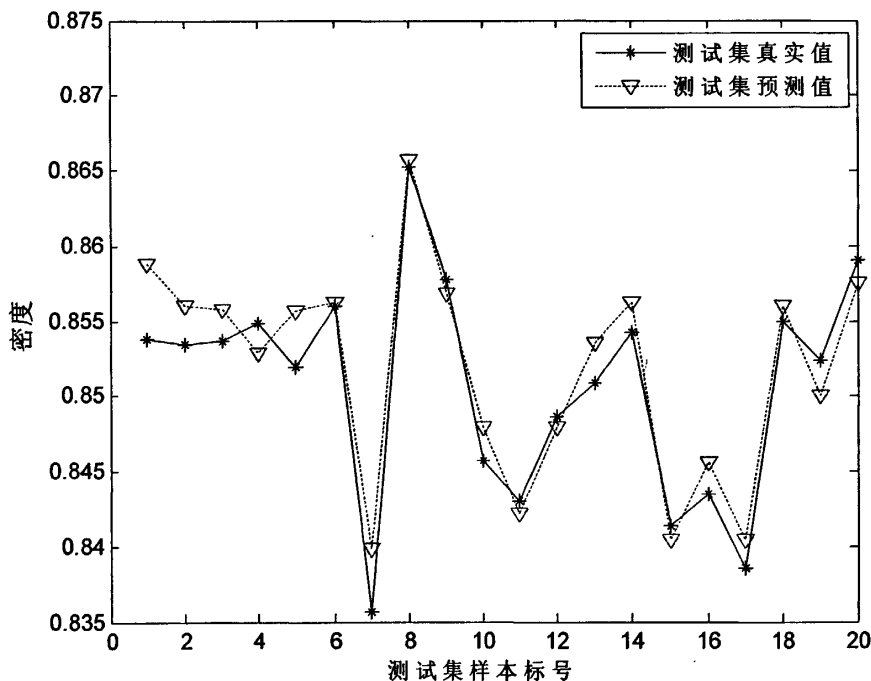


图 5-11 L-M 模型测试集的真实值和预测值比较

Fig 5-11 The comparison between true and predictive of L-M group method

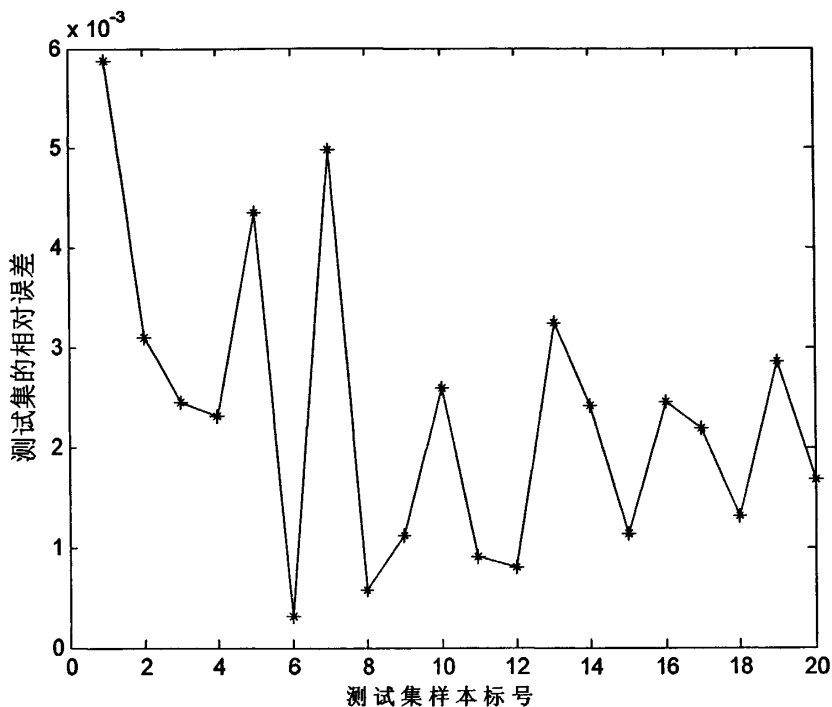


图 5-12 L-M 模型测试集的相对误差

Fig 5-12 The relative predictive error of L-M group method

## 5.4 结论比较

通过进行多元线性回归、偏最小二乘的拟合实验和利用基团选择的线性非线性实验，结果见表5-3中各评价指标：

表 5-3 各模型测试集预测结果比较

Table 5-3 The comparison among evaluating indicators at different model

评价指标	MLR	PLS	线性基团法	组合基团法	L-M基团法
PRESS	0.0005	0.0003	0.0103	0.0017	0.0001
SD	0.0052	0.0037	0.0233	0.0095	0.0024
最大e	0.0162	0.008	0.0486	0.0249	0.005
最大re	0.0193	0.0093	0.057	0.029	0.0059

通过对表中数据进行比较，可以看出对于柴油密度，非线性的L-M基团法不论从对单个样本还是群体样本的预测性方面都优于化学计量学的模型的预测效果。

## 第6章 柴油闪点的预测模型建立

### 6.1 柴油闪点测试方法及现状

闪点（flash point,  $T_{FP}$ ）是可燃性液体表面上的蒸汽和空气的混合物与火接触而初次发生蓝色火焰闪光时的温度，测定时需要缓慢加热，并有专门仪器测定。开杯式用于测定高闪点温度液体，也就是重质油品闪点和着火点的测定；闭杯式用于测定低闪点温度液体，也就是轻质油品闪点的测定。柴油属于轻质油品。

虽然闪点的火焰是一闪即灭，但它表示油品着火燃烧的危险程度，习惯上根据闪点来对危险品进行等级划分，闪点越低就越危险。闪点低于着火点。闪点和着火点都是表示可燃性液体发生着火和爆炸的可能性，是重要的安全数据，对化学品的运输、储存、和使用安全都很重要。

闪点在内燃机油等的使用过程中也有着重要的意义。若闪点显著下降，则说明该内燃机油已经受到燃料的稀释，需要及时更换处理等操作，否则会对内燃机的运行安全造成威胁。

闪点的高低表征了石油产品的蒸发性，是判断油品馏分组成的重要质量指标，馏分组成较重往往具有较高的闪点，据此可以根据闪点调整生产过程中的操作指标，可以方便生产中提出油品的所含的轻质馏分<sup>[42]</sup>。

#### 6.1.1 柴油闪点测定方法

车用柴油的现行测量是按照《GB 261-83-T 石油产品闪点测定法（闭口杯法）》进行测定的。选择不同的实验测试方法会对油品闪点测试结果准确性造成一定的影响，在测试过程中试样的油量多少、试样含水量、加热速度、测试时的点火次数和点火火焰直径大小也都会影响到闪点测试的结果，其中由环境因素决定的大气压力会对闪电的测试结果产生的影响，可以对所测得的闪点利用式（6-1）进行修正：

$$\begin{aligned}\Delta t &= 0.25(101.3 - P) \\ \Delta t &= 0.0345(760 - P)\end{aligned}\quad (6-1)$$

