

文章编号: 1007-7294(2024)01-0036-09

# 基于经验知识的船舶螺旋桨水动力样本扩充方法及其在预报代理模型中的应用

谢 硕<sup>1,2</sup>, 陈奕宏<sup>1,2,3</sup>, 强以铭<sup>1,2</sup>, 李 亮<sup>1,2</sup>

(1. 中国船舶科学研究中心, 江苏 无锡 214082; 2. 深海技术科学太湖实验室, 江苏 无锡 214082;  
3. 浙江大学 航空航天学院, 杭州 310027)

**摘要:**近几年来,越来越多学者致力于采用机器学习进行船舶螺旋桨性能预报,但代理模型的预报效果往往受训练样本的数量和质量影响较大。目前船舶螺旋桨性能数据数量少且质量也不高,即数据对应参数分布较为集中,严重不均匀,这些问题都会影响代理模型预报的准确率与可靠性。为了解决这一实际困难,本文提出一种基于经验知识的样本扩充方法,并以船舶螺旋桨水动力性能的代理模型预报为例进行应用验证。结果表明:本文使用的样本扩充方法可以快速生成数据样本,可以一定程度上改善预报代理模型的可靠性,提升预报精度。

**关键词:** 样本扩充; 经验知识; 机器学习; 船舶螺旋桨; 水动力性能; 代理模型

**中图分类号:** U661.1 **文献标识码:** A **doi:** 10.3969/j.issn.1007-7294.2024.01.004

## A method of data expansion for marine propeller hydrodynamic performance based on priori knowledge and its application

XIE Shuo<sup>1,2</sup>, CHEN Yi-hong<sup>1,2,3</sup>, QIANG Yi-ming<sup>1,2</sup>, LI Liang<sup>1,2</sup>

(1. China Ship Scientific Research Center, Wuxi 214082, China; 2. Taihu Laboratory of Deepsea Technological Science, Wuxi 214082, China; 3. School of Aeronautics and Astronautics, Zhejiang University, Hangzhou 310027, China)

**Abstract:** In recent years, more and more researchers have applied machine learning to predict the performance of ship propellers, but the prediction effectiveness of surrogate model is often affected by the quantity and quality of data used for training. At present, the quantity and quality of the ship propeller performance data are unsatisfactory, and the distribution of data corresponding parameters is relatively centralized and seriously uneven. Therefore, these facts may affect the accuracy and reliability of surrogate models. In order to solve this problem, this paper presents a sample expansion method based on empirical knowledge, and applies it to the prediction of ship propeller hydrodynamic performance. The results show that the sample expansion method can generate the data sample quickly, and improve the reliability and accuracy of the forecasting surrogate model.

**Key words:** sample expansion; experience knowledge; machine learning; ship propeller; hydrodynamic performance; surrogate model

收稿日期: 2023-07-27

作者简介: 谢 硕(1998-),男,硕士研究生,E-mail: 1917176129@qq.com。

## 0 引 言

螺旋桨是船舶推进系统的重要组成部分。一方面,随着海上航运业的发展,船舶的吨位越来越大,螺旋桨负荷越来越重;另一方面,人们日益重视船舶的节能减排和降低船舶给海洋造成的污染。因此,对螺旋桨的推进效率、噪声性能等方面都有了更高的要求。

螺旋桨性能预报是优化设计中的重要一环,直接影响推进效率和噪声性能的提升。其中,水动力性能是螺旋桨众多性能中的一个重要方面,直接影响船机桨匹配及船舶快速性。目前,对螺旋桨水动力性能研究主要有三类方法:基于真实水池的螺旋桨敞水试验、以面元法为代表的势流理论计算方法<sup>[1-2]</sup>以及粘流 CFD 数值预报方法<sup>[3-4]</sup>。近年来,机器学习凭借其强大的特征提取能力、可以随数据增长实现自生长的特性引起了研究人员的关注。越来越多学者开始采用机器学习方法生成代理模型,对螺旋桨的水动力性能进行预报<sup>[5-6]</sup>。这种方法可以大大加快螺旋桨性能预报的计算速度,有效节约时间成本,在螺旋桨的设计及优化中有良好的应用前景。机器学习模型的可靠性和准确率与数据样本的数量和质量密切相关,然而,试验数据获取成本高、周期较长、虚拟数据可靠性差等问题导致螺旋桨水动力性能可以获得的样本数量仍然较少,难以达到机器学习要求的理想数量等级。且螺旋桨参数众多,样本数据大多属于病态分布,数据参数分布较为集中,极不均衡。这也是应用机器学习训练代理模型预报螺旋桨水动力性能过程中不可忽视的痛点。

当然,小样本和不平衡样本的问题也普遍存在于机器学习应用的各个领域,当前研究主要集中在图像识别领域,除了结合传统的图像处理方法<sup>[7]</sup>外,人们还通过特征增强、深度生成模型<sup>[8-9]</sup>等方法解决缺少样本的问题。但图像识别领域的样本生成方法难以应用到船舶螺旋桨性能领域。同时,我们可以看到有学者已经提出将多项式面积最小二乘法 (polynomial area least-squares, PALS) 应用于回归分析,可以有效降低样本数据对回归多项式的影响<sup>[10]</sup>。而在船舶螺旋桨性能方面,通过前期研究已经形成了丰富的经验知识,尤其是船舶螺旋桨水动力性能的研究中有很多经验公式<sup>[11-12]</sup>,这些都为机器学习训练代理模型提供了良好的基础。

