

高斯过程回归在近红外光谱定量分析绝缘纸老化状态中的应用

李元¹, 张文博¹, 陈晓琳^{2,3}, 李含¹, 张冠军¹

1. 西安交通大学电力设备电气绝缘国家重点实验室, 陕西 西安 710049
2. 海南电网有限责任公司电力科学研究院, 海南 海口 570125
3. 海南省电网理化分析重点实验室, 海南 海口 570125

摘要 绝缘纸的老化状态决定了油浸式变压器的剩余寿命, 对绝缘纸的老化状态进行快速有效的评估具有重要意义。聚合度是表征纤维素绝缘纸老化程度最直接可靠的参量, 在实验室中通过粘度法检测获得, 但该方法需要获得设备纸样, 不但检测耗时长, 还会对变压器类设备的绝缘造成破坏。近红外光谱分析技术可以快速有效测定物质中的组分含量, 检测过程非侵入、对绝缘无损, 目前已成功应用于多个领域, 有望成为替代传统聚合度检测的新方法。然而, 现有的近红外光谱定量分析方法尚不能满足绝缘纸聚合度的预测精度需求, 建立了基于高斯过程回归(GPR)的绝缘纸老化状态定量评估方法。构建了不同老化程度绝缘纸样本—近红外光谱数据库, 使用 Savitzky-Golay 卷积平滑算法对光谱数据进行平滑处理以提高信噪比; 研究了不同核函数 GPR 模型并开展了模型预测精度分析与参数敏感性检验。结果显示, Exp 核模型泛化性能较差, Matern32 核、Matern52 核以及 RQ 核模型对参数敏感性较高、模型稳定性较差, 最终选择了 SE 核 GPR 模型作为最优模型。将 SE 核 GPR 模型与近红外分析领域常用的 PLS, SVR 与 BPNN 模型进行性能对比, 结果表明, GPR 模型对校正集以及验证集样本的预测误差均最小(RMSE 分别为 65.5 与 70.6), 且预测结果与粘度法结果相关系数最高(r 分别为 0.94 与 0.93)。与其他三种模型相比, GPR 模型的 RMSE 比 PLS, SVR 与 BPNN 模型低 54.1%, 58.8% 和 12.9%, 显示 GPR 模型在绝缘纸近红外光谱老化评估领域具有一定优势。

关键词 油纸绝缘; 近红外光谱; 老化状态; 定量分析; 高斯过程

中图分类号: O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2022)10-3073-06

引言

大型油浸式电力变压器是输变电系统的关键设备, 其健康状态直接关系到电网的本质安全^[1-2]。油纸复合绝缘是变压器的主要绝缘形式, 运行中长期承受电、热、机械、环境等多重应力作用, 导致绝缘逐渐老化甚至失效^[3-4]。绝缘油劣化后可方便地通过滤油、祛气甚至换油实现状态更新, 而绝缘纸的老化进程却不可逆转, 而且老化后难以更换, 因此绝缘纸的老化状态基本上决定了变压器的剩余寿命。聚合度是指纤维素分子链上葡萄糖单体的数量, 是表征绝缘纸老化状态最直接、可靠的特征参量^[5]。传统的聚合度检测是将纸样研磨溶解后测量溶液粘度获得聚合度, 称为粘度检测法。该方法准确度较高但检测耗时长, 还需要取得纸样, 对变压器的主绝缘具有破坏性。在电网主设备智能运检的背景下,

提出一种检测快速、无损便捷的绝缘纸聚合度评估方法具有重要意义。

近红外光谱分析技术可以对物质中的化学组分进行快速测定, 已在多个行业有效应用, 有望成为替代绝缘纸聚合度粘度法检测的新方法。光谱定量分析方法是该技术的核心, 近年来已开展了不少研究。李广茂等^[6]利用偏最小二乘法(partial least squares regression, PLS)建立了绝缘油甲醇含量与拉曼光谱之间的定量分析模型, 实现变压器油中甲醇含量检测; 王书涛^[7]利用支持向量回归机方法(support vector regression, SVR)解析柴油的近红外光谱数据, 准确预测了柴油密度、粘度与凝点; 蒋有列等^[8]通过将 PLS、反向传播神经网络(back propagation neural network, BPNN)方法应用于近红外光谱分析, 建立了绝缘油老化状态的定量分析模型, 结果表明 BPNN 模型的预测精度较高。