其中： $P$ ——实际大气压力。当  $P$  的单位为千帕斯卡（kPa）时用式（6-1）的上式修正；当  $P$  的单位为毫米汞柱（mmHg 柱）时用式（6-1）的下式修正。

#### 6.1.2 柴油闪点关联计算方法

对于单一体系的可燃液体的闪点是随分子量的增大、沸点的升高、密度的增大而升

高；随饱和蒸汽压的增大而降低；并且饱和烷烃比不饱和烷烃的闪点高；对于同分异构体，直链结构比支链结构的样本闪点高。

孙锐睿等<sup>[43]</sup>基于炼厂实际的操作数据，用 20℃ 时油品的相对密度  $d_4^{20}$  和常压恩式蒸馏下各馏程数据回归了常压塔产品的闪点预测模型。

$$T_{FP} = \alpha_1 + \alpha_2 x_1^{0.6} x_2^{0.5} + \alpha_3 \frac{x_2}{x_1^{0.1}} + \alpha_4 \frac{T_V^{0.34}}{x_1^{0.15}} + \alpha_5 \frac{T_V^{0.6}}{x_1^{1.7}} + \alpha_6 \frac{x_2^{0.67} T_V^{0.34}}{x_1^{0.15}} + \alpha_7 \frac{T_V^3}{x_1^3} \quad (6-2)$$

$$\text{其中: } T_V = \frac{x_2 + x_3 + x_4 + x_5 + x_6}{5}$$

式中:  $T_{FP}$  —— 闪点, °C;

$x_1$  ——  $d_4^{20}$ , 20℃ 时油品的相对密度;

$x_2, x_3, \dots, x_6$  —— 分别表示常压恩式蒸馏 0%、10%、50%、90%、100% 馏程温度, °C;

$\alpha_1, \alpha_2, \dots, \alpha_7$  —— 方程式回归系数, 由回归计算得到。

式 (6-2) 预测模型函数中含有 6 个参数, 8 个系数, 方程形式简单, 还能取得不错的预测精度。

刘斌等<sup>[44]</sup>提出了一个简便计算闪点的估算经验式, 在已知混合各单独组分闪点和混合物中所占体积分数的基础上可以进行混合体系闪点的快速估算。具体公式见:

$$T_{\text{混FP}} = \frac{1}{\frac{V_1}{T_1} + \frac{V_2}{T_2} + \dots + \frac{V_n}{T_n}} \quad (6-3)$$

式中:  $T_{\text{混FP}}$  —— 混合液体的闪点, °C;

$V_1, V_2, \dots, V_n$  —— 各组分的体积分数;

$T_1, T_2, \dots, T_n$  —— 各组分的闪点, °C。

将式 (6-3) 左右两边都求倒数, 可以解释这个经验式的意义为: 混合液闪点的倒数等于各组分闪点的算术平均值, 与并联电路电阻值求法类似。实验证明对于二元、三元、四元体系以及乙醇和水混合溶液这类含有不燃组分的混合体系都具有很好的适应

性。

在定量结构-性质相关性 (QSPR) 运用于有机物闪点等的可以分为拓扑学方法和基团贡献法。

潘勇等<sup>[45-46]</sup>利用烃类化合物原子类型电性拓扑状态指数, 用以表征物质的分子结构特性, 对化合物的闪点及其分子结构之间采用多元线性回归和人工神经网络构建了预测模型。在考虑到分子中基团特性和基团间化学键的连接性基于基团贡献法建立人工神经网络预测烷烃闪点模型, 揭示了有机物分子结构和有机物闪点的定量关系。

王克强等<sup>[47]</sup>按照分子结构特点, 对有机化合物的闪点和分子组成进行模型构建, 提出一种具有普遍适用性的有机化合物闪点估计方法——三基团贡献法。预测闪点关联式为:

$$T_{FP} = 27.0 + 13.5 \left\{ \sum_i [\Delta_i^{(0)} + n_i (\Delta_i^{(1)} + x_i \Delta_i^{(2)})] - 59.7 \right\}^{0.4} \quad (6-4)$$

式中:  $n_i$ ——基团  $i$  的数目;

$\Delta^{(0)}$ 、 $\Delta^{(1)}$ 、 $\Delta^{(2)}$ ——基团参数;

$x_i = n_i / \sum n_j$ ——基团分率。

对应简单基团的集团参数可以查表得到, 只要在预测化合物时确定基团数目即可。此法对于单官能团、多官能团有机物、极性、非极性物质都具有良好的适应性。

高礼久等<sup>[48]</sup>把路径数也就是两原子间相隔的化学键数和沸点引入闪点的预测公式中。实验证明, 该模型原理简单, 结果可靠, 置信度高。化学键数利用分子图的距离矩阵得出,  $P_1$  比分子中碳原子个数少 1,  $P_2$  反应分子的支化度,  $P_3$  称为极化数, 用以表征分子的形状,  $T_b$  对应沸点, 单位是  $^{\circ}\text{C}$ 。回归方程形如:

$$T_{FP} = 0.671T_b - 0.193P_2 + 1.203P_3 + 17.195 \quad (6-5)$$

实验证明, 该模型原理简单, 结果可靠, 置信度高。

柴油是一类混合可燃液体, 是完全互溶的混合物, 柴油整体表现出来的闪点大多低于各组分闪点的平均值, 接近于含量大的组分的闪点。

对于混合物的闪点 (单位为 K) 估算, Catoire 提出含有二级基团校正的 S-C-P 法, 计算式为:

$$T_{FP} = 216 + \sum n_i \Delta F_i + W \sum n_j \Delta F_j \quad (6-6)$$

基团贡献值通过查表就可得到  $\Delta F_i$  和  $\Delta F_j$  的值。但由于此法较为复杂，需要事先查看表中的附注及实例。

## 6.2 近红外法测量柴油闪点

柴油作为一类复杂有机混合物，成分含量也会因原产地、加工过程的不同而产生较大的差别，从而导致性质测量的难度加深。徐广通等<sup>[52]</sup>利用电荷偶合传感器件测得柴油样本的近红外光谱区的吸光度，实现了不同种类柴油闪点等性质的标定。将近红外分析方法引入到油品性质分析领域，有着分析快速、成本低廉、重复性好，能够实现在线分析等优点。

徐广通等<sup>[39]</sup>还对齐鲁石化公司胜利炼油厂的成品柴油的闪点等性质通过光谱长波和短波区域的波长选择建立了偏最小二乘校正模型，所得测试结果与标准测定结果比较能够满足再现性要求。

