

基于时序分析的阀冷进阀温度预测方法^①



张朝辉¹, 梁家豪¹, 梁秉岗¹, 秦冠军², 尹毅然², 丁笠³, 周宇³

¹(中国南方电网有限责任公司超高压输电公司广州局, 广州 510663)

²(南京南瑞继保电气有限公司, 南京 211106)

³(南京航空航天大学, 南京 210016)

通讯作者: 周宇, E-mail: zhouyu@nuaa.edu.cn

摘要: 电网在运行过程中, 换流阀等关键设备会不断产生热量, 当设备的热量不断聚集温度不断上升, 会影响设备的稳定性和安全性, 保证换流阀等关键设备稳定运行就显得至关重要. 阀冷系统作为冷却系统的关键设备, 以热导性较高的水为介质, 将设备的热能带出, 达到降低设备温度的目的. 可以通过监控冷却水的温度、压力技术指标来确保换流阀安全、稳定运行. 选取阀冷系统中的进阀温度为主要预测指标, 对系统的历史数据进行充分的挖掘和分析, 达到对电网运行状态预估的目的. 将传统时序模型与机器学习结合提出 ARIMA-SVM 的混合模型, 并与传统的 ARIMA 模型、SVM 模型和 GRU 神经网络模型对中国南方电网的真实阀冷数据进行时序分析预测并进行对比实验. 实验结果表明, ARIMA 模型、SVM 模型、GRU 神经网络模型和 ARIMA-SVM 混合模型都可以较好地预测进阀温度的变化趋势, 但 ARIMA-SVM 混合模型在均方根误差、均方误差和平均绝对误差 3 个评价指标上表现均更优于其他 3 个模型, 能够进一步提升进阀温度预测的精度

关键词: 阀冷; 进阀温度; 时序分析; ARIMA-SVM 混合模型

引用格式: 张朝辉, 梁家豪, 梁秉岗, 秦冠军, 尹毅然, 丁笠, 周宇. 基于时序分析的阀冷进阀温度预测方法. 计算机系统应用, 2021, 30(4):9-16. <http://www.c-s-a.org.cn/1003-3254/7874.html>

Prediction of Inlet Valve Temperature Based on Time Series Analysis

ZHANG Chao-Hui¹, LIANG Jia-Hao¹, LIANG Bing-Gang¹, QIN Guan-Jun², YIN Yi-Ran², DING Li³, ZHOU Yu³

¹(Guangzhou Bureau, CSG EHV Power Transmission Company, Guangzhou 510663, China)

²(NR Electric Co. Ltd., Nanjing 211106, China)

³(Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract: During the operation of the power grid, key devices such as converter valves continue to generate heat, affecting the stability and safety of the system. Then it is crucial to ensure the stable operation of those devices. As a major component in the cooling system, the valve cooling system releases the heat energy from the equipment with water of high thermal conductivity as the medium. The stable operation of the converter valve can be ensured by monitoring the temperature and pressure of the cooling water. Also, with inlet valve temperature in the valve cooling system as the main predictive index, the historical data of the system is fully mined for predicting the operating state of the power grid. An ARIMA-SVM hybrid model integrating the traditional time series model and machine learning is compared with the traditional ARIMA model, the SVM model and the GRU neural network model with regard to the time series analysis of the real valve cooling data from China Southern Power Grid. The comparative experimental results demonstrate that the above four models can all clearly indicate the trend of the inlet valve temperature. However, the ARIMA-SVM hybrid model behaves better in the evaluation indicators including the root mean square error, the mean square error and the

① 基金项目: 江苏高校“青蓝工程”; 中央高校基本科研业务项目 (NS2019055)

Foundation item: “Qinglan Project” of Universities in Jiangsu Province; The Fundamental Research Funds for the Central Universities of China (NS2019055)

收稿时间: 2020-08-25; 修改时间: 2020-09-15; 采用时间: 2020-09-25; csa 在线出版时间: 2021-03-30

mean absolute error than the other three, with a more accurate prediction of inlet valve temperature.

Key words: valve cooling system; inlet valve temperature; time series analysis; ARIMA-SVM hybrid model

1 引言

目前,国内用电需求随着社会的发展而不断增长.直流输电技术^[1,2]正在朝着高电压大容量的技术方向发展,由于国土辽阔,对其在长距离输电、跨区域联网及调度灵活等方面有着现实需求.电网在运行过程中,保证换流阀等关键设备稳定运行就显得至关重要.换流阀等关键设备在电网运行过程中会不断产生热量,当设备的热量不断聚集,温度就会不断上升时则会影响电网的正常运行.电网中的换流阀等关键设备出现的故障对电力系统造成的影响也越来越大^[3].因此,需要对电网中关键设备检测数据进行预测,提高工作人员对设备的感知能力,能对关键设备未来出现的变化趋势做相应的准备.这对电网的正常运行和提供可靠的供电具有实际意义.