上述方法中, PLS 是一种线性建模方法, 在油、绝缘纸

收稿日期: 2021-06-03, 修订日期: 2022-03-15

基金项目: 国家自然科学基金项目(51607139)资助

作者简介: 李元, 1984年生, 西安交通大学电力设备电气绝缘国家重点实验室副教授 e-mail: liyuan8490@xjtu.edu.cn

及水分共存时,多种物质的光谱混叠,增加了系统非线性,导致 PLS 难以应对混合体系的定量分析。SVR 本质上是一种二分类方法,在用于绝缘纸聚合度回归分析(多分类)时效果较差。BPNN 则容易陷入局部最优,当数据库中样本数量较少时,建立的神经网络模型很容易出现过拟合问题。

在对比 PLS, SVR 以及 BPNN 定量分析模型的基础上,提出一种高斯过程回归(Gaussian process regression, GPR)的绝缘纸聚合度预测方法,建立不同老化状态纸样的近红外光谱与其聚合度之间的准确关联。通过模型精度分析与参数敏感性校验,从不同核函数 GPR 模型中确定了最优模型;进一步地,将建立得到的 GPR 模型与通用的 PLS, SVR 与 BPNN 模型进行预测性能对比。研究结果为绝缘纸老化状态的近红外光谱检测技术提供理论基础与模型参考。

1 实验部分

1.1 样本制备

通过加速热老化实验制备获得绝缘纸样本。加速热老化实验平台由真空干燥箱、干燥氮气以及老化皿等组成。在分别对绝缘纸与绝缘油进行干燥、除气处理后,通过真空浸渍获得油纸绝缘样品。在 130 °C 条件下进行油纸绝缘样品的加速热老化实验,定期取样获得 478 份不同老化程度的绝缘纸样本。

1.2 纸样近红外光谱采集

对制备得到的纸样采集漫反射近红外光谱,如图 1 所示。入射光在纸样和积分球表面发生多次漫反射,反射光通过测光窗、光纤传递至光谱仪进行分光,与参比白板的标准光谱比对后得到纸样的近红外光谱。光谱采集系统由 Sol 2.2A 便携式近红外光谱仪(美国 B&W Tek 公司)以及 BWSpec™ 光谱采集软件构成。其中,光谱仪检测单元为 256 单元铟镓砷阵列,通过半导体制冷;探头积分球与参比白板材质为聚四氟乙烯(PTFE)。考虑到光谱仪自身暗噪声会影响光谱质量,在光谱采集前需要记录暗电流水平,设置光谱积分时间为 600 μ s,扫描次数为 32 次。

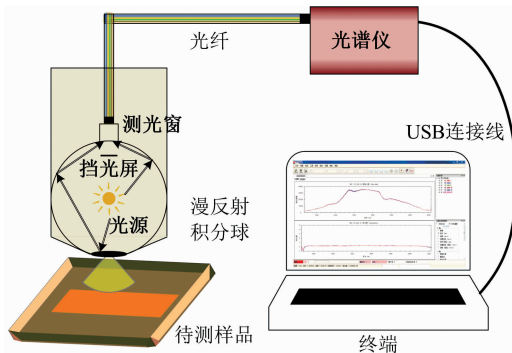


图 1 绝缘纸样近红外光谱采集

Fig. 1 NIRS acquisition of insulating paper sample

采集得到的典型绝缘纸样本的近红外光谱如图 2 所示,可以发现样本光谱在 1 220, 1 340, 1 540 以及 1 703 nm 处具有明显的特征峰。其中位于 1 220 nm 的 C—H 二级倍频、

1 703 nm 的 C—H 一级倍频归属于纤维素^[4],当纤维素链受热断裂时,上述特征峰会发生明显变化,这是利用近红外光谱数据分析绝缘纸老化状态的理论基础。但是,位于 1 340 nm 处的 C—H 弱合频、1 540 nm 处的 O—H 一级倍频可同时归属于纤维素、绝缘油与水分子,组分信息混叠,导致光谱特征峰与绝缘纸老化程度的关系并非肉眼可辨;同时,样品中混入的无关组分会影响预测结果,因此需要借助光谱定量分析方法建立绝缘纸近红外光谱与老化程度之间的准确关联。

