

甲醇汽油、乙醇汽油定性判别及其醇含量测定模型研究

胡 军, 刘燕德*, 郝 勇, 孙旭东, 欧阳爱国

华东交通大学机电与车辆工程学院, 江西 南昌 330013

摘 要 甲醇汽油和乙醇汽油都为清洁能源, 但甲醇汽油和乙醇汽油的优缺点各有不同, 其中甲醇、乙醇的含量决定了汽油性能的优劣, 对甲醇汽油和乙醇汽油进行判别区分以及醇类汽油中醇含量进行定量测定非常重要。通过中红外光谱技术对醇类汽油的类型进行判别并对其含量进行定量分析。首先通过对比分析甲醇汽油和乙醇汽油的中红外光谱图, 采用随机森林(random forest, RF)对甲醇汽油和乙醇汽油样品进行判别; 在建立甲醇汽油和乙醇汽油样品定性判别模型之后, 分别建立甲醇汽油和乙醇汽油的定量测定模型, 从而精确测定汽油中对应的醇的含量。为减小在实验过程中实验仪器振动、噪声等原因导致的光谱漂移、光散射等现象, 对中红外光谱进行预处理。首先采用不同预处理, 如(savitzky-golay, S-G)卷积平滑、多元散射校正(multiplicative scatter correction, MSC)、标准正态变量变换(standard normal variable transformation, SNV)、导数(derivatives)等方法进行校正, 分别建立适合甲醇汽油和乙醇汽油的检测模型。预处理后的数据分别建立甲醇汽油、乙醇汽油的最小二乘支持向量机(least square support vector machine, LS-SVM)模型。采用随机森林(random forest, RF)对甲醇汽油和乙醇汽油样品进行判别, 发现当决策树个数为 61 时, 判别正确率达到 98.28%。对于 LS-SVM 模型, 比较建模结果可知: 无论是甲醇汽油还是乙醇汽油, 标准正态变量变换(SNV)预处理效果最好, 经 SNV 校正处理后建立的甲醇汽油甲醇含量测定 LS-SVM 模型的预测相关系数 R_p 为 0.9519, 均方根误差(root mean square error of prediction, RMSEP)为 1.766 3; 经过标准正态变量变换后建立的乙醇汽油乙醇含量测定 LSSVM 模型的预测相关系数 R_p 为 0.951 5, 均方根误差 RMSEP 为 1.770 3。该研究可为甲醇汽油、乙醇汽油的定性判别和其含量测定提供技术参考和理论依据, 为甲醇汽油产业提供测量醇类汽油检测的新方法, 具有较为重要的现实意义, 也为其他类型的化工产品的检测奠定了基础。

关键词 中外光谱; 醇类汽油; 最小二乘支持向量机; 随机森林

中图分类号: O657.61 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)05-1640-05

引 言

汽油、柴油等传统的燃料在燃烧过程中会产生对环境有害的气体及一些固体颗粒, 不仅污染环境, 而且会直接影响人类的健康。因此世界各地都在探索新的能源或尝试减少燃料在燃烧过程中有害物的产生, 以此减少对环境的污染^[1]。甲醇汽油和乙醇汽油的使用可以提高汽油的含氧量, 从而使得汽油燃烧更加充分, 并且能够有效降低有毒气体的排放^[2]。

甲醇汽油和乙醇汽油虽都可作为燃料, 但是由于甲醇和乙醇的理化性质并不完全相同, 配制而成的甲醇汽油和乙醇汽油也存在很多不同特性^[3]。例如, 相对于乙醇汽油, 甲醇汽油的热量值高, 作为燃料使得汽车的动力更足, 但是甲醇具有腐蚀性, 乙醇汽油热量值虽低于甲醇汽油但却更加稳定^[4-5]。但若汽油所含的甲醇或乙醇含量过高, 其燃烧易造成气阻, 使得发动机的气密性下降, 导致动力不足。因此, 评估该醇类燃料质量的关键之一在于确定其醇类的含量。不同的甲醇汽油和乙醇汽油对其醇的含量有着不同的规定, 因此, 探索一种精准度高、快速性好的方法来判定醇类燃料汽