王艳斌等<sup>[49]</sup>对不同馏程的宽范围的柴油进行基于近红外光谱的柴油闪点预测，解决了由于样本范围较广带来的与性质非线性的问题，神经网络方法的预测结果对光源改变、样本池、温度等影响因素的敏感程度明显低于主成分回归和偏最小二乘法，抗干扰性较强。

## 6.3 柴油闪点性质基团法预测模型

对于柴油密度性质的非空样品，将对应样本的原始光谱数据进行小波去噪、平滑求导处理后，从预处理后的柴油近红外光谱数据集(共含 395 组样本)中随机选择 60 组作为校正集用来建立预测模型，从剩余样本中随机选择 20 组作为测试集采用预测集残差(e)、预测集相对误差(re)、预测残差平方和(PRESS)及预测样本的标准偏差(SEP)作为模型评价标准。

### 6.3.1 多元线性回归模型

将上述选定的 60 组柴油样本的近红外光谱数据值及密度性质测定值进行多元线性回归(MLR)。但是这涉及到需要知道与性质对应的特征光谱区域，并且在选择拟合的方程维数方面，考虑到参与回归的变量数(波长点数)不能超过校正集的样本数目，也就是说波长点数选取有限制，这样难免丢失部分有用的光谱信息，也会造成选取信息的不全面，存在无用信息的选取有用信息丢失的问题。还应该考虑到的就是光谱矩阵可能

存在共线的问题，导致算法程序无法求解逆矩阵。

为了解决这个参与回归变量数量以及共线性问题，利用选择合适的相关系数作为在对样本了解信息不全面的情况下确定参与建模光谱维数的方法。

选取不同的相关系数，各评价指标数值如表 6-1:

表 6-1 选取不同波长点数的 MLR 模型的评价指标

Table 6-1 The evaluating indicator at different components of MLR model

相关系数	波长点数	PRESS	SD	最大e	最大re
0.60	5	291.46	3.92	8.46	0.13
0.587	8	276.96	3.82	7.93	0.117
0.575	12	232.57	3.50	7.92	0.118
0.57	16	701.76	6.08	10.72	0.182
0.56	19	878.02	6.70	12.98	0.22
0.53	27	683.0	5.99	9.98	0.17
0.52	32	634.36	5.79	10.34	0.18

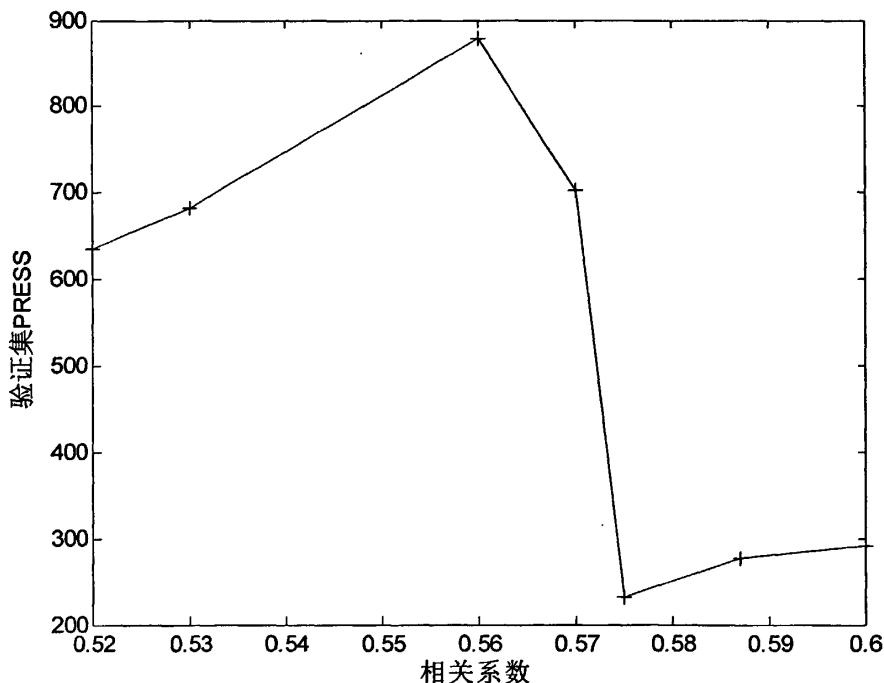


图 6-1 不同的 R 与 PRESS 的关系

Fig 6-1 PRESS value at different R

选择 12 个波长点数进行多元线性回归建模最好，预测集结果见图 6-2 和图 6-3:

