

基于机器学习的内孤立波波要素关系研究

李志鑫, 王 晶, 张 猛

(中国海洋大学 物理与光电工程学院, 山东 青岛 266100)

摘要: 内孤立波在海洋中的传播会携带能量和动量, 不同振幅的内孤立波对海洋中的能量交换及海上工程等影响也不同, 因此, 研究内孤立波振幅与半波宽度、水深、分层条件、密度等水文特征参量之间的关系显得尤为重要。以往在研究中建立内孤立波振幅与它们之间的关系时, 会受到不同理论有效适用范围的限制。本文借助实验室的水槽方法, 设计了不同的水深、分层及密度条件下的内孤立波系列综合实验, 发现内孤立波的振幅与半波宽度、水深、分层条件以及水体密度等参量之间并非简单线性关系。因此, 利用机器学习的方法建立内孤立波振幅与上述参量之间的非线性关系, 建立了支持向量机(SVM)和随机森林(RF)两种机器学习模型。将 1 266 组实验数据建立样本库, 其中包含训练集 970 组, 测试集 296 组, 对模型进行参数调优, 最终通过测试集验证, SVM 模型的平均相对误差为 17.3%, RF 模型的平均相对误差为 15.5%。该方法适用于多种不同的水文条件, 有效解决先前理论存在的适用性问题。

关键词: 内孤立波; 振幅; 水槽实验; 支持向量机; 随机森林

中图分类号: P733.1 文献标识码: A 文章编号: 1000-3096(2021)05-0113-08

DOI: 10.11759/hyxx20201105008

在过去几十年里现场实测和遥感观测的结果表明, 内孤立波是海洋中的一种普遍现象^[1], 通常由潮汐流过岛屿、海底山脊或大陆架造成密度跃层的重力塌陷产生^[2]。内孤立波在海洋中可以传播数百公里并且结构不会发生明显变化, 它们的传播伴随着海洋中巨大的能量传输^[3], 对洋流测量、海上工程、海洋航行等产生重要影响^[4-5]。因此, 正确理解内孤立波的物理特性至关重要。

弱非线性的 Korteweg - de Vries(KdV)方程^[6]是描述内孤立波传播的最简单的模型, 由于其简单的形式而被广泛使用, 但其仅适用于小振幅内孤立波。Koop 等^[7]对两层流体系统中有限振幅内孤立波的实验进行研究, 表明浅水中 KdV 理论与实验结果一致, 当水深条件改变, 理论会和实验结果发生偏差, Grue 等人^[8]在实验中固定水体分层比, 通过改变振幅进行实验, 发现 KdV 理论适用的有效振幅范围很小, 之后的大部分实验表明 KdV 理论仅适用于小振幅内孤立波的情况, 当振幅与流体厚度之比超过 0.1 时, KdV 理论不再适用^[9-10]。eKdV 理论作为 KdV 理论的一个扩展, 可以更好地描述中等振幅的内孤立波^[11], 对于相对较小的振幅, eKdV 与 KdV 之间的解存在显著差异^[1], 对于大振幅内孤立波, eKdV 理论在上下两层流体厚度差异较小时较为准确。为描述更大振幅的内孤立波, Miyata^[12]

与 Choi 等^[13]对弱非线性两层 eKdV 模型进行了扩展, 给出了完全非线性的内孤立波的解析解, 即 Miyata - Choi - Camassa(MCC)理论, 研究表明, 对于大振幅的内孤立波, MCC 理论的解与实验室结果一致^[14], 不同于 eKdV 的是, MCC 理论在两层流体厚度差异明显时较为准确。Du 等人^[15]通过实验得出特征频率(相速度/内孤立波特征波长)与内孤立波振幅之间的关系是选择以上哪种理论的最佳标准, 而特征频率又与上下层水体的厚度及密度有关。适用于不同条件下的其他理论也大多是在 KdV 理论的基础上进行扩展而得到的, 由此可见, 内孤立波振幅与上下层水体的厚度、密度及内孤立波的波形之间存在复杂的关系。

综上所述, 在建立内孤立波振幅与其他参量之间的关系时, 要充分考虑各种理论的适用条件, 由于每种理论只能很好地在各自的有效范围内对内孤立波进

收稿日期: 2020-11-05; 修回日期: 2021-02-03

基金项目: 国家重点研发项目(2017YFC1405600); 国家自然科学基金(61871353)