收稿日期: 2019-06-25, **修订日期:** 2019-10-30

基金项目: 国家高技术研究发展计划(863 计划)项目(SS2012AA101306), 江西省科技支撑计划项目(20121BBF60054), 南方山地果园智能化管理技术与装备协同创新中心(赣教高字[2014]60 号)和江西省优势科技创新团队(20153BCB24002), 国家自然科学基金项目(2002017018), 江西省教育厅科学技术研究青年项目(GJJ190348), 江西省博士研究生创新基金项目(YC2019-B106)资助

作者简介: 胡 军, 1992 年生, 华东交通大学机电与车辆工程学院讲师 e-mail: 1825868002@qq.com

* 通讯联系人 e-mail: jxliuyd@163.com

油的类型并同时测定其含量非常重要^[6]。

近年来,国内外研究工作者开始对燃料汽油的成分开展研究。姚捷等^[7]针对甲醇汽油品质的检测,利用了拉曼技术,并且使用一元线性回归法对其建模,验证了拉曼技术对甲醇汽油检测的可行性。欧阳爱国等^[8]利用近红外光谱和拉曼光谱技术,对汽油中甲醇的含量建立了偏最小二乘模型,该研究的结论证明了近红外光谱技术对甲醇汽油具有较好的检测效果。李雁如等^[9]利用蒸馏水萃取法与气相色谱分析技术相结合测定了不同规格车用甲醇汽油中的甲醇含量,结果表明:该方法可以检测,但需要对样品进行复杂的前处理,操作较为复杂。

目前,使用中红外光谱技术检测醇类汽油中醇含量的研究的报道相对较少,也很少建立甲醇汽油和乙醇汽油定性判别模型。本文采用中红外光谱检测技术结合随机森林(random forest, RF)算法^[9-10]建立甲醇汽油和乙醇汽油定性判别模型,然后分别建立偏最小二乘支持向量机的甲醇汽油和乙醇汽油定量检测模型。

1 实验部分

1.1 材料

实验用汽油(92号)在某中石化加油站购买,实验用无水甲醇和无水乙醇纯度 $\geq 99.7\%$ 。为增加实验的复杂性,综合实际情况,甲醇汽油和乙醇汽油设置不同浓度梯度。每配制一个甲醇样品,其浓度比例增加 0.6% ,甲醇的浓度在 $0.6\% \sim 18.6\%$ 的范围;每配制一个乙醇样品,其浓度比例增加 0.8% ,其浓度在 $0.8\% \sim 25\%$ 范围。先按照对应的浓度梯度,分别计算出在不同浓度梯度下所需的甲醇、乙醇体积,在容量瓶中加入之前计算出的所需的甲醇、乙醇,然后分别滴入所购买的92号汽油至容量瓶50 mL刻度处,充分摇匀,使汽油与其完全混合。依次按照同样的方法,分别配制29个浓度梯度的甲醇汽油、乙醇汽油。每个浓度梯度配制4个样品,分别配制各116个样品。

1.2 光谱采集

实验采用中红外光谱仪(型号为WQF-510A),实验室的环境湿度 $\leq 30\%$,温度设置为恒温 $25\text{ }^\circ\text{C}$ 左右,扫描范围为: $4\ 000 \sim 900\ \text{cm}^{-1}$,分辨率 $\geq 0.85\ \text{cm}^{-1}$,开机仪器预热约30 min待设备稳定后进行相应操作。利用中红外水平衰减全反射(attenuated total reflection, ATR)附件呈放待测样品,在附件上均匀涂抹对应浓度样品溶液,然后关闭仪器的舱门,分别采集甲醇汽油、乙醇汽油的中红外光谱信息。在MATLAB2014b软件中利用化学计量学方法进行数据处理。为减少测量过程中的随机误差,实验对每个样品进行5次光谱采集,为减少样本挥发带来检测误差,每个样本控制在2min内测量完毕,然后求取平均光谱。