本文以病态分布的船舶螺旋桨水动力性能样本为研究对象,提出一种基于经验知识的螺旋桨水动力样本扩充方法,同时在船舶螺旋桨水动力性能的代理预报模型中进行应用,给出经过扩充方法改良后代理模型的预报结果,并与扩充前以及经验公式结果作比较分析。

## 1 船舶螺旋桨敞水性能数据及代理模型存在的问题

### 1.1 敞水性能数据分布

通过收集 B 系列螺旋桨、MAU 系列螺旋桨以及中国船舶科学研究中心自研桨的敞水性能试验数据,形成了 412 个螺旋桨敞水性能数据样本。螺旋桨参数包括桨叶数  $Z$ 、直径  $D$ 、盘面比、 $0.7R$  处的螺距比、侧斜和纵倾  $Z_r$  等,水动力性能参数为进速系数  $J_0$ 、推力系数  $K_T$ 、转矩系数  $K_Q$ 。

敞水进速系数:

$$J_0 = \frac{v_0}{nD} \quad (1)$$

敞水推力系数:

$$K_T = \frac{T}{\rho n^2 D^4} \quad (2)$$

敞水转矩系数:

$$K_Q = \frac{Q}{\rho n^2 D^5} \quad (3)$$

由以上无因次系数可以进一步计算得到螺旋桨的敞水效率  $\eta_0$  :

$$\eta_0 = \frac{K_T}{K_Q} \cdot \frac{J_0}{2\pi} \tag{4}$$

以  $J_0$  为横坐标,  $K_T$ 、 $10K_Q$  及  $\eta_0$  为纵坐标绘制曲线, 可以得到螺旋桨敞水性征曲线。图1为中国船舶科学研究中心某自研桨敞水性能曲线。

图2为数据集中桨叶数、纵倾与直径比值以及侧斜的分布情况。

图3为数据集对应盘面比和螺距比的分布情况。图中散点图表示盘面比和螺距比联合分布的核密度, 两侧坐标轴分别为样本在盘面比与螺距比上的频率分布直方图。从图中可以看出, 数据样本点聚

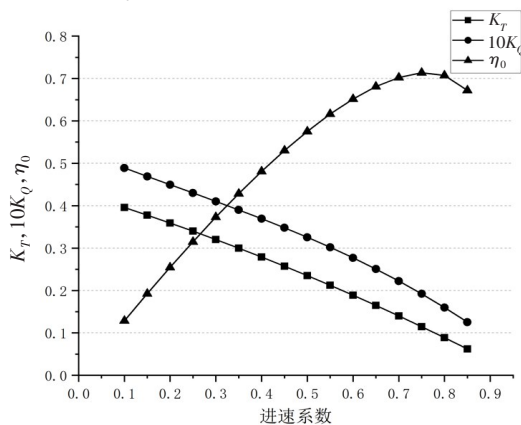
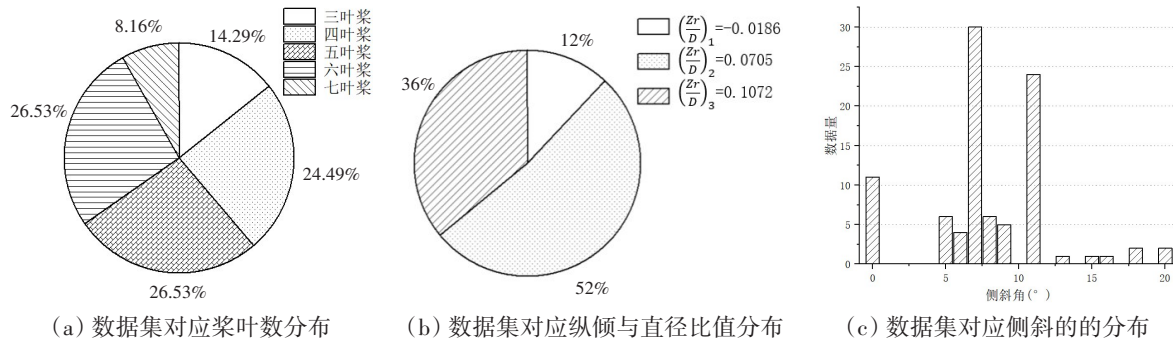