由于人工在处理电网设备时会存在经验不足和误操作等客观因素,因此对电的安全性、稳定性和智能化的要求越来越高.阀冷系统作为电网中冷却系统的关键设备,以水为传导介质,水有较高的热导性,作为“载热体”存在,将设备的热量传递到水介质并将热能带出,达到物理降低设备温度的目的.阀冷系统对高压直流系统的安全运行起着至关重要作用^[4].阀冷系统分为内冷水系统和外冷水系统两个部分.内冷水系统是密闭式循环具有水循环、内冷水处理等功能^[5].内冷水系统在换流阀运行时负责可靠高效通过水带走设备运行产生的热量,送到外冷水系统中.外冷水系统是开放式循环,具有水质软化净化、补水等功能^[6].冷却塔会对内冷水管道的喷淋散热,并通过风扇对外、内冷水交换将阀内冷水系统的热量排至大气环境中,两者共同合作完成对换流阀的持续冷却,可以达到对阀厅整个设备进行散热的目的.

阀冷系统在降低设备温度的过程中,其涉及的参数为出阀温度,进阀温度,出阀压力,内冷水导电率.所有参数中最核心和最关键的参数是温度,包括了出阀温度和进阀温度.阀冷系统中换流阀的进出阀温度在运行过程中只要维持在合理范围内,就能够保证换流阀设备的稳定运行从而保证阀冷系统的稳定运行.由于换流阀的出阀温度和进阀温度具有很强的相关性,

所以我们选取其中进阀温度作为整个设备运行状态的评价指标.如果设备热量过度聚集导致换流阀的进阀温度过高,可能会损坏相应设备.如果这个问题长时间不能够妥善解决,就会产生严重的安全隐患,造成难以估计的后果.进阀温度参数短期内有波动,但从长期来看,具有一定的变化趋势,可以利用时序分析如 ARIMA、SVM、GRU 以及 ARIMA-SVM 混合等方法对进阀温度进行预测,使运行工作人员提前预知进阀温度变化趋势.

2 相关工作

时序数据本质上是反应某个或者某些随机变量随时间不断变化的趋势,分析时序数据的核心是从这些数据里挖掘出某种规律,并利用规律对未来进行预估^[7].时序数据分析方法从最早提出到现在经过漫长的发展,国内外研究学者提出了许多基于时序数据分析方法^[8-11].时序分析方法主要有 3 类方法,从最开始提出的传统的时间序列预测方法,到将机器学习和时间序列预测结合的方法,以及要提高预测效率的基于参数模型的在线时间序列预测方法.无论是传统的时间序列预测方法中还是应用到时序领域的机器学习方法以及基于参数模型的在线时间序列预测方法,目前,没有一种单一模型能够统一地对特定的时间序列数据做出最好的预测.因此混合模型应运而生,混合模型受益于各个模型的多样性,减少单一模型使用风险,很好弥补单个模型的缺陷,可以提升时间序列预测精度.

经典的时间序列模型包括移动平均模型 (Move-Average, MA)、自回归模型 (Auto-Regressive, AR)、自回归移动平均模型 (Auto Regressive Moving Average, ARMA), 还有其对应的各种衍生出来的模型,如动态自回归模型和向量化的自回归模型^[12].而 Box 和 Jenkins^[13]提出的“Box-Jenkins 方法”非常流行.该方法有 3 步,第 1 步对给定时间序列,确定参数模型中适当的 p, q, d 值,第 2 步通过最有效的方法估计出参数模型的具体值,第 3 步检测模型的有效性并做出相应调整.

时间序列预测方法归根到底在本质上与基于机

机器学习分类中的回归分析有着许多相似之处。机器学习中许多方法都可以应用在时间序列预测方面,并且取得了不错的效果,例如经典的支持向量机 SVM, 贝叶斯网络 BN, GRU 神经网络。随着人工神经网络的不断发展,以及神经网络在时间序列中趋势性分析有着独特的特性,所以有人将神经网络应用在时间序列预测上,取得很好的效果。进一步深度学习也被看作是时间序列预测的有效工具。Kim^[14] 直接用支持向量机来预测股票价格,并通过对比试验来验证该方法的可行性, Gestel 等提出把贝叶斯证据框架^[15,16] 应用到最小二乘支持向量机^[17,18] 应用到金融领域进行时间序列预测。Das 和 Ghosh 将贝叶斯网络 (BN) 应用到气象时间序列预测,并在后续时间做了大量的工作。在 2017 年,提出基于语义贝叶斯网络的多元气象时间序列预测网络 semBnet^[19]。随着深度学习的流行,也被看作是实现时间序列预测的有效工具。机器学习本质上是多层神经网络,主要包括卷积神经网络 (Convolutional Neural Networks, CNN) 和循环神经网络 (Recurrent Neural Networks, RNN)^[20], 其中 RNN 能够解决时间序列的延续性问题。RNN 可以处理整个时间序列所包含的信息,但存在长期依赖问题,即对长序列进行学习时, RNN 会出现梯度消失和梯度爆炸问题,无法学习到长时间跨度的关系。为了解决这个问题, Schmidhuber 教授在 1997 年提出长短期记忆网络 (Long Short-Term Memory, LSTM) 能够很好的解决这个问题,默认就可以记住长期的信息,实现长程记忆。