1.3 模型评价标准

首先对甲醇汽油、乙醇汽油样品的中红外光谱进行预处理,然后分别建立后续模型。建模集与预测集的数量约为 $3:1$ 。预测相关系数 R 和预测均方根误差(root mean square error of prediction, RMSEP)参数是本模型评价的关键。检测

模型的好坏由预测集 R 和RMSEP系数共同决定,检测模型预测均方根误差越小、相关系数越高,则模型的精度越高。此外,校正均方根误差(root mean square error of calibration, RMSEC)与RMSEP越接近,说明建立的模型越稳定。

2 结果与讨论

2.1 中红外光谱响应特性

图1为92号汽油的红外原始谱图,92号汽油官能团基频振动的信息体现在波长范围为 $1\ 350 \sim 1\ 610$ 和 $2\ 850 \sim 3\ 000\ \text{cm}^{-1}$ 处。官能团 CH_3 中 $\text{C}-\text{H}$ 的弯曲振动体现在 $1\ 376, 1\ 457, 2\ 870$ 和 $2\ 960\ \text{cm}^{-1}$ 等处的强吸收峰,不饱和键 $\text{C}=\text{C}$ 的伸缩振动导致在 $1\ 604\ \text{cm}^{-1}$ 处出现多峰, CH_2 中 $\text{C}-\text{H}$ 的拉伸振动在 $2\ 925\ \text{cm}^{-1}$ 出现最强的吸收峰^[12]。

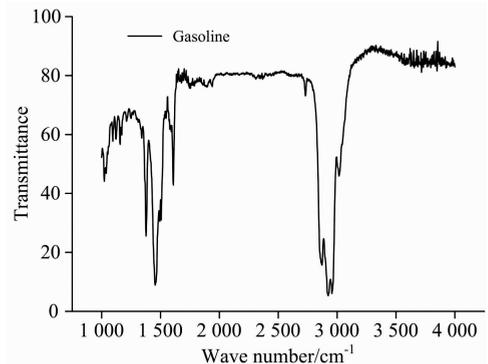


图1 汽油、甲醇汽油以及乙醇汽油的中红外光谱图
Fig. 1 The original mid-infrared spectra of gasoline

图2为甲醇汽油和乙醇汽油的原始中红外光谱,同浓度的甲醇汽油和乙醇汽油中红外原始光谱波形基本一致,只是在一定的波段内的光强有区别,同样含量的乙醇汽油中红外的透射更强。

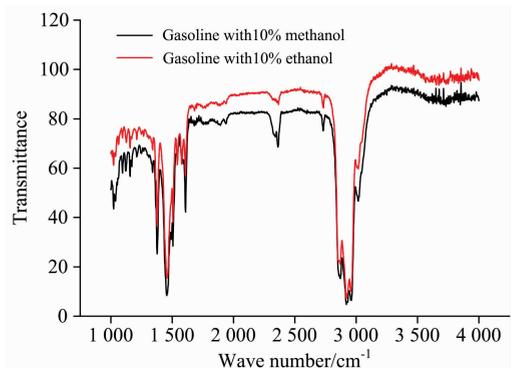


图2 汽油、甲醇汽油以及乙醇汽油的中红外光谱图
Fig. 2 The original mid-infrared spectra of methanol gasoline and ethanol gasoline

2.2 甲醇汽油、乙醇汽油随机森林判别模型的建立

在醇类汽油混合样品的原始光谱数据中,波段范围为 $1\ 000 \sim 4\ 000\ \text{cm}^{-1}$,甲醇汽油116个样品,乙醇汽油116个样品,共得到了232条中红外光谱数据,将甲醇汽油标记为