[Foundation: National Key Research and Development Project, No. 2017YFC1405600; National Natural Science Foundation of China, No. 61871353]

作者简介: 李志鑫(1996—), 男, 山东青岛人, 硕士研究生, 研究方向为海洋内波, E-mail: lizhixin_ouc@163.com; 王晶(1962—), 通信作者, 教授, 主要从事光纤光学和海洋遥感研究, E-mail: wjing@ouc.edu.cn

行描述,超出有效范围则会产生系统偏差,因此,通过各种理论给出的解析解来描述内孤立波振幅与半波宽度及水深、密度等水文参量之间的关系时,其准确性会受到不同条件的限制。为解决该问题,本文通过在实验室水槽中设计不同的实验,对内孤立波的振幅、半波宽度、分层和密度差等进行提取并生成数据库,利用两种大数据模型有效地建立内孤立波振幅与其他参量之间的关系,并且该方法可适用于多种水文条件。

1 实验设计

两层流体模型得益于考虑到密度跃层的有限厚度,给实验数据和现有理论提供了很好的一致性,实验中常用两层流体模型对内孤立波参数进行估计^[16]。实验分别在两个不同规格的长直玻璃水槽中进行,其中一个水槽长 3.0 m,宽 0.15 m,高 0.3 m,另一个水槽长 5.0 m,宽 0.35 m,高 0.8 m。在水槽中构建一个由两层流体组成的系统,将提前配置好的相应密度的盐水抽入水槽中作为下层水体,通过在下层水体表面铺设气泡膜,用蠕动泵缓慢地将密度 $\rho_1 = 0.998 \text{ g/cm}^3$ 的淡水放入水槽中作为上层水体,上层与下层水体的高度分别为 h_1 、 h_2 ,密度分别为 ρ_1 、 ρ_2 ,总水深 $H = h_1 + h_2$,电荷耦合器(CCD)位于水槽侧面,并通过计算机控制以每秒 35 帧的采样频率拍摄垂向剖面上内孤立波的传播过程。

在水槽实验中,内孤立波通常利用分层流体的重力塌陷产生^[17],生成原理如图 1 所示,在水槽中垂直插入挡板至密度跃层下方并确保挡板与水槽底部留有部分空间,挡板左侧形成一个造波区,向造波区内缓慢注入一定量的淡水,下层盐水会通过挡板下方流向右侧以保持流体静力平衡,随上层淡水注入,造波区一侧的密度跃层下降,与另一侧形成高度差,可通过设置不同的塌陷高度,生成不同振幅的内孤立波。实验时快速平稳地抽离挡板,两侧流体由于存在重力势能差发生重力塌陷,从而产生垂直剪切运动,两层水体迅速混合产生顺时针方向的扰动,生成第一模态下降型的内孤立波向右传播,同时利用计算机控制 CCD 进行拍摄。

利用两种不同规格的水槽,共设置 25 组不同初始条件的实验,包括 10 种不同的总水深、15 种不同的分层条件以及 3 种不同的密度差,其中总水深 H 在 16 cm 到 68 cm 之间,上层水深与总水深的比值 h_1/H 在 0.074 到 0.375 之间,上层水体密度固定为 $\rho_1 = 0.998 \text{ g/cm}^3$,下层水体密度采用 1.041 g/cm^3 、 1.061 g/cm^3 和 1.081 g/cm^3 三种。

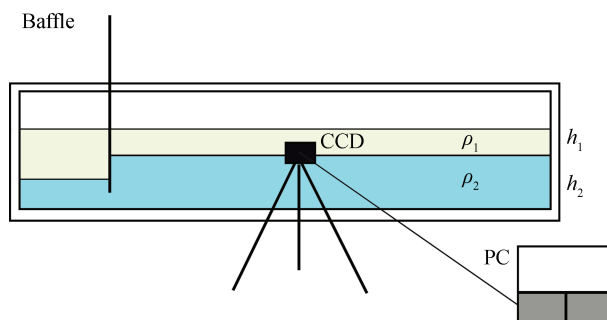


图 1 实验仪器示意图

Fig. 1 Schematic diagram of experimental apparatus

注: h_1 : 上层水深, h_2 : 下层水深, ρ_1 : 上层水体密度, ρ_2 : 下层水体密度, CCD: 电荷耦合器件, PC: 计算机, Baffle: 挡板