图1 螺旋桨敞水性能曲线

Fig.1 Open water performance of a propeller



(a) 数据集对应桨叶数分布

(b) 数据集对应纵倾与直径比值分布

(c) 数据集对应侧斜的分布

图2 数据集部分参数分布情况

Fig.2 Part of the parameter distribution of the dataset

集在图3的中间区域, 对应盘面比为0.5~0.7和螺距比为0.8~1.0的范围, 属于典型的病态分布。

### 1.2 代理模型存在的问题

将数据集按照7:3的比例划分为训练集和验证集, 分别是288条训练集数据和124条验证集数据。训练集数据将作为机器学习训练代理模型, 而验证集数据则留作验证用。

本文使用的代理模型为随机森林<sup>[13]</sup>。随机森林算法可以有效地映射非线性关系, 是一种兼具准确性与稳定性的预报模型, 能够有效地处理不平衡数据, 在数值预报领域有广泛的应用<sup>[14-15]</sup>。使用训练集进行螺旋桨敞水性能预报模型的训练, 并使用验证集进行测试。输入特征为螺旋桨的桨叶数、盘面比、螺距比和进速系数。预报目标为推力系数与转矩系数。模型在验证集上的平均误差随盘面比以及螺距比分布的结果如图4所示。使用该模型预报验证集中某螺旋桨敞水性能, 其结果详见图5。

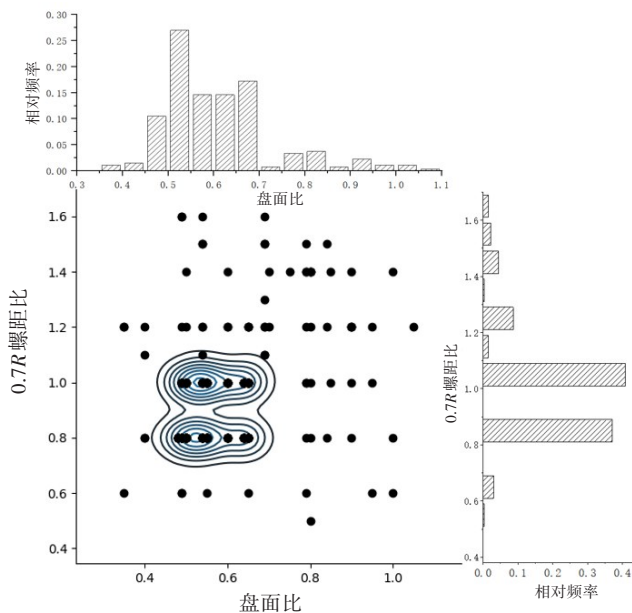


图3 训练集数据随盘面比、螺距比分布情况

Fig.3 Distribution of training data with area ratio and pitch ratio

模型在验证集上的平均误差随盘面比以及螺距比分布的结果如图4所示。使用该模型预报验证集中某螺旋桨敞水性能, 其结果详见图5。

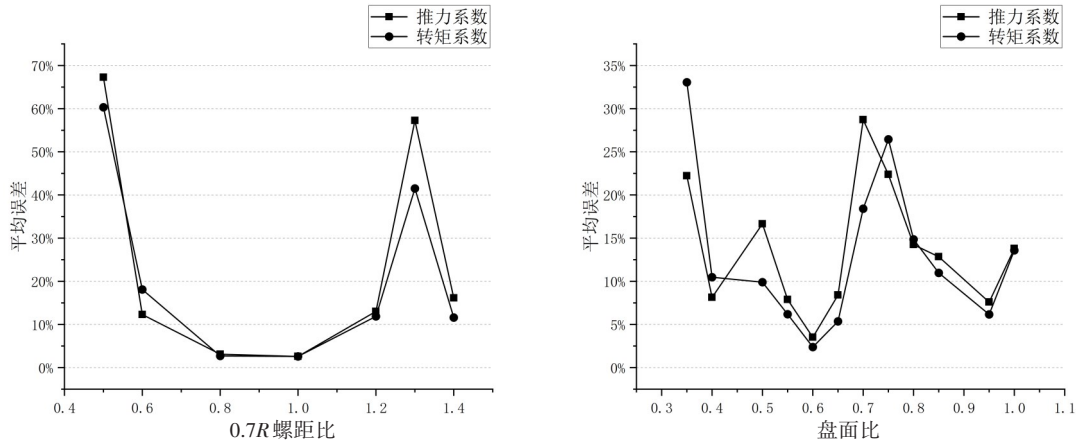


图4 代理模型预报误差

Fig.4 Prediction error of surrogate model

观察图4可知,当盘面比为0.5~0.7以及螺距比为0.8~1时,平均误差与最大误差远小于其他情况。图5中,代理模型预报误差远大于5%,与试验结果偏差较大。由以上结果可知,数据样本的不均匀分布明显影响了代理模型预报的结果。数据样本分布集中的区间,代理模型的预报精度以及稳定性较好;训练样本分布不充分时,代理模型预报结果并不可靠。

## 2 船舶螺旋桨水动力性能的经验知识

### 2.1 船舶螺旋桨敞水性能经验公式

对于B系列螺旋桨,推力系数和转矩系数可以使用多项式来表示<sup>[16-17]</sup>,详见式(5)~(6)。

$$K_T = \sum_{n=1}^{39} C_{T_n} J_0^{s_n} \left(\frac{P}{D}\right)^{t_n} \left(\frac{A_E}{A_0}\right)^{u_n} Z^{v_n} \quad (5)$$

$$K_Q = \sum_{n=1}^{39} C_{Q_n} J_0^{s_n} \left(\frac{P}{D}\right)^{t_n} \left(\frac{A_E}{A_0}\right)^{u_n} Z^{v_n} \quad (6)$$