随着计算机技术的发展和互联网的普及,以及数据的急速增长给传统时间序列预测方法和批处理机器学习预测方法带来了极大的挑战。在面对大数据的时候,传统的方法的效率低下,因此利用在线学习方法来对时间序列数据预测成为新的趋势,而且取得了不错的效果。和传统批处理学习方法相比,在线学习方法不需要处理整个数据集,只需要处理这个新的样本,大大节约了时间提高效率。Anava 等^[21] 提出了基于参数模型的在线时间序列预测算法 ARMA-ONS, 能够将传统的时间序列模型和在线学习有效结合起来。

3 合并单元设备状态研究

此章节主要针对阀冷系统的监测数据进行筛选和提取,并在此基础上对阀冷数据进行简单时序分析。

3.1 数据提取

本文工作的数据来源于中国南方电网从 2017 年到 2019 年收集的将近两年的真实数据,包括 PWR, 出阀温度, 进阀温度, 进阀压力, 冷却水流量, 冷却水导电率等相关记录。进一步筛选和内冷水系统有关的数据, 筛选和提取出进阀温度, 出阀温度, 进阀压力, 冷却水导电率相关数据。因此针对这个 4 个指标, 从 MySQL 数据库混乱的原始数据里提取出来, 这就需要一些数据预处理步骤, 最后可以构建出一个适合进行时间序列分析工作的数据集。

中国南方电网的设备众多, 每个阀冷设备都有相应的监测数据, 就需要对每个阀冷设备的监测数据进行分析。因此需要把数据按照各个设备进行划分, 并按照日期时间排列, 以保证每个设备对应数据的完整性和时序性, 最终构建出一个包含 4 个监测指标的数据表。

经过上面的数据预处理之后, 获得了可以进行下一步时序分析的数据集。

3.2 正常数据趋势分析

由于整个数据的采集频率为每半个小时采集一次而且时间跨度长达两年, 导致了整个数据集比较大, 对整个数据进行趋势分析只能从宏观上观察数据变化趋势, 我们对 4 个指标的数据分别按周、月对数据进行采样, 分析其内在的规律性和数据的趋势变化。对预处理的后数据集采用随机抽样, 随机选取一周的数据进行分析, 其数据随时间变化如图 1 所示。

图 1 包含了 4 个子图, 分别描绘了进阀温度, 出阀温度, 进阀压力, 冷却水导电率的变化趋势。进阀温度在 32°C 到 40°C 之间变化, 相差 8°C, 而出阀温度在 38°C 到 49°C 之间变化, 温度提升 6°C 到 9°C。进阀压力和冷却水导电率, 会随机波动, 值的波动幅度并不是很大, 且进阀压力和冷却水导电率均在正常值范围之内变化。进阀温度, 出阀温度, 进阀压力和冷却水导电率都在某个均值附近波动, 可以初步判断各个时间序列数据是平稳的, 可以通过计算迪基-福勒检验进行单位根校验, 计算出相对应的 p 值为远小于 0.01, 可以认为阀冷数据是时间平稳序列。

为了进一步发现阀冷变化趋势以及数据的平稳性, 我们对一个月的阀冷数据进行分析。从图 2 可以看出, 进阀温度和出阀温度的变化趋势很相似, 但总体的变化范围有些不同, 进阀温度的变化范围在 22°C 到 36°C, 而出阀温度则在 24°C 到 42°C 之间波动。相差了 2°C 到 6°C。