2 实验结果与分析

利用 CCD 拍摄内孤立波的传播过程,如图 2 所示,从图中提取内孤立波的振幅 A 和半波宽度 L ,在 KdV 方程的推导中,非线性系数作为一个尺度参数,意味着波振幅与总水深的比值 A/H ,其大小代表了振幅的大小^[18]。为研究内孤立波振幅与半波宽度及水深、分层条件和密度之间的关系,设置了若干组对照实验,改变水深、密度差和分层等初始实验条件,由于我们需要分析内孤立波振幅与半波宽度之间的关系,因此将半波宽度与总水深的比值 L/H 作为半波宽度的尺度。通过观察振幅与总水深的比值 A/H 和半波宽度与总水深的比值 L/H 之间关系的变化来反映振幅与其他水文特征参量之间的关系。

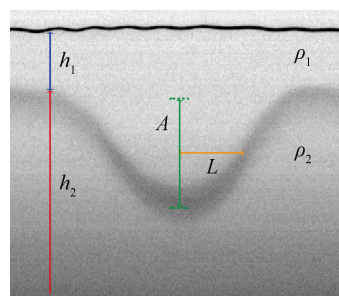


图 2 内孤立波垂向剖面的时间序列图

Fig. 2 Time series diagram of vertical profile of internal solitary wave

注: h_1 : 上层水深, h_2 : 下层水深, ρ_1 : 上层水体密度, ρ_2 : 下层水体密度, A : 内孤立波振幅, L : 内孤立波半波宽度。

2.1 不同分层条件下内孤立波振幅与半波宽度的关系

选取 3 组实验,保持总水深 $H = 24 \text{ cm}$,下层水

体密度 $\rho_2=1.081 \text{ g/cm}^3$, 研究在不同条件下内孤立波振幅与半波宽度的关系。图 3a、3b、3c 3 组实验中, 上层水深与总水深之比 h_1/H 分别为 0.125、0.208 和 0.292, 根据图 3 可以发现当分层条件改变时, 振幅与半波宽度的关系也随之发生改变, 不同分层条件下拟

合相关系数分别为 0.080 0、0.029 5 和 0.290 2, 线性相关性较低, 当分层比为 0.125 和 0.292 时, 拟合斜率为负值, 分层比为 0.208 时, 拟合斜率为正值, 两者相关性随条件的改变而改变。因此两者之间可能存在复杂的非线性关系, 这与 Du 等人的实验也吻合^[15]。

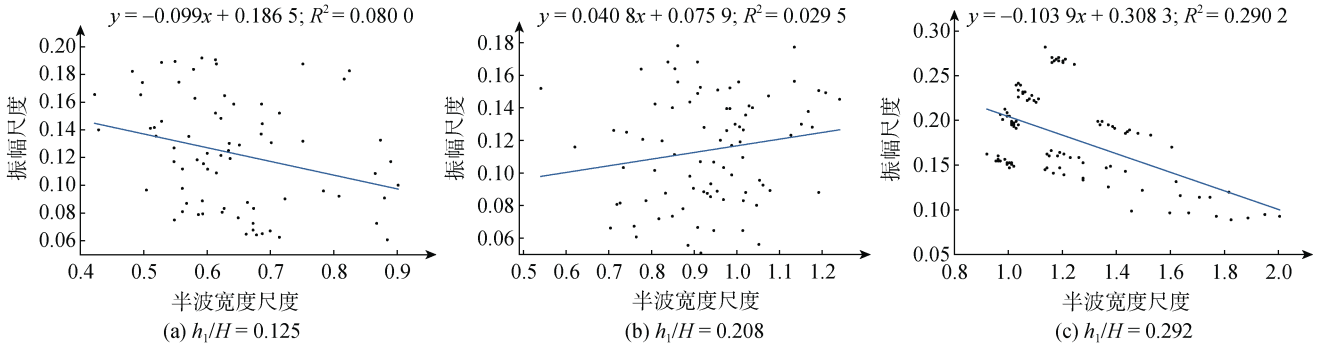


图 3 不同分层条件下内孤立波振幅与半波宽度散点图

Fig. 3 Scatter plots of amplitude and half-wave width of internal solitary waves under different stratified conditions