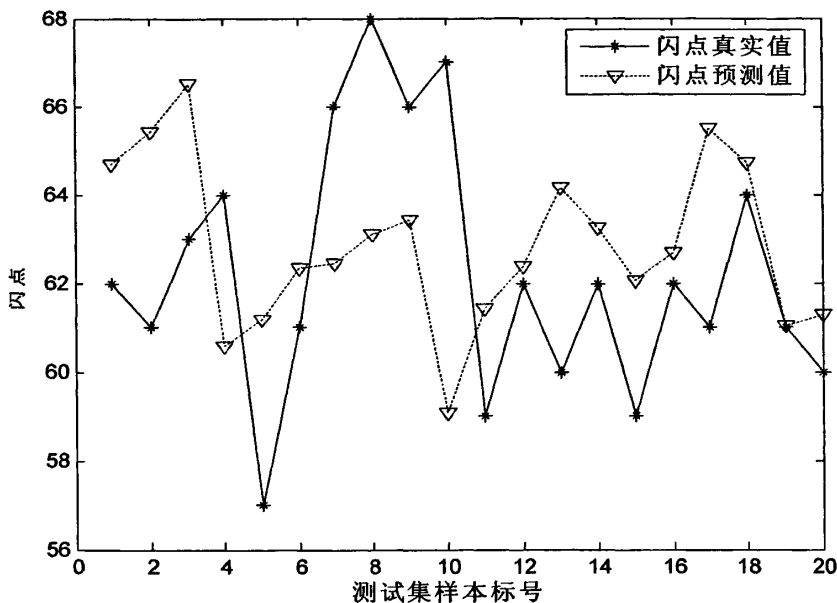


图 6-2 MLR 模型测试集真实值和预测值比较

Fig 6-2 The comparison between true and predictive of MLR method

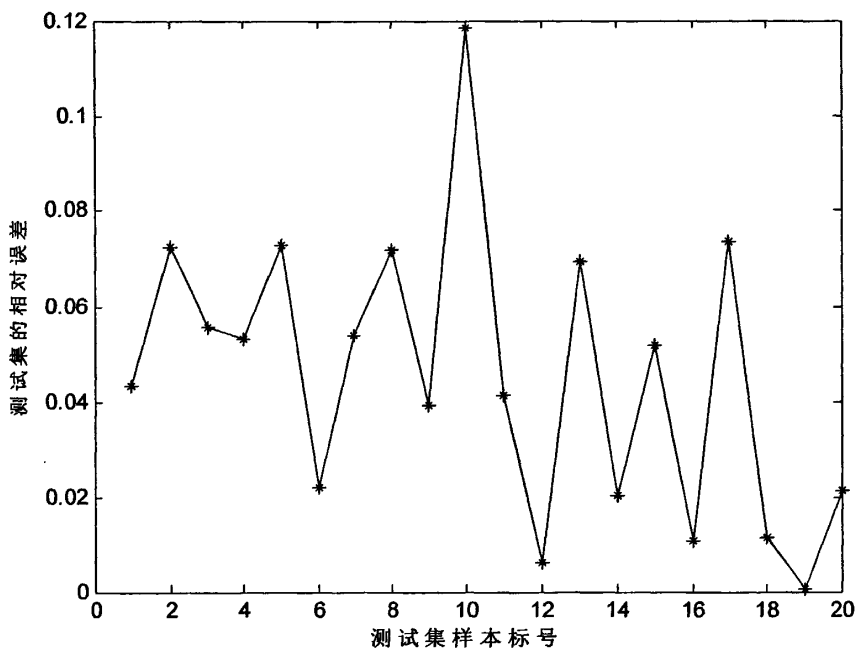


图 6-3 MLR 模型测试集的相对误差

Fig 6-3 The relative predictive error of MLR method

### 6.3.2 偏最小二乘模型

将上述选定的 60 组柴油样本的近红外光谱数据值及密度测定值进行偏最小二乘法



(PLS) 回归。

选取不同的主成分数进行建模，用预测残差平方和、标准偏差、预测集残差、预测集相对误差来评价所建模型的好坏。

表 6-2 选取不同主成分的 PLS 模型的评价指标

Table 6-2 The evaluating indicator at different principal components of PLS model

主成分数	PRESS	SD	最大 e	最大 re
3	431.10	4.76	8.54	0.133
4	377.84	4.56	8.73	0.128
5	290.77	3.91	7.54	0.111
6	297.50	3.96	7.58	0.124
7	264.91	3.73	8.44	0.138
8	191.26	3.17	6.32	0.102
9	165.03	2.95	7.3	0.124
10	188.14	3.15	9.08	0.154
11	232.68	3.50	8.05	0.136

不同的主成分数与 PRESS 的关系图见图 6-4:

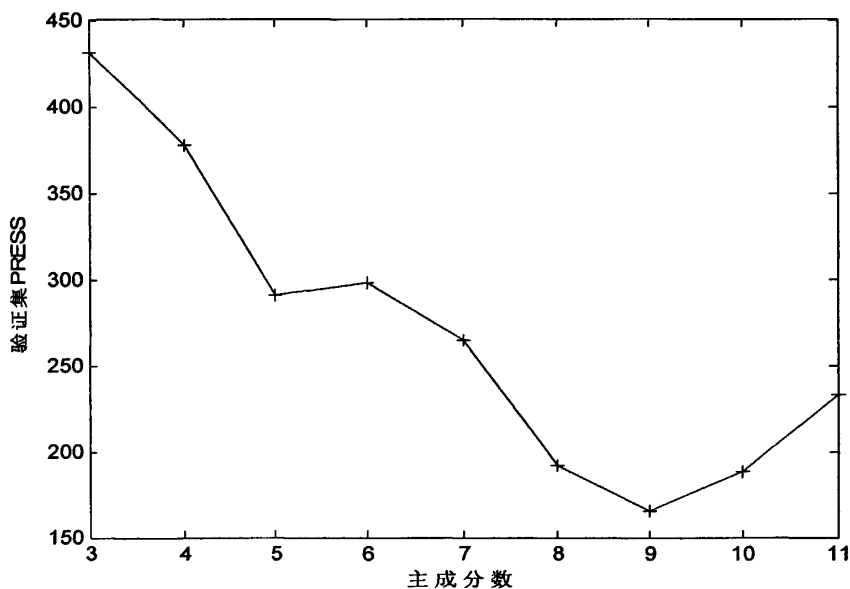


图 6-4 不同的主成分与 PRESS 的关系

Fig 6-4 PRESS value at different principal components

选择主成分数为9时，预测结果如图6-5和图6-6:

