

基于PCA和GA算法优化最小二乘支持向量机的开关柜温度预测

张好勇, 张东亮, 赵 雨, 刘景宇, 王 正

(国网廊坊供电公司, 河北 廊坊 065000)

摘要: 高压开关柜在变、配电系统中承担着保护、控制和分配电能的作用, 其运行稳定性对电网安全稳定运行至关重要。针对影响开关柜温度因素众多以及最小二乘支持向量机 (LS-SVM) 回归预测时最优参数的选择问题, 提出基于主成分分析 (PCA) 和遗传算法 (GA) 优化的最小二乘支持向量机开关柜温度预测模型。主成分分析通过变量间相关性降低影响开关柜温度因素变量维数, 加快模型训练速度; 同时, 利用遗传算法进行LS-SVM模型参数优化, 提高开关柜温度预测准确度, 并通过实例分析进行验证。

关键词: 开关柜温度; 主成分分析; 遗传算法; 最小二乘支持向量机; 预测准确度

Switchgear temperature prediction based on least square support vector machine optimized by PCA and GA algorithms

ZHANG Haoyong, ZHANG Dongliang, ZHAO Yu, LIU Jingyu, WANG Zheng

(Langfang Grid Co., Langfang 065000, China)

Abstract: The high-voltage switchgear plays the role of protecting, controlling and distributing electrical energy in the power distribution system, and its operational stability is essential for the safe and stable operation of the power grid. Aiming at the problem of multiple influencing factor of switchgear temperature and the optimal parameter selection of least square support vector machine, a least squares vector machine switchgear temperature prediction model based on principal component analysis and genetic algorithm optimization is proposed. Principal component analysis can speeds up model training by reducing the variable dimension of influencing factor of switchgear temperature; At the same time, the parameters of LS-SVM model are optimized by genetic algorithm, in oder to improve the accuracy of switchgear temperature prediction and verified by case analysis.

Key words: switchgear temperature; principal component analysis; genetic algorithm; least square support vector machine; prediction accuracy

0 引言

开关柜温度是表征开关柜运行状态的重要信息, 也是开展状态检修的重要依据。快速准

确的开关柜温度预测能够预判设备故障, 并据此合理安排检修, 避免定期检修存在的盲目性, 以提高电力系统运行稳定性。

影响开关柜温度的因素包括季节、负荷、天气、环境温度和日期类型等十余种，且不同因素之间存在一定的相关性，要想快速、准确实现开关柜温度预测，首先要降低不同因素间的相关性，确定影响开关柜温度的主要因素。本文利用主成分分析方法^[1]将影响开关柜温度的因素变量进行降维，以确定其主要因素，从而提高模型训练速度，且能够有效避免依靠经验选择方法存在的主观性问题。

目前，人工智能算法在回归预测方面得到了广泛研究，其中较为普遍的是神经网络和支持向量机^[2-4]。神经网络算法具有良好的自学习能力和强大的非线性逼近能力，能够大规模并行处理数据，但是神经网络算法存在过拟合学习、局部极小值等问题，且对于“维数灾难”没有很好的解决办法。而支持向量机算法是一种新兴的机器学习方法，它对于样本空间的高度非线性问题具有强大的处理能力，且能够解决神经网络、网格搜索等智能算法存在的过学习、高维数和局部极小值等问题。因此，本文通过支持向量机的改进算法LS-SVM进行开关柜温度预测。但是，由于LS-SVM模型参数（惩罚因子 γ 和核参数 σ ）^[5-6]直接影响模型预测结果，如依靠经验选取，存在很大的主观性和随意性。目前对于参数选择，使用较多的是利用智能算法寻优，如网格搜索、粒子群和遗传算法等^[7-8]；其中网格搜索算法用于取值范围大、参数数量多的模型时，存在搜索时间长、准确度不高等问题；而遗传算法采用整体搜索和优化搜索策略，具有强大的全局搜索能力、并行处理数据和高效性等优点，因此本文采用遗传算法对LS-SVM模型进行参数寻优^[9-12]，以提高开关柜预测准确度。