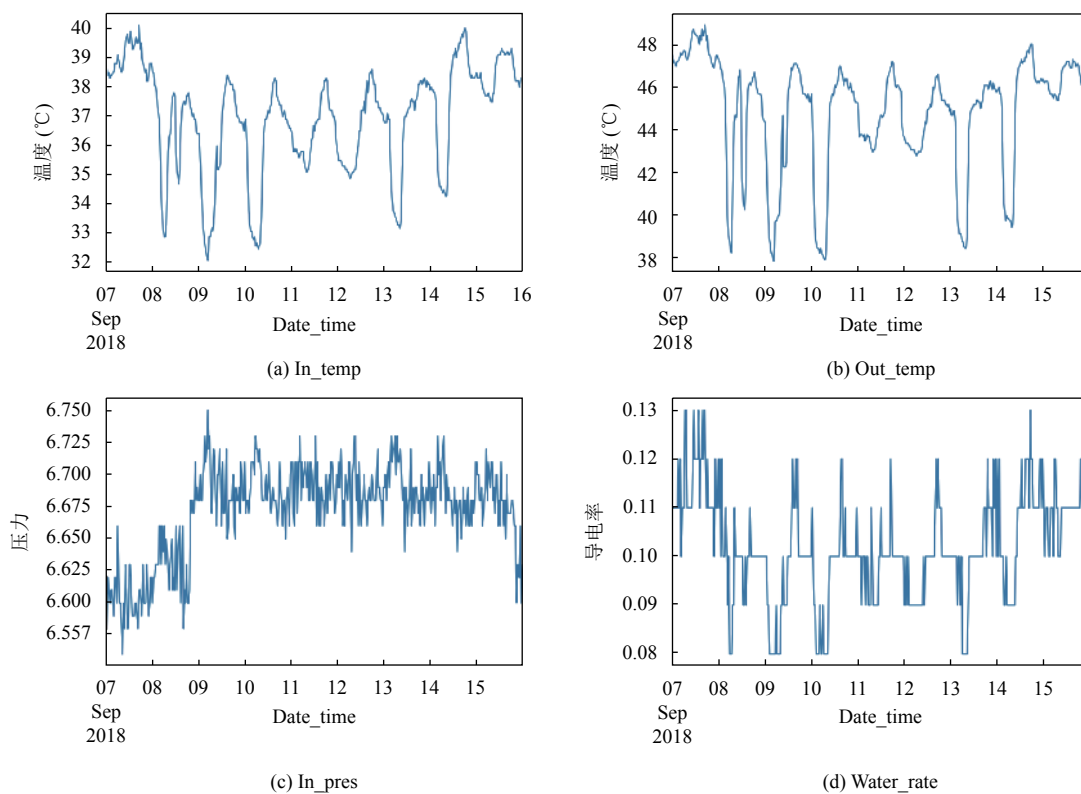


图1 阀冷4个参数的周变化趋势

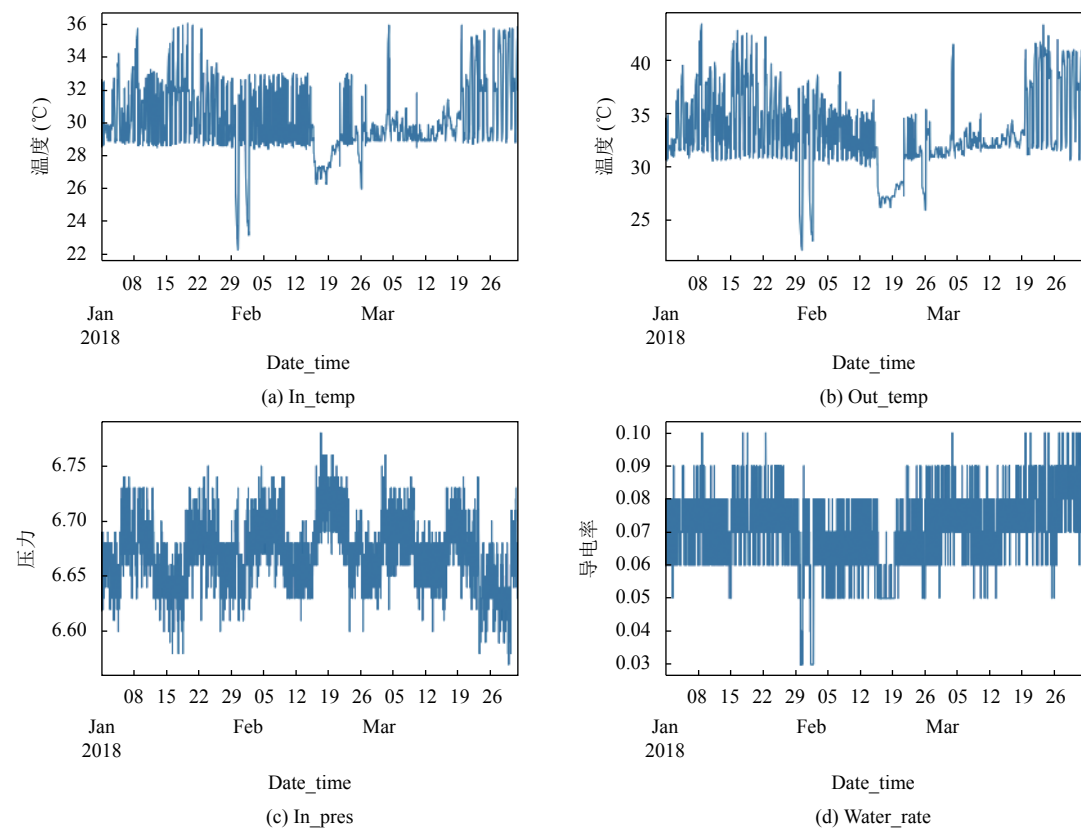


图2 阀冷4个参数的月变化趋势

从图1和图2可以发现进阀温度在22°C到36°C的范围内变化,出阀温度都在24°C到42°C范围内变化并且在各自对应的范围内变化趋势很是接近。通过相关性分析得到两者关系是正相关,相关性系数接近于1。出阀温度和进阀温度的数据波动都在正常范围之内,但是观察数据出阀温度略高一些,原因在于阀冷系统的循环水将带走其产生的热量。通过初步判断各个时间序列数据是平稳的,再进一步通过计算迪基-福勒检验进行单位根校验,计算出 p 值为远小于0.01,最终判断阀冷数据是时间平稳序列。