式中, $P/D$ 为螺距比, $A_E/A_0$ 为盘面比, $Z$ 为叶数, $C_{T_n}$ 、 $C_{Q_n}$ 、 $s_n$ 、 $t_n$ 、 $u_n$ 、 $v_n$ 均为常数。

### 2.2 经验公式适用性修改及精度分析

上述公式是基于B系列螺旋桨的敞水试验数据得到的,但现有数据来源还包括MAU系列螺旋桨以及中国船舶科学研究中心的自研桨,需要检验该公式的适用性。使用训练集数据计算经验公式结果与实验结果的误差以及相关系数,部分验算结果如图6所示,经验公式计算结果与大部分验证集样本试验结果吻合较好,但与一部分螺旋桨样本相差较大。但是相关系数计算结果(表1)表明,其计算结果与试验结果相关性高。综合考虑,该经验公式可以适用于后续的本体扩充。

此外,使用经验公式在验证集上计算敞水性能并统计计算误差对于螺距比及盘面比的分布情况,具体结果及与代理模型的预报结果对比详见图7。我们可以看出,当螺距比为0.8~1时,代理模型的预报平均误差略好于经验公式,为防止引入更多误差,故不在此范围内使用经验公式进行样本扩充。

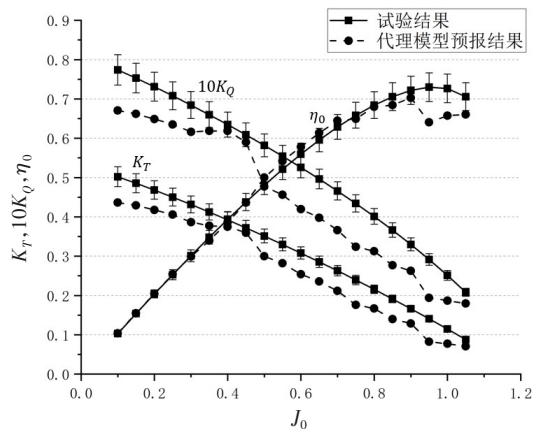


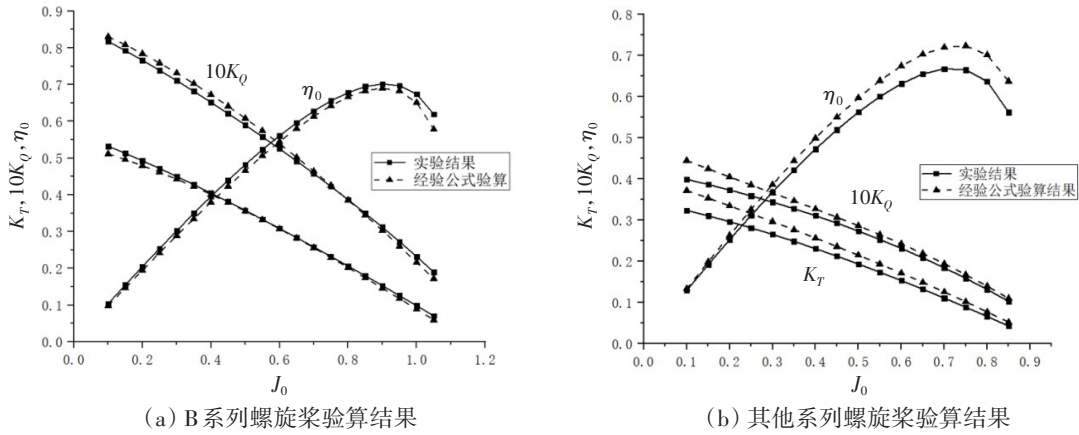
图5 验证集某螺旋桨预报结果

Fig.5 Prediction results of a propeller in the validation set

表1 扩充方程相关系数

Tab.1 Correlation coefficient of extended equation

$K_T$ 相关系数	$K_Q$ 相关系数
0.9962	0.9975



(a) B系列螺旋桨验算结果

(b) 其他系列螺旋桨验算结果

图6 部分训练集验算结果

Fig.6 Checking results of some training sets

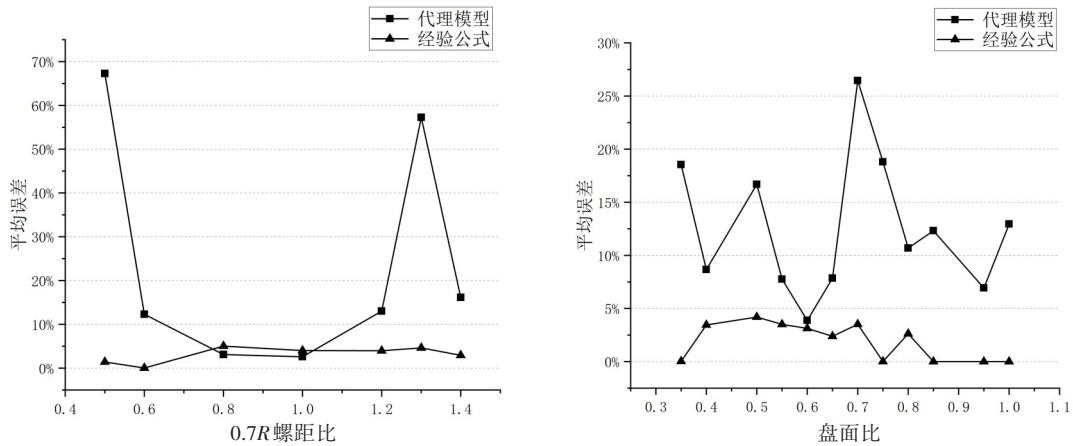


图7 经验公式与代理模型平均误差情况对比

Fig.7 Comparison of average error between empirical formula and surrogate model

