

DOI: 10.20040/j.cnki.1000-7709.2023.20220418

基于 ARIMA-SVM 方法的梯级泵站机组运行趋势预测

徐存东^{1,2,3}, 王鑫^{1,3}, 田俊姣^{1,3}, 刘子金^{1,3}, 赵志宏^{1,3}, 陈家豪^{1,3}, 胡小萌^{1,3}

(1. 华北水利水电大学水利学院, 河南 郑州 450046; 2. 浙江省农村水利水电资源配置与调控关键技术重点实验室, 浙江 杭州 310018; 3. 河南省水工结构安全工程技术研究中心, 河南 郑州 450046)

摘要: 针对多因素参与下梯级泵站机组运行趋势预测建模困难且预测准确度低、适应性差的问题, 以宁夏盐环定扬黄工程为研究对象, 引入时间序列分析法, 提出了基于 ARIMA 与 SVM 模型组合的泵站机组运行趋势预测方法, 即选择机组运行技术参数中的能源单耗和平均负荷作为试验样本, 由 ARIMA 建模对处理后的数据进行线性拟合, 通过 SVM 模型对残差进行预测处理, 补偿机组运行中的非线性变化, 综合二者预测结果得到组合模型预测值。结果表明, 最优模型为 ARIMA(1, 1, 3)、ARIMA(2, 1, 1), SVM 模型最优参数分别为 $c=38, g=0.06$ 和 $c=68, g=0.18$; 组合模型对试验样本的预测拟合优度分别达到 0.999 2、0.998 4, 均方根误差分别为 1.67×10^{-5} 、 3.9×10^{-8} , 平均绝对百分比误差分别为 0.036 1%、0.074 7%, 说明该组合模型预测泵站机组运行趋势精度较高、效果良好, 可为泵站机组运行状态监测系统优化升级提供理论基础。

关键词: 机组运行趋势; 时间序列; ARIMA-SVM; 差分自回归移动平均; 组合模型; 预测

中图分类号: TV675; TV93

文献标志码: A

文章编号: 1000-7709(2023)02-0133-04

1 引言

水泵机组是梯级泵站中水流传递能量的载体, 是泵站运行的关键设备, 开展泵站机组运行趋势预测研究, 对保障水泵机组安全运行和实现泵站节能降耗具有重要意义。近年来, 对于泵站设备故障预测和泵站运行监测及其优化已做了大量研究^[1-5], 但总体来看, 主要集中于泵站管道振动趋势预测、泵站异常状态的诊断和机组正常运行时的监测及效率优化, 而对梯级泵站机组的未来运行趋势预测尚未进行深入研究。泵站机组运行状态呈纵向的时间相关性, 可使用时间序列法预测泵站机组运行趋势, 差分自回归移动平均—支持向量机 (ARIMA-SVM) 组合模型可很好地在预测过程中体现其线性自相关性, 同时对残差值进行非线性修正建模, 弥补了 ARIMA 模型的非线性捕捉缺陷^[6-8], 具备抗噪能力和推广能力强等优点。鉴此, 本文通过收集分析宁夏盐环定扬黄灌区泵站机组长序列监测数据, 选取能源单耗和

平均负荷两个泵站机组运行参数进行建模, 将 ARIMA-SVM 预测模型引入泵站机组运行趋势预测, 并将预测结果与单一 ARIMA 模型预测结果进行比较, 验证了该模型在机组趋势预测中的可行性, 旨在为机组运行监测模式优化提供理论基础, 达到在线监测泵站机组安全运行的目的。

2 基本理论与模型构建

2.1 差分自回归移动平均模型

差分自回归移动平均模型 (ARIMA) 是利用随机干扰项和时间序列自身的滞后序列来分析描述时间序列发展规律的方法, 该模型可将非平稳的时间序列差分转化为所需的平稳时间序列, 模型的实际预测过程包含平稳化检验、差分过程、模型选择等步骤。

ARIMA 模型由自回归系数 p 、差分阶数 d 及滑动平均系数 q 来决定。模型中 p 阶自回归模型 (AR) 为:

$$y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \varepsilon_t \quad (1)$$

收稿日期: 2022-03-07, **修回日期:** 2022-05-20

基金项目: 国家自然科学基金项目 (51579102); 河南省高校科技创新团队支持计划 (19IRTSTHN030); 中原科技创新领军人才支持计划 (204200510048); 河南省科技攻关项目 (212102310273); 河南省高等学校重点科研项目计划 (20A570006); 浙江省重点研发计划 (2021C03019); 浙江省基础公益研究计划项目 (LZJWD22E090001)