1 基本原理

1.1 主成分分析

主成分分析（Principal Component Analysis, PCA）是一种考察多个变量相关性

的统计方法，即通过几个主成分来表征原始数据空间的内部结构，且能够尽可能多地保留原始数据的信息，从而降低数据空间的维数，提高算法计算速度。

设定 F_1 为原始数据空间的第一个主成分， F_1 的选取为原始数据中方差 $\text{Var}(F_1)$ 最大。若 F_1 不足以代表原始数据空间信息，则选取第二个主成分 F_2 ，且 F_1 、 F_2 的协方差满足 $\text{Cov}(F_1, F_2)$ 。以此类推，构造原始数据空间第 k 个主成分，计算公式为

$$F_k = a_{1i}Z_{X1} + a_{2i}Z_{X2} + \cdots + a_{ki}Z_{Xk} \quad (1)$$

式中， $a_{1i}, a_{2i}, \cdots, a_{ki} (i=1, 2, \cdots, m)$ 为 X 的协方差阵 Σ 的特征值所对应的特征向量； $Z_{X1}, Z_{X2}, \cdots, Z_{Xk}$ 为对原始数据进行去量纲处理后的值。

1.2 遗传算法

遗传算法（Genetic Algorithm, GA）是由美国教授Holland提出来的，它是一种模拟生物进化过程中自然选择和遗传学机理的人工智能算法。GA是从串集开始问题解的搜索，覆盖面大，能够并行处理种群中的多个个体，利于问题解的全局寻优。

算法流程如图1所示。

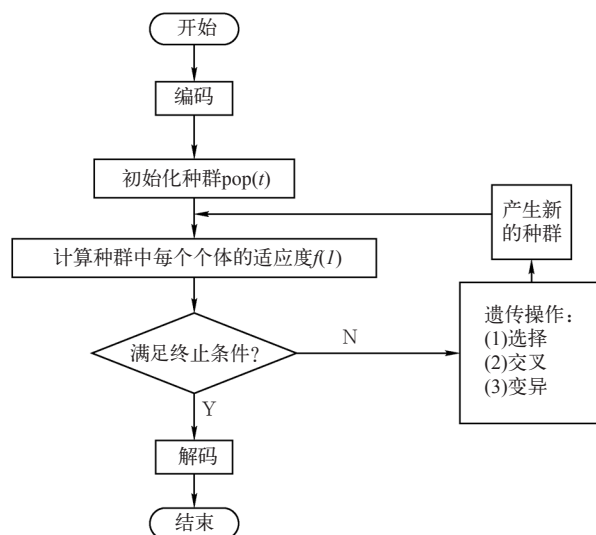


图1 遗传算法流程

Fig. 1 Genetic algorithm process

1) 编码及初始化种群。采用二进制编码或实数编码方式对样本数据进行编码，并随机生

成种群 $\text{pop}(t)$ 。

2) 计算种群 $\text{pop}(t)$ 中每个染色体 $\text{pop}_i(t)$ 的适应度 f_i 。

3) 选择操作。即通过优胜劣汰的方法将优质的个体遗传到下一代。常用的选择算子是轮盘赌选择法 (Roulette Wheel Selection, RWS)，该方法将计算选择概率 $p_i = f_i / \sum_{i=1}^N f_i$ 较大的个体进行遗传，构成新的种群。

4) 交叉操作。将染色体以设置的概率 p_c 进行交配，得到新的种群 $\text{crosspop}(t)$ 。

5) 变异操作。将染色体以设置好的概率 p_m 发生基因变异，形成新的种群 $\text{mutpop}(t)$ 。

6) 若迭代结果达到给定的个体适应度阈值或者达到预设迭代次数，则寻得种群中最优染色体，流程结束；否则返回步骤2) 重新执行。

1.3 最小二乘支持向量机

最小二乘支持向量机 (Least Squares Support Vector Machine, LS-SVM) 是支持向量机 (Support Vector Machine, SVM) 的一种改进算法，它是将低维空间的非线性样本数据映射到高维特征空间 (Hilbet空间)，通过构造最优决策面，将其转化为可线性处理问题，以提高模型运算速度及预测准确度。

设样本数据集 $T = [x_i, y_i]_{i=1}^N$ ， N 为样本总数；则样本的预测模型为

$$y(x) = w^T \varphi(x) + b \quad (2)$$

式中， $\varphi(x)$ 为从样本空间到 Hilbet 空间的非线性映射； w 和 b 为模型参数。

根据机构风险最小化原则，LS-SVM 模型的优化问题为

$$\begin{aligned} \min J_{\text{LS}}(w, \xi) &= \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^n \xi_i^2 \\ \text{s. t. } w^T \varphi(x_i) + b + \xi_i &= y_i \quad i = 1, 2, \dots, N \end{aligned} \quad (3)$$

式中， γ 为惩罚因子。

定义拉格朗日函数

$$L(w, b, \xi_i, a_i) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^n \xi_i^2 - \sum_{i=1}^n a_i [w^T \varphi(x_i) + b + \xi_i - y_i] \quad (4)$$