### 3 基于经验知识的螺旋桨水动力样本数据扩充

#### 3.1 基于经验知识的样本扩充方法

为了利用这些经验公式背后的试验数据,现提出一种基于经验公式的螺旋桨样本扩充方法,流程图详见图8。首先根据扩充的目标样本选择相应的经验公式,然后结合现有数据集进行适用性分析,如果适用则进入下一步,否则对经验公式进行修改,如增加经验公式变量、加入修正项等,直至经验公式适用。确定适用后,根据现有数据、专家知识以及工程实际情况等确定样本扩充范围以及约束条件,剔除不满足约束条件的扩充数据,最终获得扩充样本。

#### 3.2 螺旋桨水动力样本数据扩充

本文选取的经验公式与适用性检验过程在前文已介绍,不再赘述。下面确定样本扩充范围,为了符合实际情况,样本扩充时螺旋桨桨叶数取值为3、4、5、6、7;同时,原数据集随进速系数分布较均匀,故不对进速系数进行扩充。针对数据集缺少的盘面

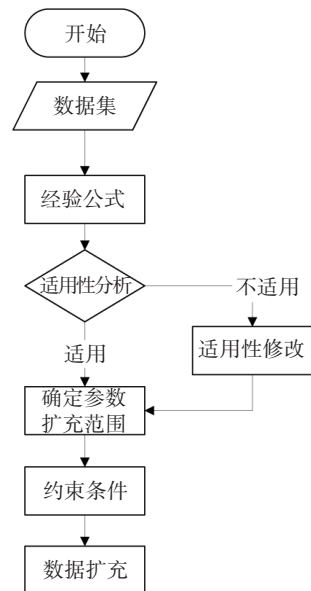


图8 基于经验公式的样本扩充方法流程

Fig.8 Process of sample expansion method based on empirical formula

比、螺距比进行补充,具体的扩充范围如表2所示。为了防止扩充数据与真实数据偏差过大,取原数据集中 $K_T$ 和 $K_Q$ 极值并浮动10%作为约束条件,具体的取值范围如式(7)~(8)所示,超出该范围的扩充样本会被剔除。

$$0.047 < K_T < 0.643 \tag{7}$$

$$0.081 < 10K_Q < 1.322 \tag{8}$$

表2 螺旋桨数据样本扩充范围

Tab.2 Extended range of propeller data samples

桨叶数	扩充的盘面比	扩充的螺距比	进速系数
3	0.4,0.45,0.7,0.75,0.8,0.9	0.7,1.1,1.2,1.3	0.1~1.4
4	0.6,0.65,0.75,0.8,0.9	0.55,0.65,0.95	0.1~1.4
5	0.55,0.7,0.8,0.85,0.9,1,1.1	0.55,0.65,0.7,0.85	0.1~1.66
6	0.6,0.75,0.8,0.9	0.7,1.1,1.2	0.1~1.66
7	0.6,0.65,0.75,0.8	0.7,1.1	0.1~1.35

经过样本扩充,共获得3078条螺旋桨敞水性能数据样本扩充数据。将扩充样本与原样本混合,形成新的训练集,绘制此时数据集的分布情况,详见图9。对比原测试集,扩充了盘面比为0.6~1.0处样本数据和螺距比0.6~0.7以及1.0~1.5处的数据样本,总体分布更为健康。从核密度图像来看,扩充以后的峰值区域明显增加,同时数据较少的边缘区域数据扩充量并未过多,以防止扩充过程中的误差累计。