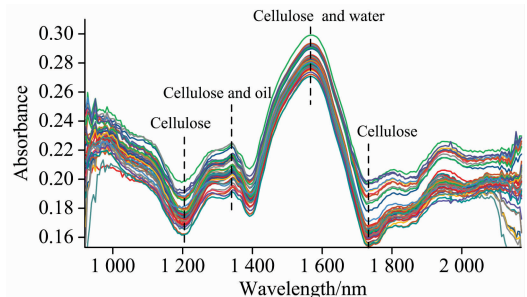


图 2 典型绝缘纸样本近红外光谱

Fig. 2 Spectra of typical insulating paper samples

1.3 样本聚合度检测

对制备得到的不同老化程度绝缘纸样进行聚合度检测,检测方法为粘度滴定法^[9]。对每个样本进行两次聚合度检测,取其平均值作为最终检测结果。聚合度检测值作为标准值,用于模型训练与性能评估。

1.4 方法

1.4.1 高斯过程回归

本质上,GPR 算法是在给定样本光谱数据分布的前提下得到对应样本聚合度值分布,所得分布函数的数学期望即为 GPR 模型的聚合度预测结果。GPR 中假定绝缘纸聚合度 y 是由高斯分布函数 $f(x)$ 以及噪声 ϵ 组成^[10],同时有式(1)成立

$$y = f(x) + \epsilon \quad (1)$$

式(1)中,高斯分布函数 $f(x)$ 由其期望 $m(x)$ 以及方差 $k(x, x')$ 决定,即

$$f(x) \sim GP(m(x), k(x, x')) \quad (2)$$

式(2)中, $k(x, x')$ 的具体形式为核函数。

高斯过程同样假设噪声 ϵ 符合高斯分布,即

$$\epsilon \sim N(0, \sigma_n^2) \quad (3)$$

式(3)中, σ_n^2 为噪声分布函数的方差。

由于任何有限的高斯序列可以组成一个独立的高斯过程^[11],因此 y 的分布同样符合高斯过程,且该分布函数可通过 $f(x)$ 以及 ϵ 求解得到

$$y \sim GP(m(x), k(x, x') + \sigma_n^2 \delta(x, x')) \quad (4)$$

式(4)中, $\delta(x, x')$ 为克罗内克函数^[12],当且仅当 $x = x'$ 时 $\delta(x, x') = 1$,否则 $\delta(x, x') = 0$ 。

当输入待预测纸样的近红外光谱数据 x 时,基于式(4)可以计算得到训练集绝缘纸样本聚合度与待测绝缘纸样本聚合度的联合分布,如式(5)所示

$$\begin{bmatrix} y \\ y_* \end{bmatrix} \sim N\left(0, \begin{bmatrix} \mathbf{K} & \mathbf{K}_*^T \\ \mathbf{K}_* & \mathbf{K}_{**} \end{bmatrix}\right) \quad (5)$$

式(5)中, \mathbf{K} 为方差矩阵, 通过式(6)求解

$$\mathbf{K} = \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & \cdots & k(x_1, x_n) \\ k(x_2, x_1) & k(x_2, x_2) & \cdots & k(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_n, x_1) & k(x_n, x_2) & \cdots & k(x_n, x_n) \end{bmatrix} \quad (6)$$

利用式(5)得到的联合分布可得边缘密度分布函数 y_* , 即待测绝缘纸样本聚合度的分布函数

$$y_* | y \sim N(\mathbf{K}_* \mathbf{K}^{-1} y, \mathbf{K}_{**} - \mathbf{K}_* \mathbf{K}^{-1} \mathbf{K}_*^T) \quad (7)$$

求解得到的待测绝缘纸样本聚合度分布函数的数学期望即样本聚合度的预测结果。

1.4.2 模型建立过程

利用 K-S(Kennard Stone)方法^[13]将样本的光谱数据按 8 : 2 分为训练集与验证集, 其中训练集样本 386 个、验证集 92 个。在模型训练前利用 SG(Savitzky-Golay)平滑方法^[14]对光谱数据进行预处理, 降低光谱数据的噪声水平, 提高数据质量。