4 阀冷系统的分析方法

阀冷系统检测到的数据,其具有多维度、实时性、线性和非线性并存等一些特点。以往传统的时序分析方法是根据所研究时间序列的基本规律和自身特点,选取适当的时序模型并先确定时间序列模型的参数,在此基础上,利用已求解出参数的模型对未来的时间序列进行预测。传统时序模型在线性预测方面取得了很好的成果,但在面对处理复杂数据的时候,往往显得力不从心,因此需要和基于机器学习的时间序列预测方法相结合。与神经网络相比,SVM具有丰富的理论基础,并且神经网络非常依赖参数,参数的好坏直接影响神经网络的学习效果。SVM可以很好弥补传统时序分析方法的缺点的同时对线性预测效果并不理想。在对阀冷系统产生的时序数据预测,单独的ARIMA模型或者SVM模型很难完全把握时序数据的变化规律,考虑将二者结合起来的混合模型是个更好的解决方法,并和神经网络GRU进行对比实验,以验证混合模型的有效性。

在本文工作中,我们认为阀冷时间序列 y_t 可以分解为线性部分 L_t 和非线性部分 N_t 。

$$y_t = L_t + N_t \quad (1)$$

线性部分 L_t 可以由ARMA(p, q)模型根据阀冷数据时间序列中过去的值计算得出,计算公式如下:

$$L_t = a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2} + \dots + b_q \varepsilon_{t-q} + \varepsilon_t \quad (2)$$

传统时序模型ARMA只能处理平稳时间序列,比AR模型与MA模型有较精确的估计,但其参数估算也相对比较繁琐。如果是非平稳时间序列,则难以求得ARMA模型参数的准确估值。非平稳时间序列并不能用ARMA直接处理,需要经过差分处理将非平稳时间

序列转化成平稳时间序列。ARIMA模型在ARMA模型的基础上多了一个差分步骤,如ARIMA(p, d, q)模型中参数 d 为0就是ARMA模型,差分用参数 d 表示。式(2)里线性部分 L_t 由一个平稳序列 $\{y_t\}$ 和是一个白噪声序列 $\{\varepsilon_t\}$ 组成, a_i 与 b_j 分别是 $\{y_t\}$ 序列和 $\{\varepsilon_t\}$ 序列的参数。平稳序列 $\{y_t\}$ 中的 p 称为自回归阶数, $\{\varepsilon_t\}$ 序列中的 q 称为移动平均阶数。在实际应用中,很多时间序列是非平稳时间序列时,往往达不到严平稳的要求,而序列平稳性是时间序列变形分析建模的重要前提。需要对非平稳序列进行差分运算,一阶差分是相邻的时序值相减,二阶差分是对一阶差分后的序列再做一次差分运算,以此类推,可以得到 n 阶差分。非平稳序列往往一次到两次差分之后,就会变成平稳序列,达到可以建模分析的前提。如果时间序列在一阶差分平稳化后,效果不够明显,再进行二阶差分,直到获得一个较好的平稳序列。自回归滑动平均模型(ARMA)和差分整合移动平均自回归模型(ARIMA)时序分析总的策略主要包含3步,第1步确定合适的 p, d, q 值,第2步,通过最有效的方法评估模型中具体的参数值,第3步要检验预测数据的准确性,拟合模型的适当性,最终不断的适当改进模型达到准确预测数据的目的。 AIC : 赤池信息准则(Akaike Information Criterion)。

$$AIC = 2K - 2\ln(L) \quad (3)$$

BIC : 贝叶斯信息准则(Bayesian Information Criterion)。

$$BIC = K \ln(n) - 2\ln(L) \quad (4)$$

AIC 准则和 BIC 准则用于确定ARIMA模型中 p, q 值。其中, k 为模型参数个数, n 为样本数量, L 为似然函数。由于中国南方电网收集的阀冷数据本身就是时间平稳序列,因此可直接建立ARIMA(p, d, q)模型中将 d 设为0,ARIMA中 d 为0其实就是ARMA模型。首先初步确定适当的 p, d, q 的值,从低阶到高阶逐步试探方法来识别模型的参数,然后结合赤池信息度量(AIC)和贝叶斯信息度量(BIC)有效方法估计出ARIMA模型的具体参数值,最后验证拟合模型的适当性,并适当改进改进该模型。 AIC 和 BIC 是为了找到ARIMA模型的合适参数,要在模型的参数个数和模型的拟合精度做出适当的取舍,能够比较准确预测未来数据。最终确定ARIMA模型参数 $p=1, d=0, q=2$ 。ARIMA模型对线性数据具有独特的优势,但缺点也比较明显只能

捕获时间序列中线性关系,不能捕获非线性关系. ARIMA 模型在捕获线性关系过程中,还获得线性关系之外时间序列的残差信息,为接下来 SVM 分析做准备,可以进一步优化预测值.