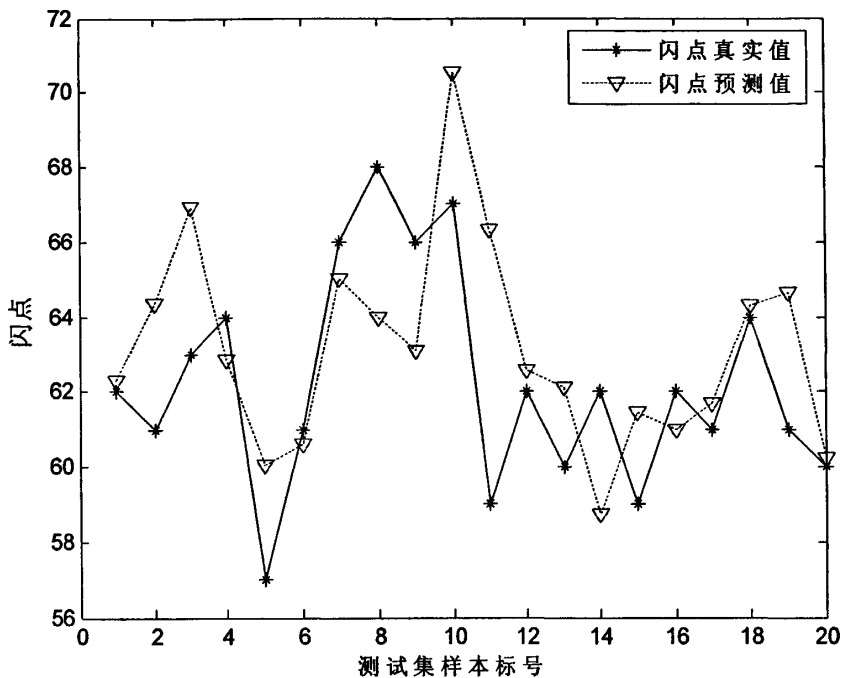


图 6-5 PLS 模型测试集真实值和预测值比较

Fig 6-5 The comparison between true and predictive of PLS method

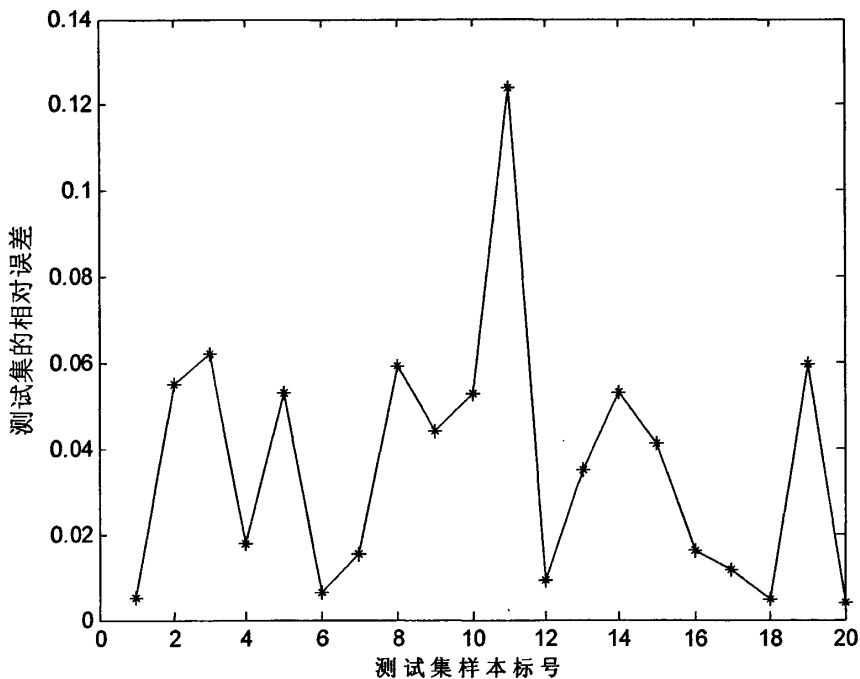


图 6-6 PLS 模型测试集的相对误差

Fig 6-6 The relative predictive error of PLS method

### 6.3.3 柴油闪点性质的基团法预测模型

经过小波去噪和二阶微分处理后的光谱数据图与前章类似。由于闪点是体现了柴油的轻组分的一种性质，所以在建模时不把芳烃这类重组分的对应波长的吸光度选入建模的范围内。

选择甲基、亚甲基和烯羟基团对应的波长建立简单线性模型，模型见式（6-7）：

$$y = -12535.6911X_{11} - 1337.4414X_{12} + 48536.7129X_{13} - 143592X_{21} + 26223.8724X_{22} - 8480.3199X_{23} - 23500.5719X_{24} + 22221.2941X_4 \quad (6-7)$$

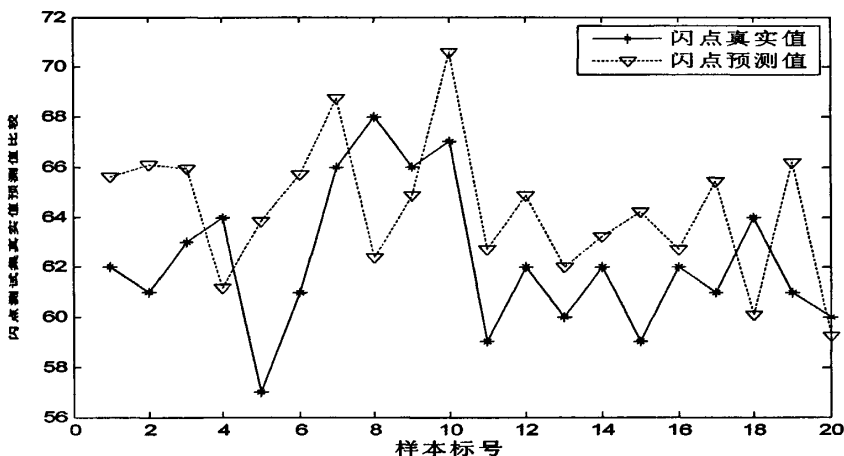


图 6-7 简单线性模型测试集真实值和预测值比较

Fig 6-7 The comparison between true and predictive of simple group method

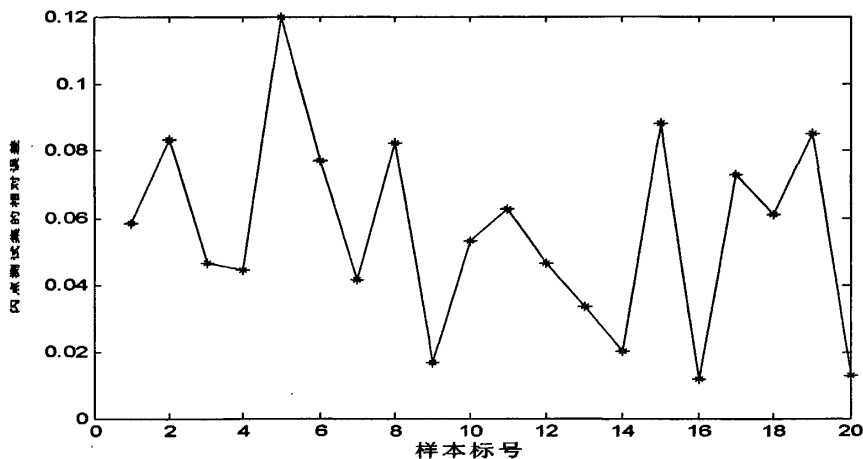


图 6-8 简单线性模型测试集的相对误差

Fig 6-8 The relative predictive error of simple group method

下面是对它进行简单改进的算法，添加了  $X_5$  和  $X_6$  表征支链化和烯烃度的项，具体拟合公式 (6-8)：

$$y = -3440.4035X_{11} + 427.0555X_{12} - 12670.5603X_{13} - 143485.7X_{21} + 10105.5141X_{22} + 12520.2906X_{23} - 16697.2911X_{24} + 226136.0979X_4 + 18.2284X_5 + 6660.065X_6 \quad (6-8)$$