作者简介: 徐存东 (1972-), 男, 教授, 研究方向为水工结构优化设计, E-mail: xcundong@126.com

式中, y_t 为待求序列的值; c 为常数项; φ_1 为自回归系数; y_{t-1} 为待求的滞后阶; ε_t 为残差白噪声序列。

q 阶滑动平均模型(MA)为:

$$y_t = c + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (2)$$

式中, θ_i 为移动平均系数。

自回归—滑动平均模型(ARMA)为:

$$y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (3)$$

对于泵站机组运行状态数据生成的时间序列,通过 ADF 检验判断数据的平稳性,若数据呈非平稳性,则通过 d 阶差分得到所需的平稳性数据。结合数据自相关函数(ACF)和偏自相关函数(PACF)图,判断拖尾截尾性得到 p, q ; 若 p, q 数值偏大,可运用 Matlab 等软件,通过 AIC、BIC 等准则暴力选定阶数。

2.2 支持向量机

支持向量机(SVM)的基本原理是采用核函数 $\varphi(x)$ 代替传统线性方程中的线性项,将低维线性空间通过非线性变换映射到高维空间。表达式为:

$$f(x) = w\varphi(x) + b \quad (4)$$

式中, $f(x)$ 为待拟合预测函数; w 为权值量; b 为偏置量。

SVM 的核心为寻找“最优超平面”,求解模型为:

$$\begin{cases} \min_{w,b} \|w\|^2 / 2 \\ \text{s. t. } y_i (w^T x_i + b) \geq 1 \end{cases} \quad (5)$$

在处理多维非线性问题时,可在式(5)中引入在错分样本比例与算法复杂度之间实现折中的惩罚系数 c 和将原始空间映射到高维空间实现线性可分的核函数 $K(x, x_i)$, 其中高斯核函数是最常用的函数之一,表达式为:

$$K(x, x_i) = \exp(-g \|x - x_i\|^2) \quad (6)$$

式中, g 为高斯核函数的一个伽玛参数,其决定了数据映射至高维空间后的分布。

因此,惩罚系数 c 和伽玛参数 g 的选择,会直接影响 SVM 模型的性能。

2.3 ARIMA-SVM 组合预测模型

ARIMA 模型和 SVM 模型在解决预测问题时各有优缺点,无法适应所有特征结构的数据。为兼顾泵站机组运行中的线性与非线性特征,提出基于 ARIMA-SVM 的泵站机组运行趋势预测模型,以增强预测模型对样本数据的适应性,提高预测精度。组合模型为:

$$Y_t = L_t + N_t \quad (7)$$

式中, Y_t 为时间序列数据; L_t 为线性自相关部分; N_t 为非线性部分。

由 ARIMA 模型预测初始泵站机组运行趋势,再通过 SVM 模型对产生的残差修正预测,最终预测值计算公式为:

$$\hat{Y}_t = \hat{L}_t + \hat{h}_t \quad (8)$$

式中, \hat{Y}_t 为组合模型预测值; \hat{L}_t 为 ARIMA 模型初始预测值; \hat{h}_t 为 SVM 模型残差预测值。

具体建模步骤见图 1。

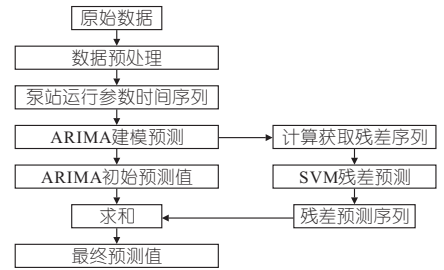


图 1 ARIMA-SVM 组合模型预测流程

Fig. 1 ARIMA-SVM combination model forecasting process

3 实例应用

盐环定扬黄灌区横跨陕、甘、宁三省,由三省(自治区)联合兴建,灌区建有 24 座泵站,总扬程为 391.2 m,总提水流量为 11 m³/s,干、支渠总长 480 km,供水管网总长 1 127 km,泵站输水过程是多梯级、长距离、流量变化频繁的多因素参与过程,灌区泵站运行时间为 4~8、11 月。能源单耗和平均负荷能直观反映机组工作的能耗情况与提水用电量,因此选取盐环定扬黄灌区一泵站 2018 年 4~8 月的能源单耗(kW·h/(kt·m))和平均负荷(10⁴ kW)两个机组运行技术参数作为研究对象,每个参数有 50 个试验样本。

3.1 平稳性检验

预处理数据,判断原时间序列是否具有平稳性。如图 2 所示,两个泵站机组运行参数的时间序列不具备零均值的特点,其方差也不断变化,所以初步判断原运行参数的时间序列具有不平稳性。