2.2 不同密度条件下内孤立波振幅与半波宽度的关系

为研究密度差对振幅和半波宽度之间关系的影响, 选取总水深 $H=22 \text{ cm}$, 上层水深与总水深之比 $h_1/H=0.227$ 的 3 组实验, 通过改变下层水体密度 ρ_2 , 对振幅与半波宽度的关系进行分析。如图 4 所示, a—c 三组实验的 ρ_2 分别为 1.041 g/cm^3 、 1.061 g/cm^3 和

1.081 g/cm^3 , 其相关系数分别为 0.031 3、0.076 1 和 0.184 3, 线性相关性很弱, 并且当下层水体密度为 1.041 g/cm^3 和 1.061 g/cm^3 , 振幅与半波宽度的拟合斜率为正值, 而当下层水体密度变为 1.081 g/cm^3 时, 两者拟合斜率变为负值。因此仅当密度差发生改变时, 内孤立波振幅与半波宽度之间的关系也会发生改变, 但不能认为是简单的线性关系。

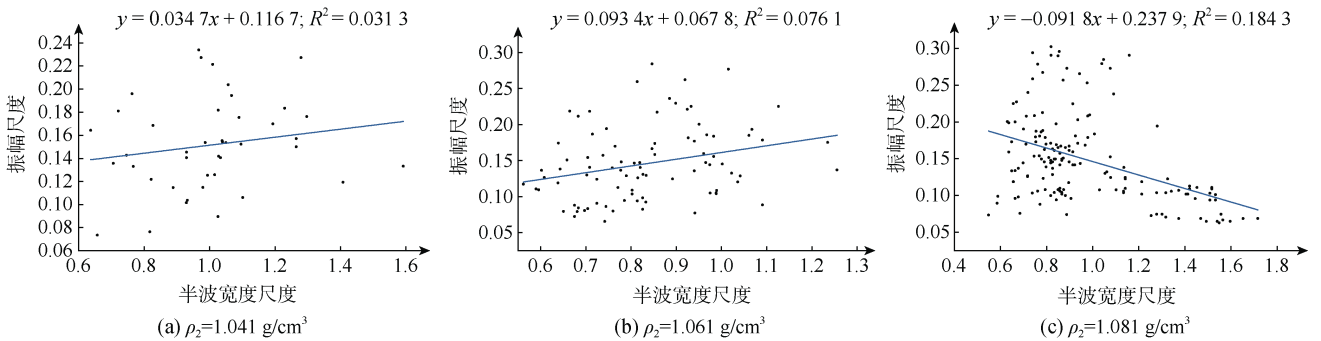


图 4 不同密度条件下内孤立波振幅与半波宽度散点图

Fig. 4 Scatter plots of amplitude and half-wave width of internal solitary waves under different density conditions

2.3 不同水深条件下内孤立波振幅与半波宽度的关系

保持下层水体密度 $\rho_2=1.081 \text{ g/cm}^3$, 上层水深与总水深之比 $h_1/H=0.227$ 不变, 研究总水深对内孤立波振幅和半波宽度之间关系的影响。图 5a、5b 分别为总水深 $H=22 \text{ cm}$ 和 $H=44 \text{ cm}$ 的实验, 对比两图可以发现, 两种不同的水深条件下, 内孤立波振幅