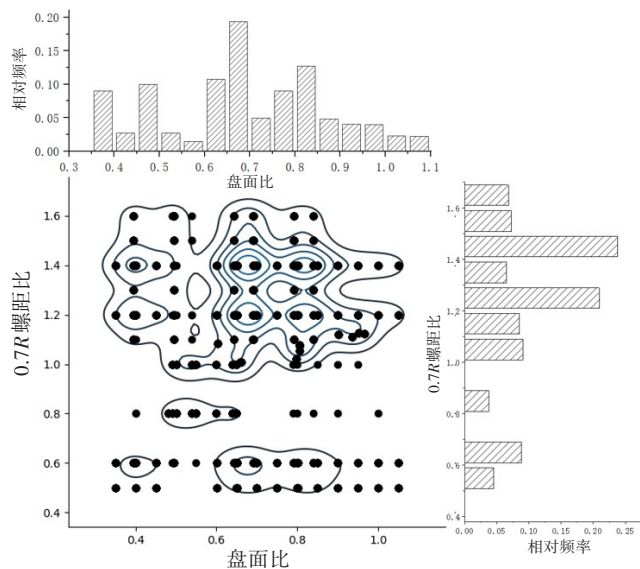


图9 扩充后训练集数据随盘面比及螺距比分布

Fig.9 Distribution of expanded training set with area ratio and pitch ratio

### 4 螺旋桨水动力性能样本数据扩充的应用

使用扩充后的数据集进行螺旋桨敞水性能预报的代理模型训练。保持代理模型、输入特征、预测目标不变,并在相同的验证集上验证预报效果。图10中展示的是随机森林的模型内部结构和主要超参数。使用网格搜索方法<sup>[18]</sup>进行模型超参数调优,调参结果详见图10。

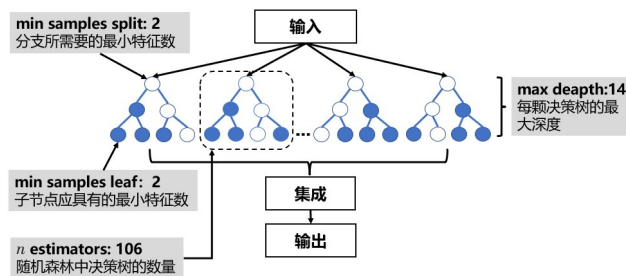


图10 随机森林结构及超参数调整结果

Fig.10 Random forest structure and super parameter adjustment results

利用验证集的数据对扩充后代理模型预报精度进行验证。图11为扩充后代理模型预报结果与扩充前代理模型预报结果的预报误差对比图。从图11可以发现,扩充前后预报误差在原误差较小的区

域基本不变,在原误差较大的区域预报误差显著降低。图12为用扩充后代理模型预报的某螺旋桨水动力性能曲线,其预报结果与实验结果偏差为5%,较扩充前有显著提升。对于单个螺旋桨,在我们关心的工况区间( $J_0=0.4\sim 0.8$ ),扩充后的代理模型预报结果更加吻合试验结果。综上所述,基于经验知识的样本扩充方法可以有效缓解训练样本对预报代理模型的不良影响,且样本扩充的过程中没有累计额外误差。