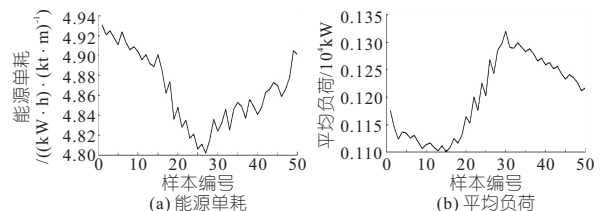


图 2 运行参数值

Fig. 2 Run parameter values

通过原始序列的自相关图(ACF)和偏自相关图(PACF)(图 3)可看出,ACF 未出现拖尾或

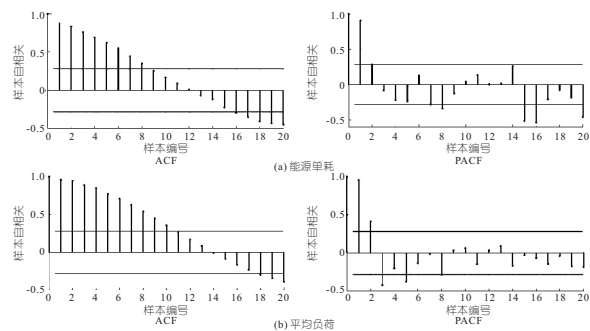


图 3 运行参数值 ACF 和 PACF 图

Fig. 3 Operating parameter values ACF and PACF graphs
截尾的痕迹,而是存在类似三角对称的现象,说明原始数值序列不具有平稳性,属于一个非平稳序列。

3.2 线性部分 ARIMA 预测

取 4~7 月份数据作为训练集,8 月份数据作为测试集。对数据进行一阶差分后数据趋于平稳,但自相关系数 p 和滑动平均系数 q 的取值偏大。因此,通过 AIC、BIC 等准则暴力选定阶数,得到能源单耗和平均负荷的最优模型为 ARIMA(1,1,3)、ARIMA(2,1,1),预测结果见图 4。由图 4 可知,ARIMA 模型未能很好地追踪 2 个泵站机组运行参数的趋势变化,且波动偏差较大。

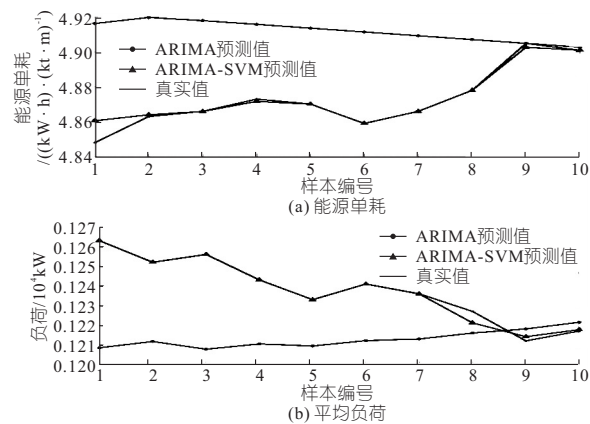


图 4 预测结果对比

Fig. 4 Comparison of prediction results

3.3 组合模型预测

计算 ARIMA 模型预测值残差,并对残差序列 $\{h_t\}$ 进行归一化处理,计算公式为:

$$h_t = y_t - L_t \quad (9)$$

式中, h_t 为所求残差; y_t 为实际值; L_t 为预测值。

$$\hat{x} = (N_{\max} - N_{\min}) \left(\frac{x - x_{\min}}{x_{\max} - x_{\min}} \right) + N_{\min} \quad (10)$$

式中, x 为原始数据; N_{\max} 、 N_{\min} 分别为归一化区间的最大值、最小值,取值为 $[-1, 1]$; x_{\max} 、 x_{\min} 分别为样本数据的最大值、最小值。

对 SVM 模型参数进行寻优,得到能源单耗和平均负荷的惩罚系数 c 和伽玛参数 g 分别为 $c=38, g=0.06$ 和 $c=68, g=0.18$ 。将参数代入

SVM 模型反归一化后得到预测残差值,由式(8)得到组合模型预测值,预测结果见图 4。

3.4 预测效果分析

对比 ARIMA-SVM 组合模型和 ARIMA 模型对泵站机组运行趋势的曲线拟合效果(图 4),发现组合模型的预测曲线相较于 ARIMA 模型与实际值的契合度更高,基本与实际值的变化趋势一致,说明应用 SVM 模型降低了预测误差,提高了组合模型预测精度,运用 SVM 模型对 ARIMA 模型预测值进行二次残差预测处理,通过高维空间映射解决了梯级泵站机组运行趋势预测中的非线性特征问题。