式中， a_i 为拉格朗日乘子。

根据 Karush-Kuhn-Tucker 条件，分别对 w 、 b 、 ξ 和 a_i 求导，消去 w 和 ξ_i ，得到

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 & E^T \\ E & \Omega + \gamma^{-1} E \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ y \end{bmatrix} \quad (5)$$

式中， $y = [y_1, y_2, \dots, y_n]^T$ ； $a = [a_1, a_2, \dots, a_n]^T$ ； $E = [1, 1, \dots, 1]^T$ 。

根据 Mercer 定理， Ω 矩阵的元素可由式 (6) 表示为

$$\Omega_{i,j} = \varphi^T(x_i) \varphi(x_j) = k(x_i, x_j) \quad (6)$$

式中， $k(x_i, x_j)$ 为核函数。

由于验证构造核函数对输入空间内的任意格拉姆矩阵比较困难，因此本文选择使用径向基核函数，即

$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (7)$$

式中， σ 为核参数宽度。

综上，LS-SVM 回归预测模型计算公式为

$$f(x) = \sum_{i=1}^n a_i K(x_i, x_j) + b \quad (8)$$

2 PCA和GA算法优化LS-SVM预测模型

根据前文所述，PCA 和 GA 算法优化 LS-SVM 预测模型 (PCA-GA-LS-SVM) 计算流程如图 2 所示。

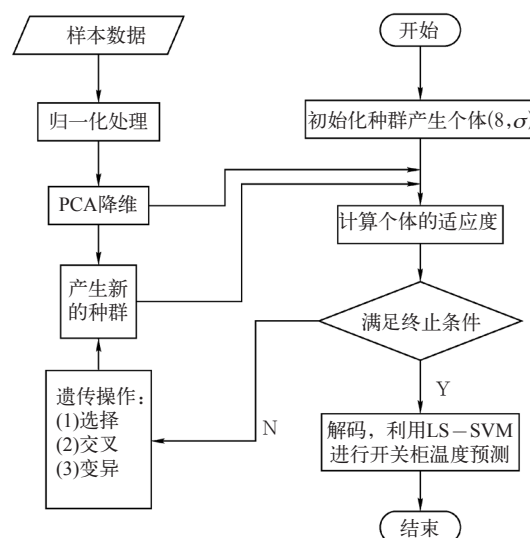


图2 PCA-GA-LS-SVM 计算流程

Fig. 2 PCA-GA-LS-SVM calculation process

1) 初始化设置。如种群大小、迭代次数和交叉概率等,其中染色体由 (γ, σ) 组成。

2) 利用LS-SVM对经PCA降维处理的样本数据进行训练,并计算个体适应度。为保证训练准确度,本文将训练的均方差(MSE)作为目标函数值,即个体适应度。

3) 判断是否满足终止条件。若满足,则解码得到最优参数,并进行开关柜温度预测;若不满足,则进行遗传操作,返回步骤2)重新执行。

3 实例分析

3.1 数据采集及模型初始化

本文以某110 kV变电站1号主变压器10 kV侧受总201开关柜对象,每10 min对实测温度和环境温度、负荷电流、季节、天气和日期类型等数据进行采集(每24 h共144个采样点),时间范围为2019年1月1日~2019年7月31日,将以上采集数据作为样本数据,利用PCA-GA-LS-SVM模型进行训练,并对2019年8月1日该开关柜温度进行预测。

PCA-GA-LS-SVM模型参数初始化:设定种群群体大小popsize为20,迭代次数=200,交叉概率 $p_c=0.6$,变异概率 $p_m=0.005$,惩罚因子 $\gamma \in [0.1, 100]$,核参数 $\sigma \in [0.01, 1\ 000]$,遗传代沟 $p_g=0.5$ 。

3.2 预测结果分析

PCA和GA算法优化的LS-SVM开关柜温度预测模型经过115次迭代终止,最优参数 $\gamma=98.686\ 5$, $\sigma=0.126\ 0$,2019年8月1日开关柜温度预测结果如图3所示。

从均方差(MSE)、训练时长两方面对LS-SVM、GA-LS-SVM、PCA-LS-SVM和PCA-GA-LS-SVM四种模型进行对比,结果如下表所示。

由下表可以看出,PCA-GA-LS-SVM模型均方差仅为1.28%,预测准确度最高;训练时长方面,与PCA-GA-LS-SVM模型相比,虽然PCA-LS-SVM模型快了3 s,但