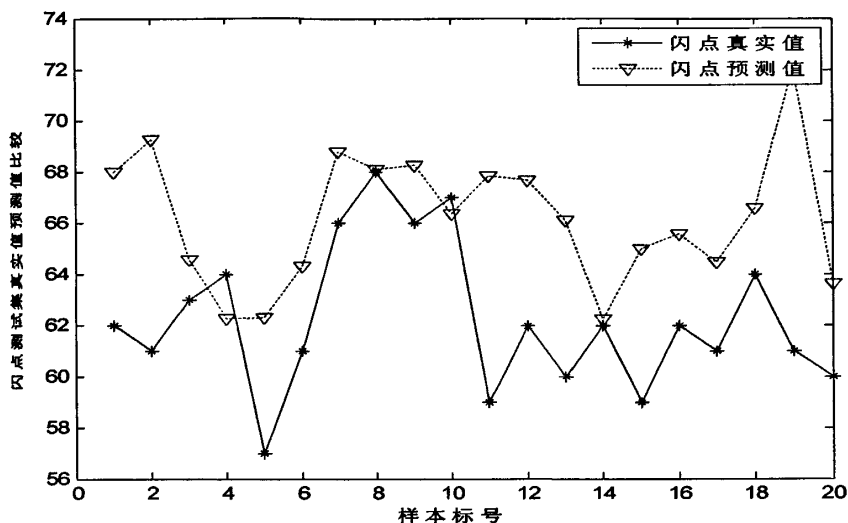


图 6-9 基团组合模型测试集真实值和预测值比较

Fig 6-9 The comparison between true and predictive of combined group method

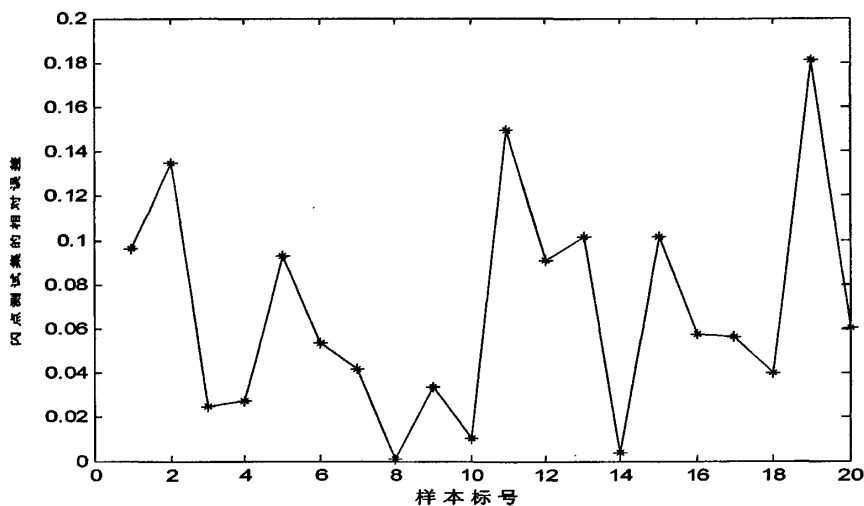


图 6-10 基团组合模型测试集的相对误差

Fig 6-10 The relative predictive error of combined group method

式 (6-7) 预测结果 PRESS 为 294.5536, 最大的预测误差为 6.8368, 最大预测的相对误差为 0.1199。式 (6-8) 预测结果 PRESS 为 508.6195, 最大的预测误差为 11.0403, 最大预测的相对误差为 0.181。可以组合的基团法所见模型优于只是选择特征波长的基团法模型。

考虑到  $X_{11}$ 、 $X_{12}$ 、 $X_{13}$ 、 $X_{21}$ 、 $X_{22}$ 、 $X_{23}$ 、 $X_{24}$ 、 $X_4$  是光谱吸光度数据, 而  $X_5$ 、 $X_6$  是表征组分结构的数据, 对吸光度数据进行 e 的指数运算, 采用 L-M 算法, 建立形如式(6-9) 的密度预测模型。

$$f(X) = (3.0133e^{X_{11}} + 1.1681e^{X_{12}} + 2.9642e^{X_{13}} + 0.6453e^{X_{21}} + 12.0042e^{X_{22}} + 2.9362e^{X_{23}} - 2.9101e^{X_{24}} + 0.8612e^{X_4} + 10.3633X_5 + 28.8615X_6)^{1.65} \quad (6-9)$$