“-1”，将乙醇汽油标记为“1”，并对其构建定性模型进行判别分析^[11-12]。

随机森林的实质是一个包含了多个决策树的分类器，随机森林中的树之间没有相互关联。在构建决策树时，采用随机选取分裂属性集的方法。Ntree 为随机森林中决策树的个数。首先，采用 K-S 算法^[13]将 232 组中红外光谱数据分成两组，分别为建模数据和预测数据，其所占的光谱数据分别为 174 个和 58 个。再预先设置好参数，决策树的个数(Ntree)设为 500 个，树节点预选变量个数(Mtry)值是光谱变量数的平方根(即向下取整)，约为 39，再构造随机森林分类模型。如图 3 所示为带外错误率(OOB)误差率与决策树个数关系，当决策树个数为 61 时，OOB 误差率获得最小值，为 0.034 5。图 4 为变量贡献率的变化图，波段范围在 1 508, 2 298 和 2 312~2 294 cm^{-1} 附近的变量对甲醇汽油、乙醇汽油的分辨具有更佳的作用。

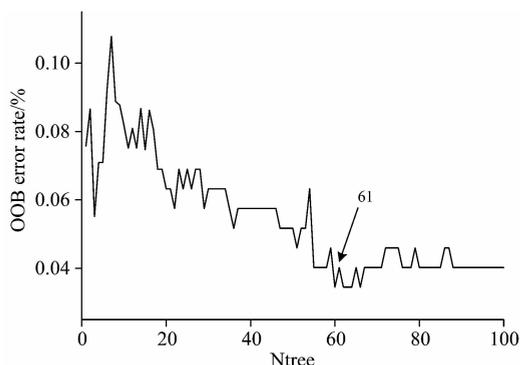


图 3 光谱带外数据误差率

Fig. 3 OOB error rate of spectra

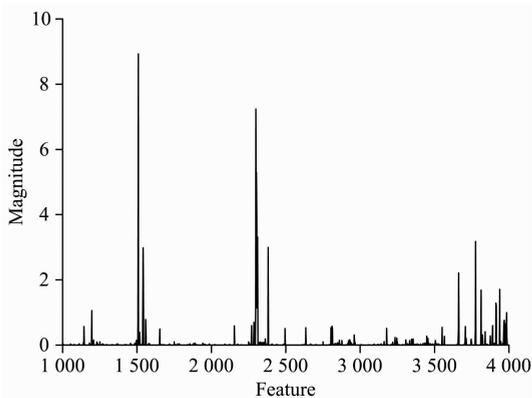


图 4 变量贡献率的变化图

Fig. 4 The contribution rate variation chart of the variables

表 1 为随机森林(RF)判别模型分类结果，甲醇汽油、乙醇汽油中模型分类的正确率为 98.28%，构造的模型对甲醇汽油和乙醇汽油样品的分类具有较好的效果。

表 1 RF 判别模型分类结果

Table 1 Classification results of RF discriminant model

样本	建模集	预测集	决策树	正确个数	正确率/%
232	174	58	61	57	98.28

2.3 甲醇汽油、乙醇汽油定量模型的建立及分析

2.3.1 甲醇汽油、乙醇汽油样品前处理

将预处理后的甲醇汽油和乙醇汽油中红外光谱数据，采用最小二乘支持向量机算法构建模型，并对不同预处理后建立的模型的预估能力进行比较。使用 KS(kennard-stone)算法按照 3:1 的比例，分别对预处理后的甲醇汽油和乙醇汽油样品数据进行划分(甲醇汽油和乙醇汽油各 112 个)，其中 84 个光谱数据用于建立了模型，28 个光谱数据用于对模型进行检验。

表 2 甲醇汽油、乙醇汽油的醇含量
在建模集和预测集的真实分布

Table 2 The actual distribution of alcohol content of methanol gasoline and ethanol gasoline in the modeling and prediction