需要指出, 式(2)中核函数 $k(x, x')$ 的类型对模型的预测能力有显著影响。建立了基于不同核函数的 GPR 模型, 选用的核函数包括平方指数核(squared exponential, SE)、指数核(exponential, Exp)、有理二次核(rational quadratic, RQ)、Matern32 与 Matern52 核。通过模型预测精度分析和参数敏感性检验, 确定最优核函数与 GPR 聚合度预测模型。将得到的 GPR 模型与通用 PLS, SVR 以及 BPNN 模型进行聚合度

预测性能横向比较。GPR 建模与性能比较流程如图 3 所示。

2 结果与讨论

2.1 不同核函数 GPR 模型性能比较

依据图 3 建立的基于不同核函数的 GPR 模型进行预测性能对比, 结果如图 4 所示, 图中纵轴表示样本均方根误差 (root mean square error, RMSE)。其中 Exp 核模型对训练集数据的 RMSE 远高于验证集, 出现严重的过拟合现象, 因此先将该核函数剔除。而其余核函数 GPR 模型的预测准确度差异尚不显著, 需要通过参数敏感性检验进一步筛选最优核函数。

通过多次试验发现, 提高式(4)中噪声分布函数方差的下限 σ_{n_low} 会显著降低模型预测准确度。参数敏感性校验是通过改变 σ_{n_low} 来检查模型 RMSE 的相对变化率 Δ , 计算方法如式(8)所示。 Δ 越小表示模型对该参数的敏感性越低, 即在不同参数条件下模型的性能可以保持稳定。图 4 所示结果是在 $\sigma_{n_low} = 60$ 的情况下得到的, 现给定参数扰动 δ 。根据 Δ 表现判定各核函数 GPR 模型的稳定性, 检验结果如图 5 所示。由

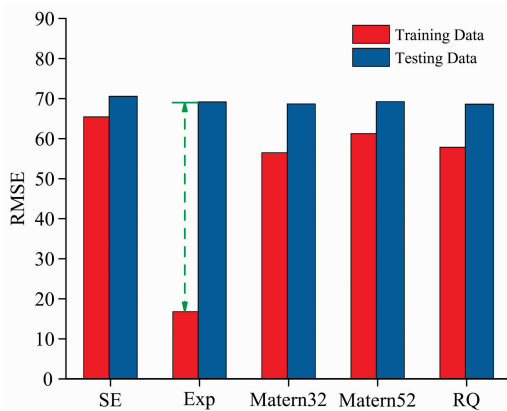


图 4 不同核函数 GPR 模型性能比较

Fig. 4 Comparison of GPR model performances with different kernels

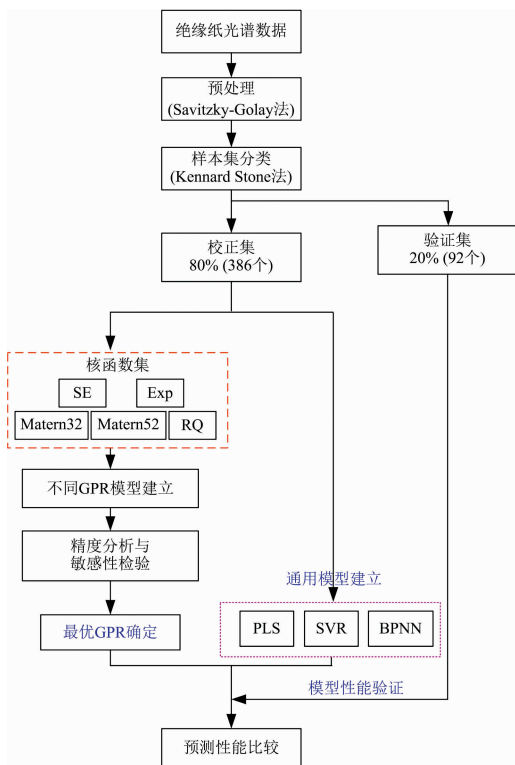


图 3 GPR 建模与不同模型性能比较流程

Fig. 3 Flowchat of GPR modelling and performance comparison among different models