残差里会包括非线性关系,而 ARIMA 模型则没有办法捕获非线性的信息而支持向量机 (SVM) 可以捕获此类信息. 所以残差是提高预测准确性的重要因素,因此残差分析就至关重要了. 支持向量机 (SVM) 在非线性预测方面能够取得很好的效果. SVM 是以统计学理论为基础的学习方法,其丰富的理论基础保证了 SVM 在解决数据高维度和非线性方面具有不可比拟的优势. ARIMA 在捕获线性部分之外获得时间序列的残差信息,残差即非线性部分,残差 N_t 作为 SVM 的输入.

$$N_t = y_t - \bar{L}_t \quad (5)$$

ARIMA 模型在 t 时刻得到 \bar{N}_t , \bar{N}_t 为 ARIMA 模型捕获的残差部分, \bar{N}_t 为 SVM 模型的输入,其表示为:

$$\bar{N}_t = g(N_{t-1}, N_{t-2}, \dots, N_{t-n}) + \Delta_t \quad (6)$$

式中, G 为非线性模型, Δ_t 是非线性部分的随机误差, n 为输入的长度. 根据以上的分析可以得到最终的预测值为:

$$\bar{y}_t = \begin{cases} \bar{L}_t + |\bar{N}_t|, N_{t-1} > 0 \\ \bar{L}_t - |\bar{N}_t|, N_{t-1} \leq 0 \end{cases} \quad (7)$$

5 实验评估

5.1 评估度量指标

根据数据分析结果,我们选取、确定了 ARIMA、SVM、GRU 和提出的 ARIMA-SVM 混合时序性分析方法,尝试建立预测模型,对阀冷系统时序数据进行预测分析. 根据预测的结果需要评估模型的好坏,我们采用多个度量指标来评估数据的是否平稳时间序列以及模型的准确性.

均方根误差公式表示如下:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_{\text{real},i} - X_{\text{predict},i})^2} \quad (8)$$

均方误差公式表示如下:

$$MSE = \frac{1}{n} \sum_{i=1}^n (X_{\text{real},i} - X_{\text{predict},i})^2 \quad (9)$$

平均绝对误差表示如下:

$$MAE = \frac{1}{n} \sum_{j=1}^1 |(x_{\text{real},i} - x_{\text{predict},i})| \quad (10)$$

DF 检验 (Dikey-Fuller test). DF 检验是一种常见的用于判断时间序列平稳性的单位根检测方法. 若通过计算不存在单位根,时间序列就是平稳时间序列,反之,需要进行差分处理. 时间序列的平稳性检测是时间序列建模过程中非常重要的一步,使用 DF 检测方法可以检验之.

6 实验

首先我们选取中国南方电网的一个阀冷系统真实数据,从 2017 年 12 月份到 2019 年 7 月份的真实数据中选取连续 3 个月的数据,我们选取的 2018 年 1 月到 3 月的数据,对这 3 个月的数据量进行实验分析评估,阀冷数据采集的时间间隔为半个小时,数据的变化趋势图 2 所示.

从图 2 中可以看到阀冷系统在平常正常工作状态中,进阀温度,出阀温度,进阀压力,冷却水导电率趋于平稳,会产生随机波动,但波动都在正常范围之内. 由于进阀温度比较重要,所以我们预测输入以进阀温度为预测标准.

ARIMA 模型. 第 1 步确定合适的 p, d, q 值,在初步确定 ARIMA(p, d, q) 由于阀冷数据都是时间平稳序列,所以 ARIMA 模型中 $d=0$. 第 2 步,通过根据 AIC 和 BIC 检测方法确定 ARIMA 模型中的 p 和 q . 第 3 步要检验预测数据的准确性,拟合模型的适当性,最终不断的适当改进模型达到准确预测数据的目的,最终 ARIMA 模型参数为 $p=1, d=0, q=2$.

实线为数据真实值,虚线为模型在测试集上的预测. 从图 3 而言,ARIMA 模型的进阀温度预测结果和阀冷进阀温度数据很是接近,但在峰值区域的预测效果有些欠缺.