与半波宽度的线性相关系数都很低, 分别为 0.009 8 和 0.100 9, 该组对照实验中振幅与半波宽度之间的关系变化不明显。

2.4 全条件下内孤立波振幅与半波宽度的关系

对所有实验数据进行分析, 在全条件下, 振幅与半波宽度之间的关系如图 6 所示, 结合以上 3 组

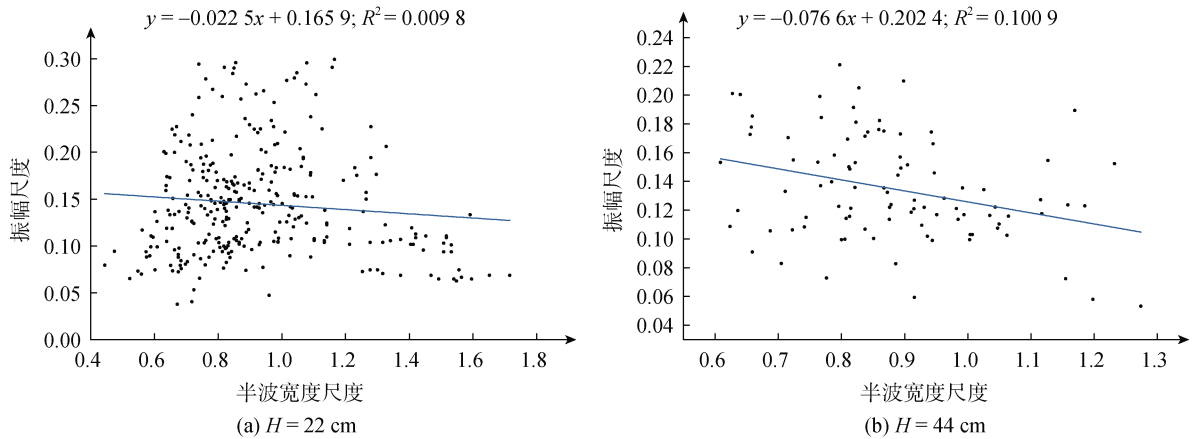


图 5 不同水深下内孤立波振幅与半波宽度散点图

Fig. 5 Scatter plots of amplitude and half-wave width of internal solitary waves in different water depths

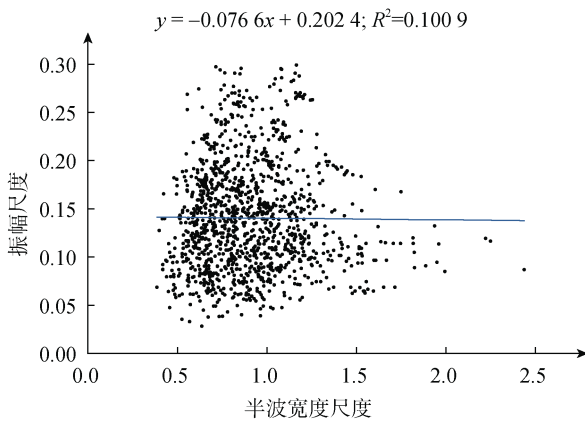


图 6 内孤立波振幅与半波宽度散点图

Fig. 6 Scatter plot of amplitude and half-wave width of internal solitary wave

对照实验,可以发现,振幅与半波宽度之间的线性相关性很低,改变某种水文条件,两者关系会发生变化,说明振幅与半波宽度之间存在的是一种复杂的非线性关系。针对不同初始条件下产生的内孤立波,传统的非线性理论给出了各自的稳态解,一旦超出适用的范围,会与实际波形产生偏差,对于各种水文条件都包含的情况,可以利用机器学习的方法建立内孤立波振幅与半波宽度及水文特征参量之间的非线性关系。

3 模型建立

目前,机器学习领域的研究取得巨大进展,Warren Mcculloch 和 Walter Pitts 在 1943 年提出了神经网络的结构模型^[19],为机器学习的发展奠定了重要基础。机器学习是从给定训练样本中寻找输入与输出之间

的最优关系,使其能够对无法观测的输出做出尽可能准确的预测^[20],20 世纪 90 年代以来,多种基于统计学习规律的浅层机器学习模型的出现,使机器学习的研究取得很大的进展^[21]。

3.1 机器学习模型

支持向量机(support vector machine)是 Cortes 和 Vapnik 在 1995 年首次提出的一种基于统计学习理论的有监督的机器学习方法^[22],广泛用于处理统计分类和回归分析问题,SVM 具有很好的鲁棒性和泛化能力,能够很好地建立数据之间的非线性关系。利用 SVM 处理回归问题的基本思想如图 7a 所示,在处理回归问题时,我们希望能够找到最优超平面,使得所有样本点与超平面之间的偏差最小,即让更多的数据落在边界内,当数据足够多时,就可以将超平面作为回归结果,为防止模型过拟合,通常采用 ϵ -不敏感损失函数(见图 7b)来界定超平面与边界的间隔距离,当模型结果与真值之间差值的绝对值大于 ϵ 才计算损失^[23]。在处理非线性回归问题时,由于低维空间无法进行处理,需要将样本空间的数据映射到高维特征空间(又称希尔伯特空间)进行处理。由于非线性 SVM 在升维过程中会使计算量增大,延长数据处理的时间,且容易引发维度爆炸,因此需要引入合适的核函数 $K(x, x_i)$ 使其在低维空间进行计算,将实际效果映射到高维空间中,避免了在高维空间对数据进行直接计算。SVM 中不同的核函数会构造出不同的高维特征空间,在处理此类问题时,通常采用径向基核函数(RBF),SVM 的模型结构如图 8a 所示。