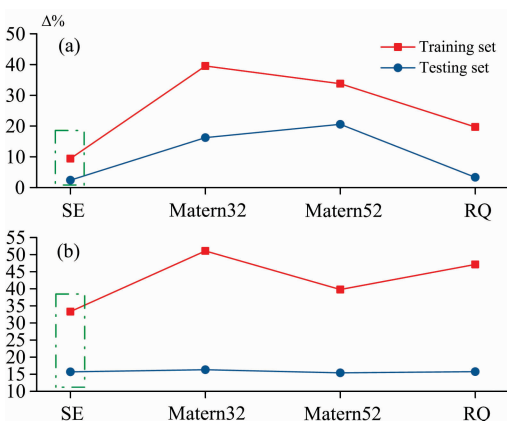


图 5 模型参数敏感性检验结果

Fig. 5 Results of model parameter sensitivity tests

图可知, $\delta=30$ 或 60 时, 四种核函数模型的 RMSE 均有不同程度的增加, 其中 SE 内核模型的 Δ 最小, SE 内核回归模型相较于其他模型具有更高稳定性, 因此将 SE 核函数作为 GPR 模型的最优核函数。

$$\Delta = \frac{d(\text{RMSE})}{d(\sigma_{n_{\text{low}}})} = \frac{\text{RMSE}(\sigma_{n_{\text{low}}} + \delta) - \text{RMSE}(\sigma_{n_{\text{low}}})}{\text{RMSE}(\sigma_{n_{\text{low}}})} \quad (8)$$

式(8)中, δ 为参数扰动。

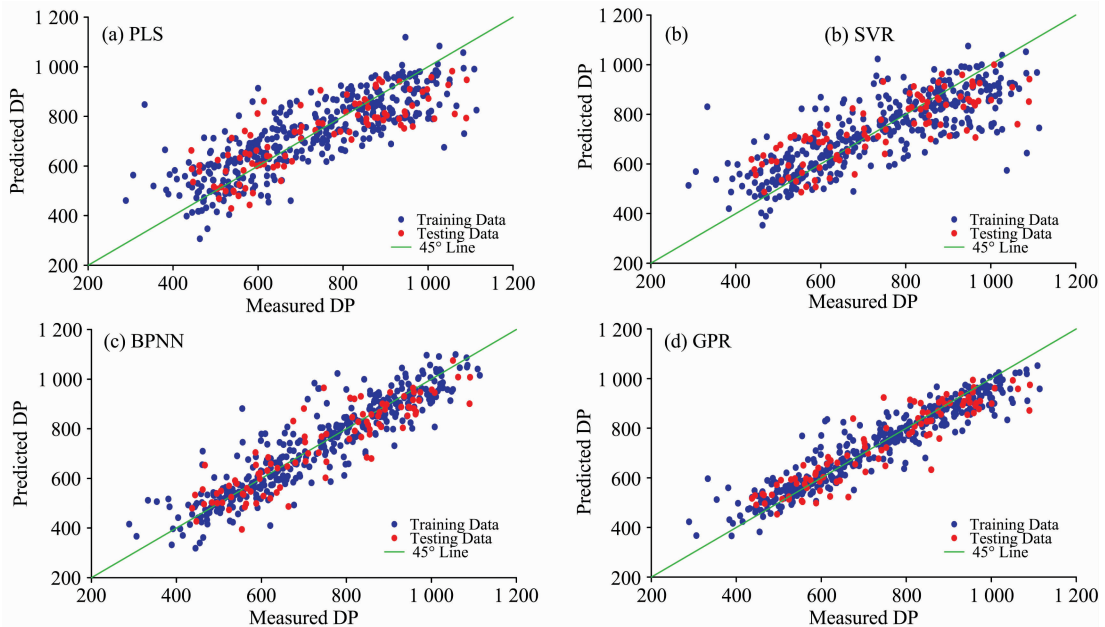


图 6 不同定量分析模型聚合度预测结果

Fig. 6 DP prediction results of different quantitative analysis models

利用 RMSE 与相关系数 r 定量表征四种模型的预测效果, 结果如表 1 所示。其中 SVR 模型性能最差, 该模型对校正集样本以及验证集样本的预测 RMSE 分别为 110.6 与 105.5。GPR 模型的效果最优, 对校正集以及验证集样本的预测结果均具有最小的 RMSE(65.5 与 70.6)。四种模型的 RMSE 与 r 曲线如图 7 所示, 其中 RMSE_c 与 RMSE_v 分别表示校正集与验证集的均方根误差, r_c 与 r_v 表示校正集与验证集的相关系数, 图中表明四种模型对绝缘纸聚合度的预测准确度排序为: $\text{GPR} > \text{BPNN} > \text{PLS} > \text{SVR}$ 。