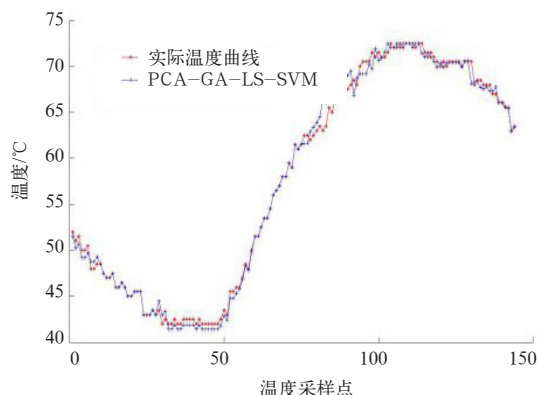


图3 2019年8月1日开关柜温度预测曲线

Fig. 3 Temperature prediction curve of switchgear on August 1, 2019

表 四种模型预测结果对比

Tab. Comparison of prediction results of four models

预测模型	均方差/%	训练时长/s
LS-SVM	3.95	195.69
GA-LS-SVM	1.74	89.65
PCA-LS-SVM	3.89	42.90
PCA-GA-LS-SVM	1.28	45.97

是其均方差是PCA-GA-LS-SVM模型的3倍,预测准确度较差。因此,基于PCA和GA算法优化的LS-SVM模型有效提高了预测准确度和训练时长。

4 结束语

由于最小二乘支持向量机能够很好地处理高维数、非线性数据,因此利用最小二乘支持向量机进行开关柜温度预测具有可行性。考虑到最小二乘支持向量机的预测准确度对模型参数有很大依赖性,本文提出基于主成分分析和遗传算法优化的最小二乘支持向量机开关柜温度预测模型,并进行实例验证,得到以下结论:

1) 利用主成分分析方法对样本数据进行降维处理,大大提升了预测模型的运算速度。

2) 提出的遗传算法对最小二乘支持向量机预测模型进行参数优化,有效提高了开关柜温

度预测准确度。

参考文献

- [1] 孙平安, 王备战. 机器学习中的PCA降维方法研究及其应用[J]. 湖南工业大学学报, 2019, 33(1): 80-85.
- [2] 张健美, 周步祥, 林楠, 等. 灰色Elman神经网络的电网中长期负荷预测[J]. 电力系统及其自动化学报, 2013, 25(4): 145-149.
- [3] 拜润卿, 何新, 赵红, 等. 基于粒子群优化方法的短期电力负荷预测[J]. 电子测试, 2015(4): 46-48.
- [4] 马晓博. 基于小波变换和BP神经网络的短期风电功率预测[J]. 电力科学与技术学报, 2015, 30(2): 92-97.
- [5] 姚晔. 最小二乘支持向量机的参数优化及其应用研究[J]. 科技经济导刊, 2016(29): 70-71.
- [6] 郭辉, 刘贺平, 王玲. 最小二乘支持向量机参数选择方法及其应用研究[J]. 系统仿真学报, 2006, 18(7): 2033-2036, 2051.
- [7] 顾凯成. 改进网格搜索的支持向量机参数优化研究及应用[D]. 兰州: 兰州理工大学, 2016.
- [8] 徐龙秀, 辛超山, 牛东晓, 等. 基于自适应粒子群参数优化的最小二乘支持向量机用电量预测模型[J]. 科学技术与工程, 2019, 19(6): 136-141.
- [9] 王克奇, 杨少春, 戴天虹, 等. 采用遗传算法优化最小二乘支持向量机参数的方法[J]. 计算机应用与软件, 2009, 26(7): 109-111.
- [10] 万源, 童恒庆, 朱映映. 基于遗传算法的多核支持向量机的参数优化[J]. 武汉大学学报(理学

报), 2012, 58(3): 255-259.

- [11] 张伟. 基于GA优化的SVM算法的股票趋势预测[D]. 长春: 吉林大学, 2010.

- [12] 郎凡. 基于遗传算法的改进支持向量机财务预警研究[D]. 北京: 北京交通大学, 2008.

作者简介



张好勇

张好勇(1986-), 男, 工程师, 硕士研究生, 主要从事高压断路器、隔离开关等电力设备评估、验收工作。

张东亮(1976-), 男, 工程师, 主要从事高压断路器、隔离开关等电力设备调试和检修工作。

赵雨(1982-), 男, 工程师, 主要从事高压断路器、隔离开关等电力设备调试和检修工作。

刘景宇(1984-), 男, 工程师, 主要从事高压断路器、隔离开关等电力设备调试和检修工作。

王正(1993-), 男, 研究生, 助理工程师, 主要从事高压断路器、隔离开关等电力设备调试和检修工作。

(收稿日期: 2019-10-19) **EA**