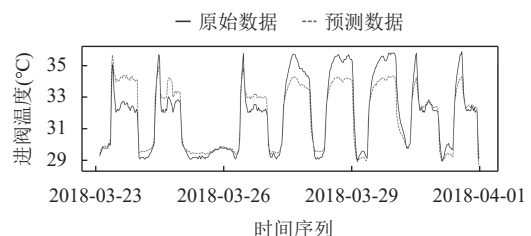


图 3 ARIMA 预测图

SVM 模型. SVM 模型可以检测非线性关系,以 VC 理论和结构风险最小原理为基础的,通过选择适当的判别函数,希望 SVM 的测试误差较小. 用 SVM 来

单独预测进阀温度数据中的非线性关系。

SVM 的预测性能主要和误差惩罚因子 C 和选取核函数有关。优化误差惩罚因子 C 和核函数能够很好提升 SVM 的性能, 确定惩罚因子 $C=10$, 选用径向基函数 (rbf) 作为核函数。 C 惩罚因子是确定数据子空间时调节置信区间范围。采用遗传算法, 对两个参数优化。

实线为数据真实值, 虚线为模型在测试集上的预测。从图 4 而言, SVM 模型的进阀温度预测很好拟合了真实进阀温度的变化趋势, 但是和真实进阀温度的真实值仍有差距。

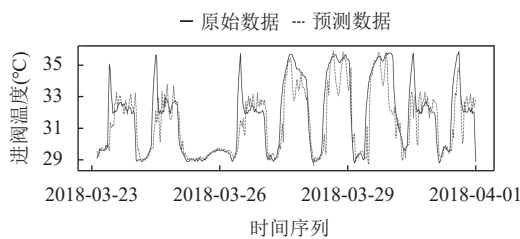


图 4 SVM 预测图

GRU 神经网络模型。为了进一步说明混合模型能够在单一模型的基础上提高预测精度, 将混合模型和 GRU 神经网络做对比实验。GRU 在循环神经网络的基础上改进的变体, 引入门控机制, 自适应控制信息流动, 可以捕捉时序数据之间的相互依赖关系, 具有一定的优势。GRU 神经网络模型在训练时涉及到许多超参数的设定, 神经元数量 m , 时间步长 T , 批数据大小 $batchsize$, 迭代次数 $epoch$ 轮。神经元个数决定了神经网络对时序数据的拟合程度, 时间步长和批数据大小决定了模型训练的结果。选择 ADAM 优化器对参数进行优化调整。

实线为数据真实值, 虚线为模型在测试集上的预测。从图 5 而言, GRU 模型的进阀温度预测结果比 ARIMA 模型预测结果更加接近真实的进阀温度数据, 在 3 个峰值区域的预测效果有些欠缺。

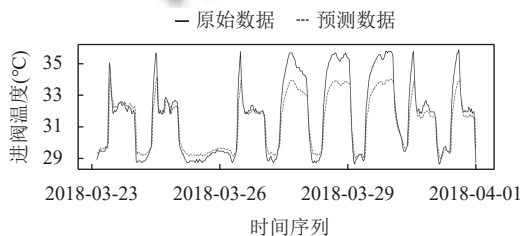


图 5 GRU 预测图

ARIMA 模型可以很好捕获时间序列中线性关系, 但时序数据仍有非线性关系没有捕捉。利用 SVM 可以

很好捕捉非线性关系, 从 ARIMA 获得线性关系之外时间序列的残差信息获得非线性关系。进一步优化预测值。从而提出 ARIMA-SVM 混合模型。

ARIMA-SVM 混合模型结合两个模型的优点。首先使用 ARIMA 对进阀温度时间序列进行预测, 捕获进阀温度时间序列的线性变化趋势, 然后利用 SVM 对进阀温度时间序列的非线性, 以提高预测的准确性。在 ARIMA 模型预测的同时获得了残差数据, 将获得的残差部分作为 SVM 的输入, 根据 SVM 模型的输出进一步调整 ARIMA 的预测值, 使得预测值更加接近真实值。

实线为数据真实值, 虚线为模型在测试集上的预测。从图 6 而言, ARIMA-SVM 混合模型的进阀温度预测结果比 ARIMA 模型预测结果更加接近真实的进阀温度数据。ARIMA-SVM 混合模型结合两个模型的优点。首先使用 ARIMA 对电网时间序列进行预测, 捕获电网时间序列的线性变化趋势, 然后利用 SVM 对电网时间序列的非线性, 以提高预测的准确性。在 ARIMA 模型预测的同时获得了残差数据, 将获得的残差部分作为 SVM 的输入, 根据 SVM 模型的输出进一步调整 ARIMA 的预测值, 使得预测值更加接近真实值。

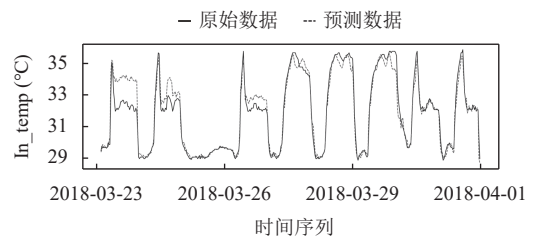


图 6 混合模型预测

为了能够从整体评估各个模型的预测能力, 计算出各个模型的均方根误差、均方误差和平均绝对误差进行比较。结果如表 1 所示, 根据本文选取的均方根误差、均方误差和平均绝对误差 3 个评价指标来衡量预测结果的准确性, 3 个评价指标可以很好的衡量各个模型预测值和真实值之间的偏差和拟合精度, 值越小说明精度越高, 预测效果也越好。从表 1 可以看出 ARIMA-SVM 混合模型在整体的误差上相比其他 3 个模型都有一定的优势, 能够更好的预测出进阀温度, 提高预测精度。