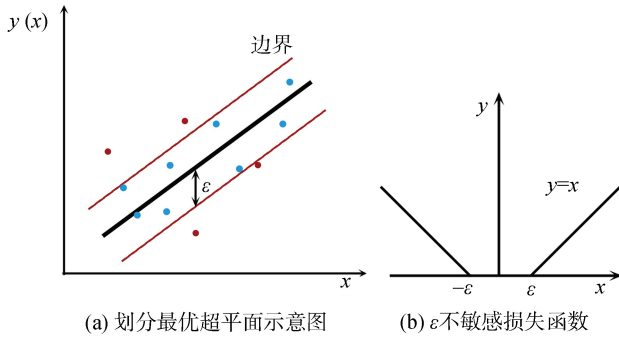
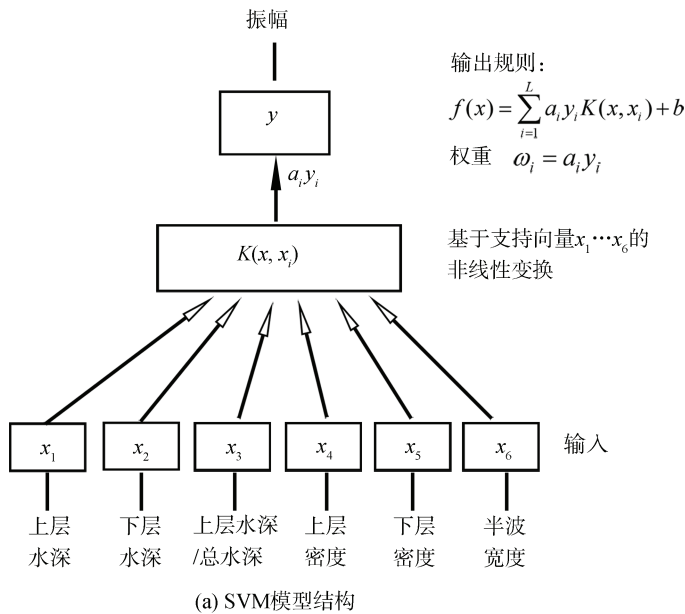


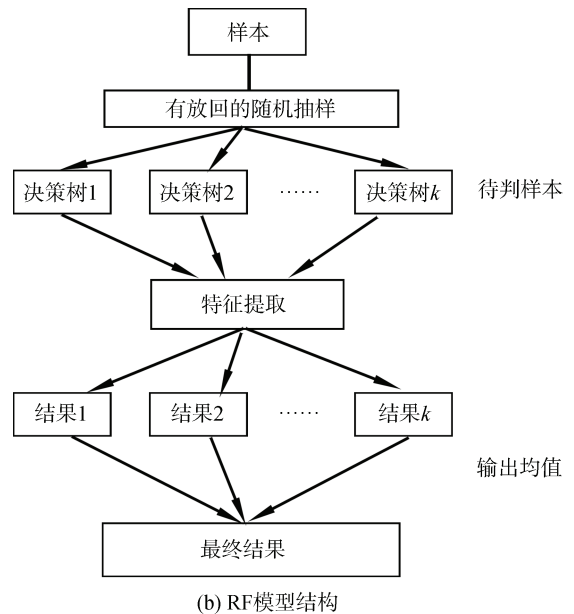
图 7 SVM 处理回归问题原理图

Fig. 7 Principle diagram of SVM dealing with regression problem

注: ϵ : 超平面与边界的间隔距离



(a) SVM模型结构



(b) RF模型结构

图 8 SVM 与 RF 模型结构图

Fig. 8 SVM and RF model structure diagram

注: x_1 : 上层水深, x_2 : 下层水深, x_3 : 上层水深与总水深的比值, x_4 : 上层水体密度, x_5 : 下层水体密度, x_6 : 内孤立波半波宽度, $K(x, x_i)$: 核函数, $a_i y_i$: 权重, y : 输出振幅值。