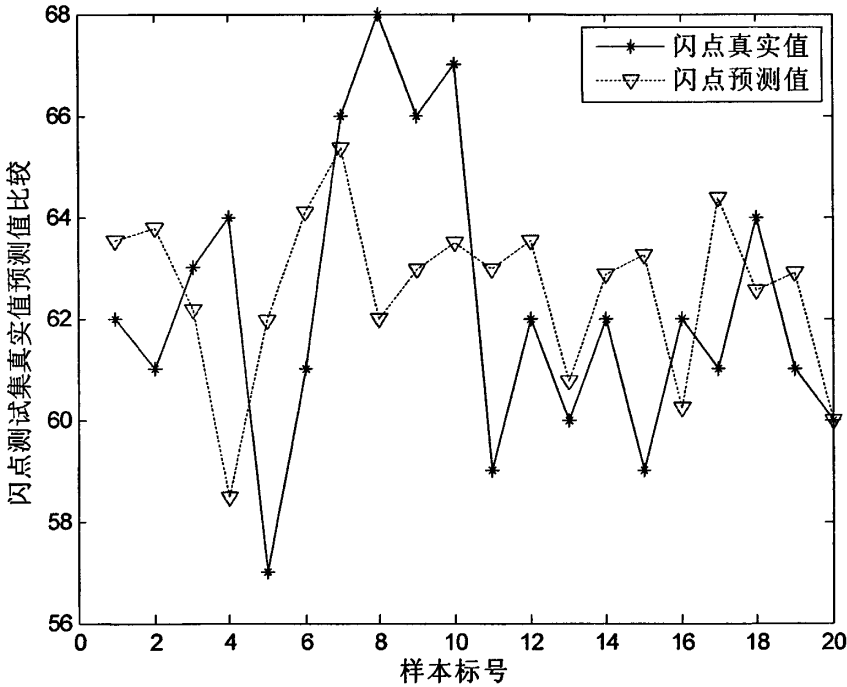


图 6-11 L-M 模型测试集的真实值和预测值比较

Fig 6-11 The comparison between true and predictive of L-M group method

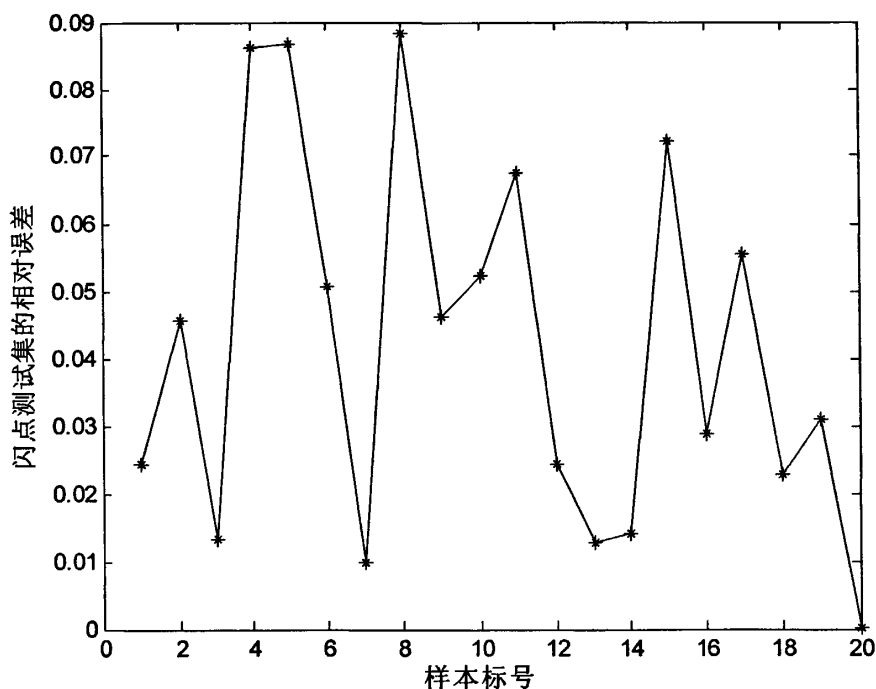


图 6-12 L-M 模型测试集的相对误差

Fig 6-12 The relative predictive error of L-M group method

## 6.4 结论比较

通过进行多元线性回归、偏最小二乘的拟合实验和利用基团选择的线性非线性实验，结果比较可见表6-3：

表 6-3 各模型测试集预测结果比较

Table 6-3 The comparison among evaluating indicators at different model

评价指标	MLR	PLS	线性基团法	组合基团法	L-M基团法
PRESS	232.57	165.03	294.55	508.62	190.81
SD	3.50	2.95	3.94	5.17	3.17
最大e	7.92	7.3	6.84	11.04	6.00
最大re	0.118	0.124	0.12	0.181	0.088

通过对表中数据的比较，可以看出非线性的 L-M 基团法在整体样本的预测性方面比 PLS 稍差一些，但对单个样本预测性方面优于其他各模型的预测效果。

## 总结与展望

本文通过对近红外光谱吸收带对应油品中的特征基团进行分析，研究了柴油的黏度、密度、闪点性质与柴油成分的组成结构的关系，建立了由样本的基团成分含量和结构信息与性质之间的机理模型。由于近红外光谱的吸收带产生主要是由于含有各种 C-H 基团的组成分子在吸收了近红外辐射的光子或能量后引发的化学键的振动引起的，而石油产品尤其是组成柴油产品的各种化合物都可以看作是由这些基本结构基团构成的，它们之间相互作用，共同决定整个柴油产品样本的一些性质。本文主要的研究工作和取得的成果可总结为：

通过查阅学习相关的中外文资料，对近红外光谱技术应用在分析领域的发展现状进行了介绍。对构建基于近红外光谱的性质模型的理论背景和现状进行了详细的描述。在对光谱样本预处理的前提下，选择对理论上对性质有贡献影响的特征吸收谱带的的数据，关联推测和一些样本的结构性质（支链化度和烯烃度），将这些光谱数据和性质数据作为自变量，用于建立与因变量性质间的关系模型，由于这些性质不是某种成分的浓度值，因而建立的非线性模型预测精度更好一些。非线性模型的回归参数采用 L-M 算法求解，避免了 N-G 算法对病态矩阵不使用情况。实验证明采用特征波长处吸光度和表征结构性质的数据共同构建的非线性模型在验证测试集的柴油黏度、密度、闪点性质的问题中得到较好的分析结果。

论文采用随机选定的一组柴油样品光谱及黏度、密度、闪点性质数据分别建立了 MLR、PLS、简单基团法、组合基团法和非线性基团法定量预测模型，并对这五种模型进行了评价。研究表明，由于近红外光谱数据信息度有限，在基团法构建的各模型选择基本基团如甲基、亚甲基、芳烃、烯烃和它们的组合表征成分分子结构特点作为参数进行对性质的预测建模。由于柴油性质和光谱数据间存在非线性的关系，由建模预测精度的结构可知，组合基团法的各项性能均优于简单基团法，L-M 基团法在黏度、密度的性质预测中，不论是整体预测能力还是个体预测精度，均是五种算法中最优的。在闪点的预测结果中，L-M 基团法对个体的预测能力也是五种算法中最好的，在整体预测能力方面与偏最小二乘方法相当。实验结论可以认定认为非线性的基团法模型具有更好的泛化性，在对未知样品进行预测的过程中会取得更好的效果。

近红外光谱分析方法作为一种快速无损的分析方法，用在组分结构复杂柴油产品等性质的检测中，能实现一些传统的校正技术难以达到分析目的，如何进一步利用基团性

质体现混合物整体性质，并对这些技术加以改进，降低预测误差，提高模型的稳健性，拓宽它的适应范围，是一个很值得深入研究的方向。



## 参考文献

- [1] 寿德清, 王从岗, 向正为等. 国产石油物性研究的进展[J]. 石油大学学报, 1993, 17(7):254-261
- [2] Roman M.Balabin, Ravilya Z, Safieva. Gasoline classification by source and type based on near infrared(NIR)spectroscopy data [N]. Fuel, 2008, 87:1096-1101
- [3] Alfred A, Christy, Sumaporn Kasemsumran, Yiping Du. The Detection and Quantification of Adulteration in Olive Oil by Near-Infrared Spectroscopy and Chemometrics [C]. Analytical Sciences, 2004, 4(20):935-940
- [4] 冯新泸, 史永刚. 近红外光谱及其在石油产品分析中的应用[M]. 北京: 中国石化出版社, 2002:4-5; 90-112; 201-233
- [5] 陆婉珍, 袁洪福, 徐广通等. 现代近红外光谱分析技术(第二版) [M]. 北京: 中国石化出版社, 2007: 29-30; 38-46
- [6] 徐广通, 刘泽龙, 杨玉蕊. 近红外光谱法测定柴油组成及其应用[J]. 石油学报, 18(4): 65-71
- [7] Jeffrey J Kelly, James B Callis. Nondestructive Analytical Procedure for Simultaneous Estimation of the Major Classes of Hydrocarbon Constituents of Finished Gasolines[J]. 1990, 62(14):1444-1451
- [8] 郭纯孝. 计算化学[M]. 北京: 化学工业出版社, 2004:288-289; 306-308
- [9] 谢小伟. 浅析多元线性回归中多重共线性问题的三种解决方法. 高校理科研究, 09(28): 117-118
- [10] Edward V, Thomas. Adaptable Multivariate Calibration Models for Spectral Applications[J]. Analytica Chemica. 2000, 72: 2821-2827
- [11] 陈锴, 张岩. 化学计量学方法在近红外光谱分析中的应用[J]. 石油学报, 20(4): 93-95
- [12] 陈朋山, 焦伟利, 贾秀朋等. 抗差 LM 算法求解遥感影像严格物理模型[J]. 科学技术与工程. 2009, 8. 9(16):4614-4618
- [13] Jerry Workman, Jr. Lois Weyer. Practical Guide to Interpretive Near-Infrared Spectroscopy [M]. 北京: 化学工业出版社, 2009:19-42
- [14] Del Fanti N A, Bradley K, Bradley M, ect. IR Spectroscopy of polymers[J].