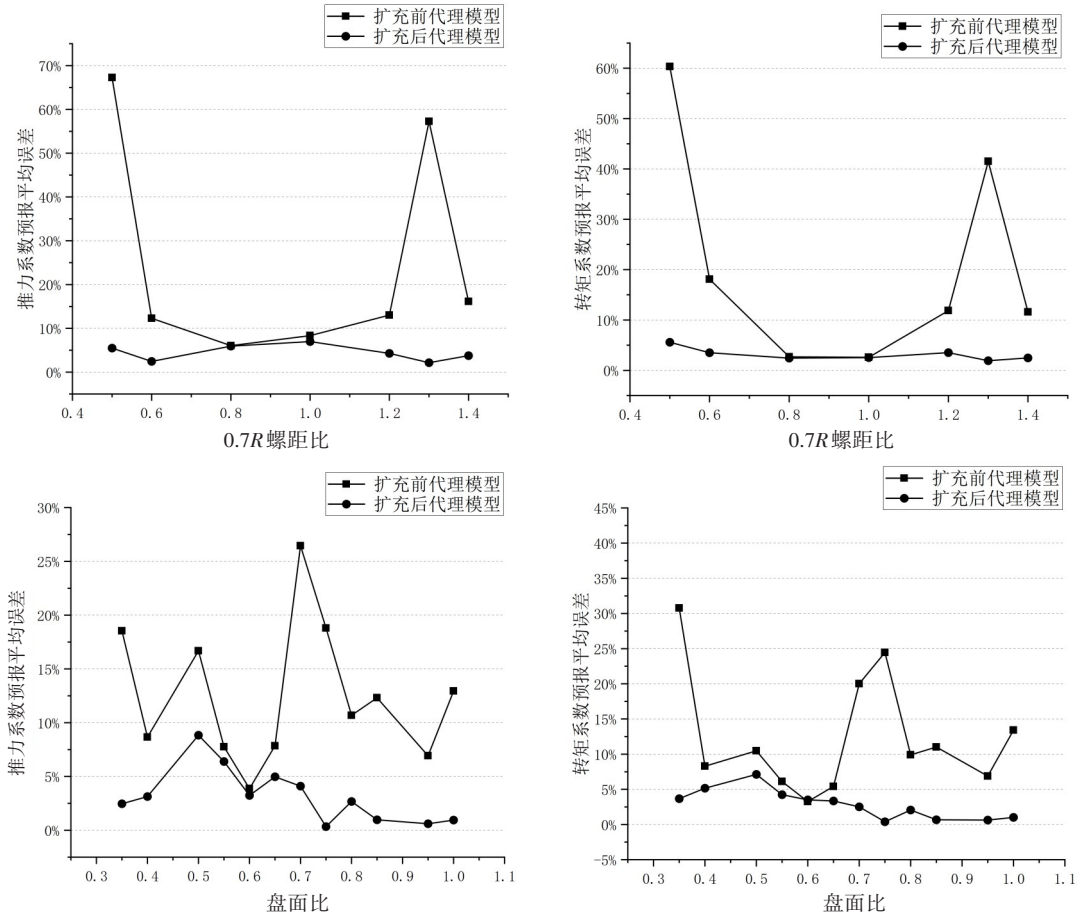


图11 扩充前后代理模型预报误差

Fig.11 Prediction error of surrogate model before and after expansion

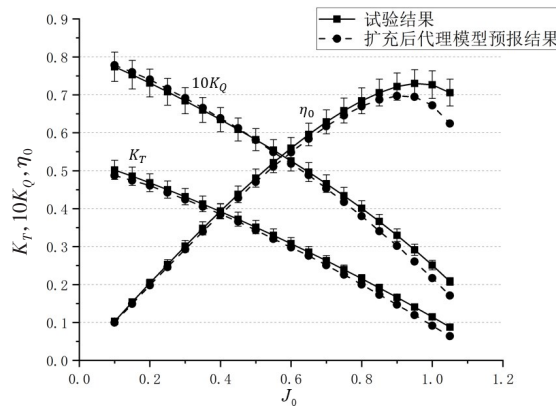


图12 扩充后代理模型预报的螺旋桨敞水性能曲线

Fig.12 Comparison of propeller open water performance predicted by surrogate model and experimental data

值得注意的是,在螺距比为0.8~1区域内,按照前文的方法进行数据扩充、代理模型的训练和预报,预报误差见图13。可以看到,扩充后的预报误差相较于未扩充时反而略微变差。主要原因是样本

扩充时带来了一定偏差,因此在后续工作中需注意检查经验公式的预报误差。在经验公式比扩充前代理模型预报误差大时,谨慎采用经验公式扩充样本。

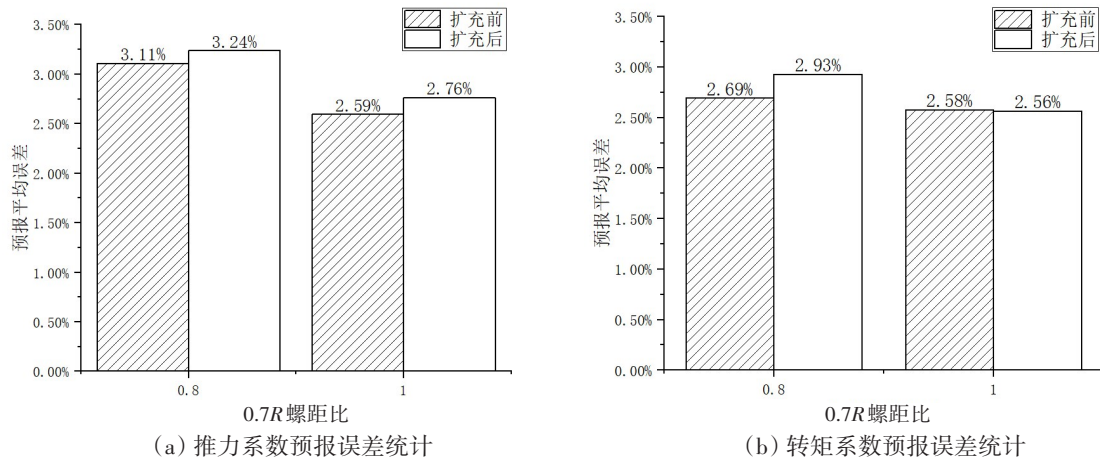


图 13 扩充前后螺旋桨水动力预报误差对比

Fig.13 Comparison of average errors of open water performance prediction before and after expansion

## 5 结 论

本文基于经验知识,建立了船舶螺旋桨水动力性能数据扩充方法,计算并分析了扩充后船舶螺旋桨水动力性能代理模型的预报精度,得到如下结论:

(1) 本文提出的基于经验公式的样本扩充方法是一种行之有效的方法。它充分利用了经验公式后的隐藏试验数据,扩充数据与真实数据相关性高,可以有效补充数据样本,缓解数据的病态分布问题,后续试验也证明了扩充样本的有效性;

(2) 使用代理模型进行螺旋桨水动力性能预报时,会受到病态数据集的不良影响。使用本文中的样本扩充方法对数据集进行针对性的数据增强,可以缓解数据分布不均衡带来的影响,降低代理模型在数据稀疏区域的误差,有效提高代理模型的准确率与稳定性;

(3) 经验公式的选取和确定是本文发展数据扩充方法的关键,需注意检查公式的预报误差。在经验公式比扩充前代理模型预报误差大时,不能采用经验公式扩充样本或需要对经验公式进行改进。