类别	样品集分类	数量 N	最小值 /%	最大值 /%	平均值 /%	标准差 STDEV
甲醇汽油	全部样本	112	0.6	19.999	9.544 5	6.156
	建模集	84	0.6	19.999	9.624 8	6.065
	预测集	28	0.6	19.999	9.295 8	6.206
乙醇汽油	全部样本	112	0.8	24.000	12.414	6.996
	建模集	84	0.8	24.000	12.545	7.074
	预测集	28	0.8	24.000	12.000	6.853

Note: N: Number of samples; STDEV: Standard deviation

2.3.2 甲醇汽油、乙醇汽油 LS-SVM 定量模型的建立及分析

最小二乘支持向量机是一种针对支持向量机算法的优化形式，本模型基于 RBF 径向基函数，此建模方法必须确定两个参数，分别为正则化参数 γ 和核参数 σ^2 。建模效果的评估是基于醇类汽油醇含量的均方根误差 RMSEP 和相关系数 R。

表 3 为不同预处理后的 LSSVM 模型预测甲醇汽油、乙

表 3 LSSVM 模型预测甲醇汽油、乙醇汽油中
醇含量的建模结果

Table 3 The modeling results of alcohol content in methanol gasoline and ethanol gasoline predicted by LSSVM

类型	处理方法	γ	σ^2	R	RMSEP/%
甲醇汽油	Origin	$1.179 4 \times 10^6$	$1.731 9 \times 10^3$	0.933 8	1.919 0
	SG	$2.273 5 \times 10^5$	$1.994 5 \times 10^3$	0.929 3	1.985 3
	MSC	$1.263 9 \times 10^5$	$2.033 4 \times 10^3$	0.951 4	1.774 1
	SNV	$4.666 0 \times 10^4$	$1.915 5 \times 10^3$	0.951 9	1.766 3
	1st derivatives	$6.010 3 \times 10^4$	$8.721 6 \times 10^4$	0.876 5	2.774 8
	2nd derivatives	$6.712 8 \times 10^4$	$1.602 2 \times 10^5$	0.823 5	3.225 8
乙醇汽油	Origin	$5.453 1 \times 10^6$	258.020 3	0.937 8	2.399 8
	SG	$9.438 1 \times 10^5$	327.988 0	0.930 1	2.643 7
	MSC	40.316 7	845.716 7	0.922 0	2.614 3
	SNV	$1.919 6 \times 10^5$	$2.025 1 \times 10^3$	0.951 5	1.770 3
	1st derivatives	$9.895 1 \times 10^5$	$1.904 7 \times 10^4$	0.909 6	3.146 8
	2nd derivatives	$3.862 0 \times 10^5$	$2.943 6 \times 10^5$	0.899 8	3.242 8