3.2 样本库

在物理海洋学理论中, 内孤立波的振幅与水深、密度及分层条件等都有关系, 因此我们选择上层密度 ρ_1 、下层密度 ρ_2 、上层水深 h_1 、下层水深 h_2 、上层水深占比 h_1/H 和半波宽度 L 作为 SVM 和 RF 的输入, 内孤立波的振幅 A 作为模型输出, 表 1 给出了模型中采用的输入输出数据。将不同组别的实验数据汇总建立样本库, 总共获得 1 266 组数据, 将数据集随机打乱划分为互斥的训练集和测试集, 训练集用来拟合模型, 测试集用来验证模型的精度。通常需要保证训练集数量在总数据集数量的 2/3 到 4/5 之间^[26],

随机森林(random forest, RF)是 Breiman 在 2001 年提出的一种基于分类树的机器学习算法^[24], 其基本原理是利用自助重抽样方法从原始样本中抽取多个样本^[25], 即多个决策树分别从训练集中抽取一部分样本, 不同决策树学习的样本是有差别的, 关注的特征也是不同的, 决策树之间没有关联, 体现了其样本和属性的双重随机性。这种方式也从一定程度上避免了过拟合, 每棵决策树都会输出一个结果, 在处理回归问题时, 最终结果由每棵决策树给出的均值综合而来, 是一种集成学习算法, 具有精确度高、调节参数少、训练速度快的优点, 是一种很好的非线性建模工具, 模型结构如图 8b 所示。

遵循该原则, 最终训练集分配 970 组数据, 测试集分配 296 组数据。

4 结果与分析

SVM 模型的精度主要受惩罚因子 C 与核函数参数 γ 影响, C 是用来调节正则化和经验风险部分之间的平衡^[27], 若选取不当会造成过拟合或欠拟合, γ 则决定了数据映射到新的特征空间中的分布, γ 越大, 则支持向量越少, 反之越多, 该参数会影响模型的训练与预测速度。由于这两个参数都会影响模型的精确度, 所以需要在这两个变量进行寻优, 常见的优化

算法有网格搜索(GS)、粒子群优化(PSO)和遗传算法(GA), 在先前的实验研究中发现 GS 方法较稳定,更适用于实验室研究^[28]。GS 就是尽可能地寻找 C 与 γ 的值来进行交叉验证, 寻找全局最优解, 本模型经

表 1 输入-输出设置

样本数量	上层水深/cm	下层水深/cm	上层水深占比	上层密度/(g·cm ⁻³)	下层密度/(g·cm ⁻³)	半波宽度/cm	振幅/cm
1	3	17	0.15	0.998	1.081	17.57	5.12
2	5	17	0.227	0.998	1.041	21.43	5
3	5	17	0.227	0.998	1.061	17.38	2.84
4	8.5	59.5	0.125	0.998	1.081	42.19	3.99
.....						
1266	5	39	0.113	0.998	1.081	34.17	7.5

过 GS 方法优化后的 $C=0.125$, $\gamma=45.2548$ 。利用训练好的 SVM 模型对测试集进行验证, 图 9a 为优化后的 SVM 在测试集上验证的结果, 总体与测试集数据吻合良好, 平均相对误差为 17.3%。