表 1 不同绝缘纸聚合度预测模型性能比较

Table 1 Comparison of different DP prediction models for insulating paper

建模方法	校正集		验证集			
	样本数量	RMSE	r	样本数量	RMSE	r
PLS ^[6]	386	105.0	0.82	92	104.7	0.82
SVR ^[7]		110.6	0.80		105.5	0.81
BPNN ^[8]		75.3	0.92		78.3	0.90
GPR		65.5	0.94		70.6	0.93

2.2 多模型聚合度预测性能对比

将建立的最优 GPR 模型与通用 PLS, SVR 和 BPNN 三种聚合度预测模型的性能比较如图 6 所示。较为明显地, BPNN 模型[图 6(c)]与 GPR 模型[图 6(d)]的预测效果较好, 而 PLS 模型[图 6(a)]与 SVR 模型[图 6(b)]预测效果则稍差。

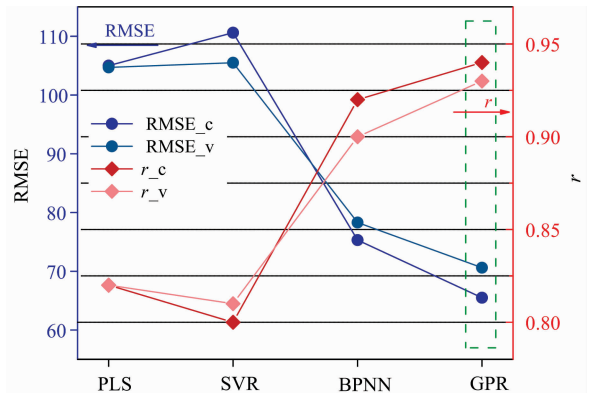


图 7 四种定量分析模型聚合度预测结果

Fig. 7 DP prediction results of four quantitative models

3 结论

通过加速热老化实验制备获得不同老化程度的绝缘纸样品, 建立了绝缘纸的近红外光谱数据库, 提出了绝缘纸聚合度的 GPR 预测模型。主要研究结论如下:

(1) 考虑到核函数的类型对 GPR 模型的预测准确度影响较大, 建立了基于多种核函数的 GPR 模型, 对模型进行参数敏感性检验, 研究结果发现 SE 核 GPR 模型不但能克服过拟

合问题, 而且模型的参数敏感性低, 稳定性最好, 由此筛选出基于 SE 核的最优 GPR 聚合度预测模型。

(2) 与近红外光谱领域常用预测模型横向比较显示, 研究提出的 GPR 聚合度预测模型准确度最高, 其预测误差分别低于 PLS, SVR 与 BPNN 模型误差 54.1%, 58.8% 和 12.9%。研究结果证实了利用 GPR 方法准确预测绝缘纸聚

合度的可行性。

(3) 基于高斯过程回归的绝缘纸聚合度近红外光谱评估模型检测精度较高, 相比于传统检测方法(粘度法)检测方便快捷, 具有更加广泛的应用场景, 能够更好的服务于变压器绝缘老化诊断与运行评价。