图 5 为 ARIMA-SVM、ARIMA 所得预测值与实际值误差对比。由图 5 可知,ARIMA-SVM 组合模型所得预测值与实际值误差趋势较为平缓,结合表 1,可以发现组合模型预测效果优于 ARIMA 模型,且 SVM 对初始预测值的修正效果也较为显著。

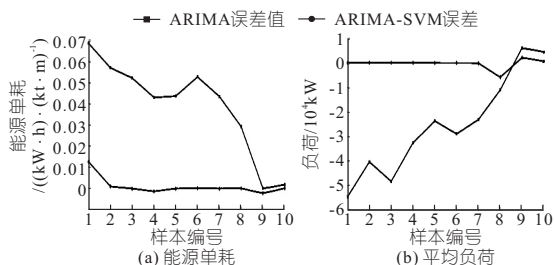


图 5 误差对比图

Fig. 5 Error comparison figure

表 1 能源单耗、平均负荷的模型预测结果误差对比

Tab. 1 Model prediction results error comparison of the energy consumption and average load

泵站机组	实际值	ARIMA 模型		ARIMA-SVM 组合模型	
		预测值	误差/%	预测值	误差/%
能源单耗	4.848	4.916 400	6.840	4.860 582	12.582
	4.863	4.919 851	5.685	4.864 008	1.008
	4.866	4.918 045	5.205	4.865 921	0.079
	4.873	4.915 908	4.291	4.871 691	1.309
	4.870	4.913 731	4.354	4.870 131	0.060
	4.859	4.911 548	5.255	4.859 073	0.073
	4.866	4.909 365	4.334	4.865 942	0.058
	4.878	4.907 182	2.918	4.878 049	0.049
	4.905	4.904 990	0.001	4.902 745	2.226
	4.901	4.902 816	0.182	4.901 076	0.076
平均负荷	0.126 3	0.120 848	0.545	0.126 305	0.005
	0.125 2	0.121 169	0.403	0.125 207	0.007
	0.125 6	0.120 776	0.482	0.125 606	0.006
	0.124 3	0.121 043	0.326	0.124 308	0.008
	0.123 3	0.120 939	0.236	0.123 295	0.005
	0.124 1	0.121 212	0.289	0.124 093	0.007
	0.123 6	0.121 296	0.230	0.123 591	0.009
	0.122 7	0.121 599	0.110	0.122 118	0.582
	0.121 2	0.121 809	0.061	0.121 412	0.212
	0.121 7	0.122 148	0.044	0.121 771	0.071

注:能源单耗、平均负荷单位分别为 $\text{kW} \cdot \text{h}/(\text{kt} \cdot \text{m}^3)$ 、 10^4kW 。

为进一步验证 ARIMA-SVM 组合模型在泵站运行趋势预测中的合理性,使用均方根误差

(R_{RMSE}), 平均绝对百分比误差 (M_{MAPE}), 拟合优度 (R^2) 共 3 项指标评价运行参数的预测结果, 结果见表 2。

表 2 预测结果指标统计

Tab. 2 Statistics of prediction results

模型	能源单耗值			平均负荷值		
	R_{RMSE}	$M_{MAPE}/\%$	R^2	R_{RMSE}	$M_{MAPE}/\%$	R^2
ARIMA	0.002 0	0.803 2	0.990 9	1.01×10^{-5}	2.187 0	0.974 4
ARIMA-SVM	1.67×10^{-5}	0.036 1	0.999 2	3.9×10^{-8}	0.074 7	0.998 4

由表 2 可得, ARIMA-SVM 组合模型的 R_{RMSE} 、 M_{MAPE} 值显著降低, R^2 值升高, 表明 ARIMA-SVM 组合模型预测精度更高。盐环定扬黄灌区一泵站 4~8 月份的能源单耗和平均负荷样本数据能够直观反映其机组的工作状态, 利用 ARIMA-SVM 组合模型对泵站机组实际运行状态进行预测, 在应对能源单耗和平均负荷 2 种类型的数据时均获得很高的预测精度, 是一种准确可行的泵站机组运行趋势预测模型。在泵站机组实际运行中, ARIMA-SVM 组合模型对机组运行趋势的预测可为明确机组潜在故障点、提前采取调控措施、保障机组良好运行提供一定的理论指导。