表 1 各模型的评价指标

评价指标	SVM	ARIMA	RNN	ARIMA-SVM
RMSE	1.33	0.72	0.75	0.60
MSE	1.78	0.52	0.57	0.36
MAE	0.83	0.65	0.53	0.39

7 结语

高压直流运行中换流阀会产生热能,作为换流阀的关键设备和重要降温设备,阀冷系统以水为介质将换流阀热能带出以达到冷却目的。换流阀若要安全稳定地运行,将设备运行时的温度维持在合理范围内至关重要。进阀温度和出阀温度是阀冷系统最为关键的两个参数。本文通过构建 ARIMA 模型、SVM 模型、GRU 神经网络模型以及 ARIMA-SVM 的混合模型对阀冷数据进行时序分析,针对进阀温度这一关键参数进行预测。实验结果表明,虽然某些地方有些许偏差,ARIMA 模型大体上能够较好地预测进阀温度但偏差比较大。SVM 模型可以预测数据的变化趋势。GRU 神经网络模型和 ARIMA 模型相比预测值更加接近进阀温度的真实值,但在 3 个评价指标上稍劣于 ARIMA-SVM 混合模型。通过在 ARIMA 模型分析线性数据的基础上结合 SVM 模型,进而提高进阀温度预测准确性。在 ARIMA 模型预测结果的基础上利用捕获的残差信息做进一步优化,实验结果表明将传统时序模型与机器学习结合提出 ARIMA-SVM 的混合模型可以达到一个优化的结果。

参考文献

- 徐德明,徐健,姚晓玲.智能电网的关键技术与功能实现.电工电气,2011,(11):1-5,41.[doi:10.3969/j.issn.1007-3175.2011.11.001]
- 孙玉娇,周勤勇,申洪.未来中国输电网发展模式的分析与展望.电网技术,2013,37(7):1929-1935.
- 崔宗清,孙仁龙,邹荣盛.换流站接地极在线监测系统电源的设计.集成电路应用,2019,36(10):46-48.
- 张望平,李家羊,谌毅,等.高压直流阀冷却系统喷淋水净化工艺.广东电力,2018,31(5):107-114.
- 罗永金,杨垒,马根坡,等.复合式换流阀冷却系统的设计与试验.自动化仪表,2018,39(7):11-15.
- 崔鹏飞,王海军,国建宝.直流输电工程阀冷系统在线仪表可靠控制策略.广东化工,2011,38(12):112-113.[doi:10.3969/j.issn.1007-1865.2011.12.057]
- 原继东,王志海.时间序列的表示与分类算法综述.计算机科学,2015,42(3):1-7.[doi:10.11896/j.issn.1002-137X.2015.03.001]
- 张美英,何杰.时间序列预测模型研究综述.数学的实践与认识,2011,41(18):189-195.
- 吴志华,韩燮,张永梅,等.基于知识发现的时序数据分析方法.测试技术学报,2002,16(S1):571-575.
- 杨海民,潘志松,白玮.时间序列预测方法综述.计算机科学,2019,46(1):21-28.[doi:10.11896/j.issn.1002-137X.2019.01.004]
- 王丽娜,肖冬荣.基于 ARMA 模型的经济非平稳时间序列的预测分析.武汉理工大学学报(交通科学与工程版),2004,28(1):133-136.
- 何永沛.时间序列数据挖掘在股市预测分析中的应用研究[硕士学位论文].重庆:重庆大学,2008.
- Box GEP, Jenkins GM. Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day, 1990.
- Kim KJ. Financial time series forecasting using support vector machines. Neurocomputing, 2003, 55(1-2): 307-319.
- MacKay DJC. Bayesian interpolation. Neural Computation, 1992, 4(3): 415-447.
- MacKay DJC. Probable networks and plausible predictions—a review of practical Bayesian methods for supervised neural networks. Network: Computation in Neural Systems, 1995, 6(3): 469-505.
- Suykens JAK, Vandewalle J. Least squares support vector machine classifiers. Neural Processing Letters, 1999, 9(3): 293-300.
- Suykens JAK. Least squares support vector machines for classification and nonlinear modelling. Neural Network World, 2000, 10(1): 29-48.
- Das M, Ghosh SK. semBnet: A semantic Bayesian network for multivariate prediction of meteorological time series data. Pattern Recognition Letters, 2017, 93: 192-201.
- 张敏.跨语言文本情感分类技术研究[硕士学位论文].厦门:华侨大学,2017.
- Anava O, Hazan E, Mannor S, et al. Online learning for time series prediction. Journal of Machine Learning Research, 2013, 30: 172-184.