## 参 考 文 献:

- [1] 王晓强, 龚正琦. 水下螺旋桨技术发展现状与展望[J]. 中国水运, 2021, 4: 74-76.  
Wang Xiaoqiang, Gong Zhengqi. Development status and prospect of underwater propeller technology[J]. China Water Transportation, 2021, 4: 74-76. (in Chinese)
- [2] 苏玉民, 黄 胜. 用面元法预报船舶螺旋桨的水动力性能[J]. 哈尔滨工程大学学报, 2001, 22(2): 1-5.  
Su Yumin, Huang Sheng. Prediction of hydrodynamic performance of ship propeller by panel method[J]. Journal of Harbin Engineering University, 2001, 22(2): 1-5. (in Chinese)
- [3] 丁文文, 李 涛. 全结构网络技术在螺旋桨水动力性能预报中的应用[J]. 舰船科学技术, 2019, 41(2A): 10-12.  
Ding Wenwen, Li Tao. Application of full structure network technology in propeller hydrodynamic performance prediction [J]. Ship Science and Technology, 2019, 41(2A): 10-12. (in Chinese)
- [4] 李 亮, 刘登成, 郑巢生, 周 斌. 螺旋桨水动力性能预报自动化程序开发及试验验证[C]//第三十届全国水动力学研讨会暨第十五届全国水动力学学术会议论文集, 2019.  
Li Liang, Liu Dengcheng, Zheng Chaosheng, Zhou Bin. Automatic program development and experimental verification of

- propeller hydrodynamic performance prediction[C]//Proceedings of the 30th National Hydrodynamic Symposium and the 15th National Hydrodynamic Academic Conference, 2019. (in Chinese)
- [5] 王波, 张以良, 熊鹰. 船用螺旋桨水动力性能优化设计[J]. 武汉理工大学学报(交通科学与工程版), 2015, 39(04): 773-777.  
Wang Bo, Zhang Yiliang, Xiong Ying. Optimization design of hydrodynamic performance of marine propeller[J]. Journal of Wuhan University of Technology (Traffic Science and Engineering Edition), 2015, 39(04): 773-777. (in Chinese)
- [6] 曾志波, 丁恩宝, 唐登海. 基于BP神经网络和遗传算法的船舶螺旋桨优化设计[J]. 船舶力学, 2010, 14(Z1): 20-27.  
Zeng Zhibo, Ding enbao, Tang Denghai. Optimization design of ship propeller based on BP artificial neural network and genetic algorithm[J]. Journal of Ship Mechanics, 2010, 14(Z1): 20-27. (in Chinese)
- [7] 邱敏娜. 基于样本扩充的小样本人脸识别研究[D]. 哈尔滨: 哈尔滨工业大学, 2014.  
Qiu Minna. Research on small sample face recognition based on sample expansion[D]. Harbin: Harbin Institute of Technology, 2014. (in Chinese)
- [8] Frid-Adar M, Diamant I, Klang E, et al. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification[J]. Neurocomputing, 2018, 321: 321-331.
- [9] 何佳美. 基于生成对抗网络的电力设备图像扩充模型及算法研究[D]. 成都: 电子科技大学, 2020.  
He Jiamei. Research on image expansion model and algorithm of power equipment based on generation countermeasure network[D]. Chengdu: University of Electronic Science and Technology, 2020. (in Chinese)
- [10] Araujo A. Polynomial regression with reduced over-fitting—The PALS technique[J]. Measurement, 2018, 124: 515-521.
- [11] Radoji D, Simi A, Kalajdi M. Fifty years of the Gawn-Burrill KCA propeller series[J]. The International Journal of Small Craft Technology, 2009, 151(Part B2): 9-17.
- [12] Ito M, Yamasaki S, Oku M. Polynomial representation of AU-CP propeller open characteristics and its application to propeller initial design[J]. Journal of the Kansai Society of Naval Architects Japan, 1981: 25-33.
- [13] Cutler A, Cutler D R, Stevens J R. Random forests[J]. Machine Learning, 2004, 45(1): 157-176.
- [14] Janitza S, Tutz G, Boulesteix A L. Random forest for ordinal responses: Prediction and variable selection[J]. Computational Statistics & Data Analysis, 2016, 96: 57-73.
- [15] Xu W, Zhang J, Zhang Q, et al. Risk prediction of Type II diabetes based on random forest model[C]//2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB). IEEE, 2017.
- [16] Molland A F. The maritime engineering reference book: A guide to ship design, construction and operation[M]. Elsevier, 2011.
- [17] 吴小平, 刘洋浩, 张磊. 基于遗传算法的船舶螺旋桨优化设计[J]. 船舶与海洋工程, 2014(4): 31-37.  
Wu Xiaoping, Liu Yanghao, Zhang Lei. Optimization design of ship propeller based on genetic algorithm[J]. Ship and Ocean Engineering, 2014(4): 31-37. (in Chinese)
- [18] Lameski P, Zdravovski E, Mingov R, et al. SVM parameter tuning with grid search and its impact on reduction of model over-fitting[M]. Springer, 2015.