References

- [1] ZOU Jing-xin, CHEN Wei-gen, WAN Fu, et al(邹经鑫, 陈伟根, 万 福, 等). Transactions of China Electrotechnical Society(电工技术学报), 2018, 33(5): 1133.
- [2] Wang M, Vandermaar A J, Strivastave K D. IEEE Electrical Insulation Magazine, 2002, 18(6): 12.
- [3] Saha T K, Purkait P. IEEE Transactions on Power Delivery, 2008, 23(1): 10.
- [4] Baird P J, Herman H, Stevens G C, et al. IEEE Transactions on Dielectrics & Electrical Insulation, 2006, 13(02): 309.
- [5] YANG Ding-kun, CHEN Wei-gen, WAN Fu, et al(杨定坤, 陈伟根, 万 福, 等). Proceedings of the CSEE(中国电机工程学报), 2021, 41(13): 4710.
- [6] LI Guang-mao, QIAO Sheng-ya, ZHU Chen, et al(李光茂, 乔胜亚, 朱 晨, 等). High Voltage Engineering(高电压技术), 2021, 47(6): 2007.
- [7] Wang S, Liu S, Yuan Y, et al. Infrared Physics & Technology, 2020, 106: 103276.
- [8] JIANG You-lie, ZHU Shi-ping, TANG Chao, et al(蒋友列, 祝诗平, 唐 超, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2020, 40(11): 3515.
- [9] ZHOU Li-jun, LI Xian-lang, DUAN Zong-chao, et al(周利军, 李先浪, 段宗超, 等). Proceedings of the CSEE(中国电机工程学报), 2014, 34(21): 3514.
- [10] Schulz E, Speekenbrink M, Krause A, et al. Journal of Mathematical Psychology, 2018, 85: 1.
- [11] Kong D, Chen Y, Li N, et al. Mechanical Systems and Signal Processing, 2018, 104: 556.
- [12] Wang B and Chen T. Chemometrics and Intelligent Laboratory Systems, 2015, 142: 159.
- [13] Tao D, Wang Z, Li G, et al. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2019, 208: 7.
- [14] Liu Y I, Sun L, Ran Z, et al. Journal of Food Protection, 2019, 82(10): 1655.

Application of Gaussian Process Regression on the Quantitative Analysis of the Aging Condition of Insulating Paper by Near-Infrared Spectroscopy

LI Yuan¹, ZHANG Wen-bo¹, CHEN Xiao-lin^{2,3}, LI Han¹, ZHANG Guan-jun¹

1. State Key Laboratory of Electrical Insulation and Power Equipment, Xi'an Jiaotong University, Xi'an 710049, China

2. Electric Power Research Institute of Hainan Power Grid Co., Ltd., Haikou 570125, China

3. Key Laboratory of Physical and Chemical Analysis for Electric Power of Hainan Province, Haikou 570125, China

Abstract As the aging condition of the insulating papers determines the remaining lifetime of the oil-immersed transformers, a fast and effective aging assessment method for insulating paper is of great significance. As it is known, the degree of polymerization (DP) is the most direct parameter to characterize the aging condition of insulating papers. However, the traditional detection method or so-called viscometry is time-consuming and destructive. Near-infrared spectroscopy (NIRS) technology, as a non-destructive detection method can rapidly determine the samples' components and contents. Until now, it has been successfully applied in many fields and will hopefully be employed as an alternative method to viscometry. However, the current spectral quantitative analysis method is still not accurate enough to predict the DP of insulating paper samples. In this paper, we introduce Gaussian process regression (GPR) to predict DP of insulating papers accurately. Firstly, the NIRS database of insulating papers under different aging conditions is established, and in this procedure, the raw spectra are preprocessed by the Savitzky-Golay method to improve the signal ratio to noise. Then GPR models with various kernels are established, and the prediction accuracy and stability of the different models are comparatively studied. The results show that the GPR model with Exp kernel is of poor generalization performance, and the models with Matern32, Matern52 and RQ kernels are highly sensitive to the model parameters. Finally, the SE kernel is selected as the optimal kernel function of the GPR model. The

DP prediction results of the SE kernel GPR model are compared with traditional PLS, SVR and BPNN models, and the results show that our established GPR model has the lowest RMSE (65.5 and 70.6) and highest correlation coefficient r (0.94 and 0.93), both for the training set and testing set. The RMSE of the GPR model is lower than PLS, SVR and BPNN models by 54.1%, 58.8% and 12.9% respectively. It is indicated that the established GPR model can be a powerful tool for the aging assessment of insulating papers by the NIRS technique.

Keywords Oil-paper insulation; Near-infrared spectroscopy; Aging condition; Quantitative analysis; Gaussian process

(Received Jun. 3, 2021; accepted Mar. 15, 2022)