醇汽油中醇含量的建模结果。从表中可以得出,无论是甲醇汽油还是乙醇汽油,运用 SNV 对其光谱数据预处理后,基

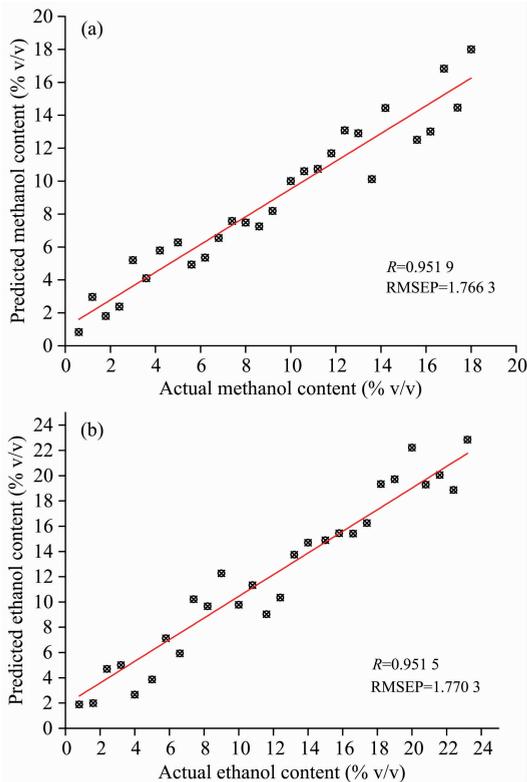


图 5 LSSVM 预测醇类汽油醇含量的散点图

(a): 甲醇汽油; (b): 乙醇汽油

Fig. 5 Scatter diagram of alcohol content in alcohol gasoline predicted by LSSVM

(a): Methanol gasoline; (b): Methanol gasoline

于 RBF 函数构造的 LS-SVM 模型效果更佳,甲醇汽油所对应的 R_p 和 RMSEP 的值分别是 0.951 9 和 1.766 3;乙醇汽油此时所对应的 R_p 和 RMSEP 的值分别是 0.951 5 和 1.770 3(见表 6)。说明使用 SNV 可以较好的消除样品表面散射以及光程差变化对中红外光谱的影响。经 SNV 预处理后,再与 LSSVM 模型相结合,可以较为精准的定量检测醇类汽油的醇含量。图 5(a,b)分别为 LSSVM 预测甲醇和乙醇汽油醇含量的散点图。

表 6 预测甲醇汽油、乙醇汽油中醇含量的模型对比结果

Table 6 Model comparison results for predicting alcohol content in methanol gasoline and ethanol gasoline

类型	Pretreatment	Model	Parameter	R_p	RMSEP/%
甲醇汽油	SNV	LS-SVM	$\gamma=4.666\ 0\times 10^4$ $\sigma^2=1.915\ 5\times 10^3$	0.951 9	1.766 3
乙醇汽油	SNV	LS-SVM	$\gamma=1.919\ 6\times 10^5$ $\sigma^2=2.025\ 1\times 10^3$	0.951 5	1.770 3

3 结 论

采用 RF 对甲醇汽油和乙醇汽油进行判别分析,正确率达到 98.23%。对所得的光谱数据使用 S-G 平滑、MSC、SNV,一阶导数以及二阶导数等多种方法进行预处理,应用 LSSVM 方法对其建模,对比模型结果可知,无论是甲醇汽油还是乙醇汽油,对于 LSSVM 模型,SNV 预处理效果最好。对于甲醇汽油模型,预测集相关系数 R_p 为 0.951 9, RMSEP 为 1.766 3,对于乙醇汽油模型,预测集相关系数 R_p 为 0.951 5, RMSEP 为 1.770 3。为今后现场快速检测醇类汽油提供了理论参考,具有较为重要的应用价值。

References

- [1] LIU Yan-de, HU Jun, TANG Tian-yi, et al(刘燕德,胡军,唐天义,等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2019, 39(2): 459.
- [2] OUYNAG Ai-guo, LIU Jun, WANG Ya-ping, et al(欧阳爱国,刘军,王亚平,等). Laser and Infrared(激光与红外), 2012, 42(8): 901.
- [3] DING Hong-yuan, HUANG Rong-hua, WANG Zhao-wen, et al(丁红元,黄荣华,王兆文,等). Journal of Huazhong University of Science and Technology · Natural Science Edition(华中科技大学学报·自然科学版), 2012, 6: 113; 2013, 1: 59.
- [4] JIN Pan-pan, ZHANG Teng, YANG Xiao-ping, et al(金盼盼,张腾,杨晓平,等). Auto Mobile Science and Technology(汽车科技), 2012, (5): 66.
- [5] LIU Yan-de, WU Ming-ming, SUN Xu-dong, et al(刘燕德,吴明明,孙旭东,等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2016, 6: 289.
- [6] HAN Zhong-zhi, WAN Jian-hua, LIU Kang-wei, et al(韩仲志,万剑华,刘康伟,等). Chinese Journal of Analysis Laboratory(分析试验室), 2015, 11: 1268.
- [7] YAO Jie, DAI Lian-kui, LIN Yi-ling(姚捷,戴连奎,林艺玲). Chinese Journal of Light Scattering(光散射学报), 2013, 25(1): 59.
- [8] OUYANG Ai-guo, LIU Jun(欧阳爱国,刘军). Journal of Southwest China Normal University · Natural Science Edition(西南师范大学学报·自然科学版), 2012, 37(9): 98.
- [9] LI Yan-ru, SUN Rui-qing, WANG Yong-miao(李雁如,孙瑞卿,王永苗). Modern Chemical Industry(现代化工), 2013, 33(2): 113.
- [10] Hutengs C, Vohland M. Remote Sensing of Environment, 2016, 178: 127.
- [11] Jog A, Carass A, Roy S, et al. Medical Image Analysis, 2017, 35: 475.