ThermoFisher Scientific Corp., 2007: 162

- [15] Murray I, Williams P C. Chemical principles of near-infrared technology in the agriculture and food industries [M]. Williams P, Norris K. American Association of Cereal chemists, St. Paul, MN, 1987: 17-34
- [16] Ricard-Lespade L, Longhi G, Abbate S. The first overtone of C-H stretchings in polymethylene chains: a conformationally dependent spectrum [J]. Chem Phys, 1990, 142: 245-259
- [17] Iwamoto R, Nara A, Matsuda T. Near-Infrared combination and overtone bands of CH in  $\text{CHX}_3$ ,  $\text{CHX}_2\text{-CHX}_2$  and  $\text{CHX}_2\text{-CX}_2\text{-CHX}_2$  [J]. Appl Spectrosc, 2005, 59(11): 1393-1398
- [18] Snavely D L, Angevine C. Near-infrared spectrum of polybutadiene [J]. J Polym Sci: Part A, 1996, 34: 1669-1673
- [19] Snavely D L, Overly A, Walters V A. Vibrational overtone spectroscopy of pyridine and related compounds [J]. Chem Phys, 1995, 201: 567-574
- [20] Bini R, Foggi P, Della Valle R G. Vibrational analysis of C-H stretching overtones in pyridine and 2,6-lutidine [J]. Phys Chem, 1991, 95: 3027-3031
- [21] Southwest Research Institute (SWRI). Near Infrared Spectra of Diesel Fuels [EB/OL]. <http://software.Eigenvector.com/Data/index.html>, 1999.
- [22] S.A. Hutzler, S.R. Westbrook. Estimating chemical and bulk properties of middle distillate fuels from near-infrared spectra [R]. TFLRF NO. 348, Washington DC: SWRI, 2000.
- [23] C. T. O' Connor, R. D. Forrester, M. S. Scurrill. Cetane number determination of synthetic diesel fuels [J]. FUEL, 1992, 71(11): 1323-1327
- [24] 陈俊武, 曹汉昌. 催化裂化过程中物料化学结构组成变化规律的探讨 (上) [J]. 石油炼制与化工. 1994, 26(8): 1-7
- [25] 盛卫心, 戎宗明, 英徐根. 基团贡献法分子设计研究的进展 [J]. 化学工业与工程. 2007, 9. 24(5): 457-465
- [26] 李会鹏, 李会举, 沈本贤. 柴油低温黏温特性和流变性研究 [J]. 化工科技. 2006, 14(2): 8-12
- [27] 马沛生. 化工数据教程 [M]. 天津: 天津大学出版社, 2008: 104-107; 141-148

- [28] Sastri S R S, Rao K K. Chem Eng J, 1992, 50:9
- [29] 韦锐, 张德志. 用便携式油料质量分析仪测定 3 号喷气燃料的性质[J]. 石油商技. 2003, 2. 21(1):48-51
- [30] 王艳斌, 袁洪福, 陆婉珍. 近红外分析方法测定润滑油基础油粘度指数[J]. 润滑油. 2001, 12. 16(6): 53-56
- [31] 冯新泸, 罗平亚, 史永刚等. 发动机润滑油粘度的近红外光谱研究[J]. 计算机与应用化学. 2003, 5. 20(3): 277-280
- [32] 宋延东. 润滑油基础油粘度性能的近红外光谱研究[J]. 化工时刊. 2008, 4. 22(4): 9-12
- [33] 杨庭栋, 冯新泸. 近红外光谱法定性分析内燃机油的粘度特性[J]. 合成润滑材料. 2004, 31(1):13-16
- [34] 冯新泸, 罗平亚, 李子存等. 自组织神经网络定量分析内燃机油粘度指数[J]. 石油炼制与化工. 2003, 12. 34(12):44-48
- [35] 王克强. 预测烷烃密度的新方法——基团键贡献法[J]. 有机化学. 1999. 19: 304-308
- [36] 冯瑞英, 王克强. 芳烃密度与其分子结构的关联及计算[J]. 成都师专学报. 2000. 4: 41-45
- [37] 徐广通, 陆婉珍. 柴油近红外光谱与性质的相关性分析[J]. 石油学报(石油化工). 2001, 4. 17(2):91-95
- [38] 徐广通, 陆婉珍, 袁洪福等. CCD 近红外光谱仪在柴油生产控制分析中的应用[J]. 石油炼制与化工. 1999, 9. 30(9):57-61
- [39] 徐广通, 袁洪福, 陆婉珍等. 近红外光谱仪技术快速测定柴油物理性质[J]. 石油学报(石油化工). 1999, 10. 15(5):63-68
- [40] 方利民, 林敏. 柴油近红外光谱的独立分量分析方法[J]. 石油学报(石油化工). 2008, 12. 24(6):726-732
- [41] 王志芳, 夏伯锴. 用基团法预测柴油密度值[J]. 化工自动化及仪表. 2009. 36(1):67-70
- [42] 周山丹. 对石油产品闪点测定的讨论[J]. 科技信息. 2009. 21:44
- [43] 孔锐睿, 代长江, 仇汝臣等. 原油常压产品闪点数学模型的研究[J]. 山东化工. 2009. 38(2):42-43
- [44] 刘斌, 安正阳. 混合可燃液体闪点的近似估算[J]. 昆明理工大学学报(理工

- 版). 2008, 10. 33(5):68-71
- [45] 潘勇, 蒋军成. 电性拓扑状态指数预测烃类物质闪点[J]. 石油学报(石油加工). 2007, 12. 23(6):70-74
- [46] 潘勇, 蒋军成, 王志荣. 人工神经网络基团键贡献法预测烷烃闪点[J]. 化学工程. 2007, 4. 35(4):38-41
- [47] 王克强, 杜仲详, 王捷. 有机化合物的闪点与分子组成的关联及预测[J]. 科技通报. 2002, 5. 18(3):234-239
- [48] 高礼久, 秦正龙. 烷烃、烯烃的定量结构与闪点的相关性研究[J]. 徐州师范大学学报(自然科学版). 2007, 12. 25(4):66-68
- [49] 王艳斌, 袁洪福, 陆婉珍等. 人工神经网络用于近红外光谱测定柴油闪点[J]. 分析化学. 2000, 9. 28(9):1070-1073

## 致谢

感谢夏伯锴教授，是他将我领入了近红外领域，并对我的研究提出了很多宝贵的意见，给予了及时的指导，使我的研究工作有了目标和方向。在这三年时间里，他对我进行了悉心的指导和教育，使我的学习和研究能够不断的提升，而且这些课题的研究成果也成为了本论文的主要素材。同时，夏老师渊博的学识、严谨的治学态度也令我十分敬佩，是我以后学习和工作的榜样。还要再次感谢夏老师对我的关心和照顾，在此表示我最诚挚的敬意和感谢。

感谢我的父母，是他们为我提供了一个在学校安心读书的机会。最后，感谢所有关心过，帮助过我的老师和同学。