4 结论

- a. 将 ARIMA-SVM 组合模型应用于泵站机组运行趋势预测中, 得到泵站机组运行技术参数能源单耗和平均负荷的 ARIMA 最优模型分别为 (1,1,3)、(2,1,1), SVM 模型最优惩罚系数 c 和伽玛参数 g 分别为 $c=38, g=0.06$ 和 $c=68, g=0.18$ 。
- b. ARIMA-SVM 组合模型对能源单耗和平均

均负荷的预测拟合优度 R^2 分别达到了 0.999 2、0.998 4, 均方根误差 R_{RMSE} 分别为 1.67×10^{-5} 、 3.9×10^{-8} , 平均绝对百分比误差分别为 0.036 1%、0.074 7%, 较 ARIMA 模型均方根误差、平均绝对百分比误差值显著降低, R^2 值升高, 说明 SVM 模型提升了组合模型预测精度, ARIMA-SVM 组合模型预测效果良好。

参考文献:

- [1] ZHANG J, HUA W, HOU G. Prediction and analysis of the vibration response trend of pumping station pipeline by IVMD[J]. Journal of vibration, measurement & diagnosis, 2019, 39(3):478-483,666.
- [2] PARK S E, ROH H W. CFD prediction on vortex in sump intake at pump station[J]. Ksfm journal of fluid machinery, 2007, 10(4):39-46.
- [3] 张建伟, 江琦, 刘轩然, 等. 基于 PSO-SVM 算法的梯级泵站管道振动响应预测[J]. 农业工程学报, 2017, 33(11):75-81.
- [4] 李宝, 李俊梅, 钮月磊, 等. 基于大数据技术的多沙水流泵站故障诊断方法[J]. 水电能源科学, 2020, 38(5):160-162,94.
- [5] 李梦轩, 田雨, 雷晓辉. 引江济淮梯级泵站事故停泵应急响应时间分析[J]. 人民长江, 2021, 52(5):212-217.
- [6] 宋国君, 国潇丹, 杨啸, 等. 沈阳市 PM_{2.5} 浓度 ARIMA-SVM 组合预测研究[J]. 中国环境科学, 2018, 38(11):4031-4039.
- [7] 杨敏, 丁剑, 王炜. 基于 ARIMA-SVM 模型的快速公交停站时间组合预测方法[J]. 东南大学学报(自然科学版), 2016, 46(3):651-656.
- [8] 刘胜. 基于 ARIMA 与 SVM 组合模型的国内旅游市场预测研究[D]. 杭州: 东华理工大学, 2017.

Operation Trend Prediction of Cascade Pumping Stations Based on ARIMA-SVM Method

XU Cun-dong^{1,2,3}, WANG Xin^{1,3}, TIAN Jun-jiao^{1,3}, LIU Zi-jin^{1,3}, ZHAO Zhi-hong^{1,3}, CHEN Jia-hao^{1,3}, HU Xiao-meng^{1,3}

(1. School of Water Conservancy, North China University of Water Resources and Electric Power, Zhengzhou 450046, China; 2. Key Laboratory for Technology in Rural Water Management of Zhejiang Province, Hangzhou 310018, China; 3. Henan Provincial Hydraulic Structure Safety Engineering Research Center, Zhengzhou 450046, China)

Abstract: Aiming at the problems of difficult modeling, low prediction accuracy and poor adaptability of the operation trend prediction of cascade pumping station units under the process of multi-factor participation, this study took Yanhuanding Yellow River Project in Ningxia as the research object, introduced the time series analysis method, and put forward the operation trend prediction method of pumping station units based on ARIMA and SVM combination model. The energy consumption and average load in the operation technical parameters of the unit were selected as the test samples. The ARIMA model was used to linearly fit the processed data, and the SVM model was used to predict the residual error to compensate for the nonlinear change in the operation of the unit. The prediction results of the combination model were obtained by combining the two prediction results. The results show that the optimal models are ARIMA (1,1,3) and ARIMA (2,1,1), and the optimal parameters of SVM model are $c=38, g=0.06$ and $c=68, g=0.18$, respectively. The goodness of fit of the combined model for the test samples were 0.999 2 and 0.998 4, R_{RMSE} were 1.67×10^{-5} and 3.9×10^{-8} , the M_{MAPE} were 0.036 1% and 0.074 7%, indicating that the combined model has high accuracy and good effect in predicting the operation trend of pumping stations. ARIMA-SVM combination model can provide a theoretical basis for the optimization and upgrading of pumping station unit operation condition monitoring system.

Key words: unit operation trend; time series; ARIMA-SVM; differential autoregressive moving average; combination model; prediction