[12] Silva M P F D, Brito L R E, Honorato F A, et al. *Fuel*, 2014, 116(1): 151.

[13] Zhang L, Li G, Sun M, et al. *Infrared Physics & Technology*, 2017, 86: 116.

Qualitative Discrimination and Quantitative Determination Model Research of Methanol Gasoline and Ethanol Gasoline

HU Jun, LIU Yan-de*, HAO Yong, SUN Xu-dong, OUYANG Ai-guo

School of Vehicle and Mechatronics Engineering, East China Jiaotong University, Nanchang 330013, China

Abstract Methanol gasoline and ethanol gasoline are both clean energy sources, but the advantages and disadvantages of them are different. Among them, the content of methanol or ethanol determines the performance of gasoline. Therefore, it is of great significance to qualitatively distinguish methanol gasoline and ethanol gasoline and quantitatively determine the alcohol content in alcohol gasoline. In this paper, the types of alcohol gasoline and its content were identified and quantitatively analyzed by mid-infrared spectroscopy. Firstly, by comparing and analyzing the mid-infrared spectroscopy of methanol gasoline and ethanol gasoline, Random Forest (RF) was used to discriminate methanol gasoline and ethanol gasoline samples. After establishing the qualitative model of methanol gasoline and ethanol gasoline, the quantitative determination model of methanol gasoline and ethanol gasoline is established to accurately determine the corresponding alcohol content in gasoline. In order to reduce the spectrum drift and light scattering caused by vibration and noise of the experimental instrument during the experiment, the mid-infrared spectrum was pretreated. In the process of analysis, different pre-treatment methods are first used for correction, such as S-G convolution smoothing, Multivariate Scattering Correction (MSC), Standard Normal Variable (SNV), derivatives (1st derivative and 2nd derivative), and then, Least Square Support Vector Machine (LSSVM) models of methanol gasoline and ethanol gasoline were established respectively. It was found that the discriminant accuracy is up to 98.23% when the number of decision trees was 61. Secondly, for the LS-SVM model, the comparison of modeling results showed that for both methanol gasoline and ethanol gasoline, SNV pre-treatment had the best effect. The predictive correlation coefficient R_p of LSSVM model after the transformation of standard normal variables for methanol content determination of methanol gasoline was 0.951 9 and RMSEP was 1.766 3. In the same situation, ethanol gasoline was 0.951 5 and 1.770 3, respectively. This research can provide technical reference and theoretical basis for the qualitative discrimination and content determination of methanol gasoline and ethanol gasoline. The detection technology can provide a new method for the measurement of alcohol gasoline in the methanol gasoline industry, which has important practical significance and lays a foundation for the detection of other types of chemical products.

Keywords Mid-infrared spectroscopy; Alcohol gasoline; Least square support vector machine; Random forest

(Received Jun. 25, 2019; accepted Oct. 30, 2019)

* Corresponding author