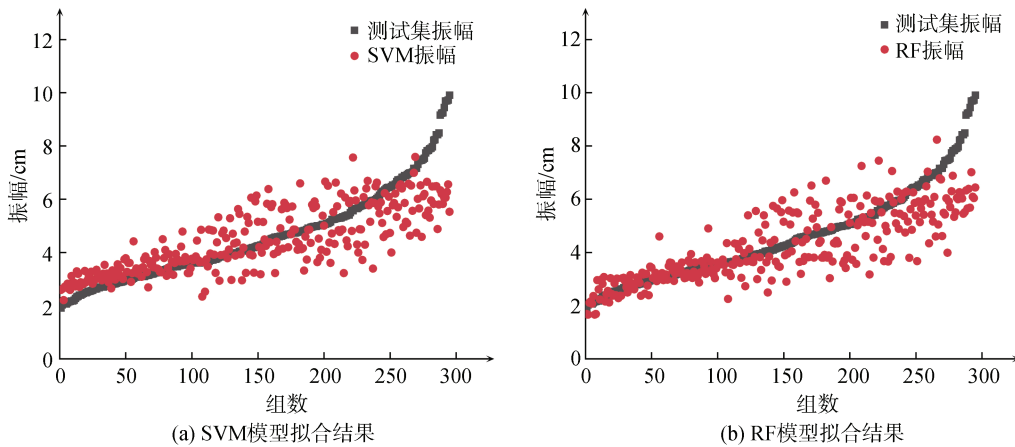


图 9 模型结果与测试集拟合图
Fig. 9 Fitting graph between model results and test set

RF 是一种基于多棵决策树进行优化决策的算法, 其参数调优主要包括两部分: 一是对 RF 框架的调优, 二是对 RF 决策树的调优。对于框架的调优一般考虑调节决策树的数量, 设置决策树数量范围为(0, 1 000], 并以步长为 10 进行个数寻优。关于 RF 决策树调优的重要参数是树的最大深度和最大特征值, 当样本数量和特征较多时, 可以对树的最大深度进行限制, 最大特征值可以提高每个子模型的拟合性能。本文设置树的最大深度范围为[1, 50], 步长为 1, 最大特征值范围设为(0, 1), 步长为 0.1。寻优之后, 树的数量为 211, 最大深度为 7, 最大特征值为 0.706, 寻优后的 RF 模型测试结果的平均相对误差为 15.5%。如图 9b 所示, 模型结果与测试集数据的吻合程度较好, 当振幅较小时, 模型拟合精度较高, 在振幅较大时也表现出优于 SVM 的能力。

5 结论

本文利用机器学习的方法, 以 SVM 和 RF 模型建立内孤立波振幅与上下层水深、密度、半波宽度之间的非线性关系, 解决了传统的非线性理论在实验室中描述内孤立波时存在的适用性问题, 为实验室中内孤立波振幅研究提供了一种新思路。

为获取足够的数据进行训练, 借助实验室的水槽方法, 设计了不同的水深、分层及密度条件下的内孤立波系列综合实验, 获取不同条件下内孤立波的振幅。对 CCD 拍摄的照片以时间序列图的形式输出, 提取内孤立波的振幅和半波宽度。设置对照实验分析实验数据, 内孤立波的振幅与半波宽度并非简单的线性关系, 其关系会随水深、分层条件以及密度差的变化而发生改变, 说明两者之间的非线性关系受

初始实验条件的影响。

为使内孤立波振幅与水深、分层条件等水文特征参量之间的关系适用于更广的范围,本文提出建立 SVM 模型和 RF 模型,以一种新的形式建立内孤立波振幅与水深、分层条件、密度、半波宽度之间的非线性关系。利用实验室获取的 1 266 组数据建立样本库,训练集分配 970 组,测试集分配 296 组。SVM 模型采用 GS 优化算法, $C=0.125$, $\gamma=45.2548$, 最终利用测试集对训练好的模型进行验证,平均相对误差为 17.3%。RF 模型通过对决策树个数、最大特征值及最大深度进行寻优,寻优之后,树的数量为 211,最大深度为 7,最大特征值为 0.706,最终平均相对误差为 15.5%,优于 SVM 模型。表明利用机器学习的方法建立内孤立波振幅与其他参量之间的非线性关系是较为准确和可行的,为扩大该模型的适用范围,提升模型精度,后续将设计更多的实验,扩充不同初始条件之下的数据量。

参考文献:

- [1] HELFRICH K R, MELVILLE W K. Long nonlinear internal waves[J]. Annual Review of Fluid Mechanics, 2006, 38(1): 395-425.
- [2] NAKAMURA T, AWAJI T. A growth mechanism for topographic internal waves generated by an oscillatory flow[J]. Journal of Physical Oceanography, 2001, 31(8): 2511-2524.
- [3] LAMB K G, FARMER D. Instabilities in an internal solitary-like wave on the oregon shelf[J]. Journal of Physical Oceanography, 2011, 41(1): 67-87.
- [4] OSBORNE A R, BURCH T L. Internal solitons in the andaman sea[J]. Science, 1980, 208(4443): 451-460.
- [5] CHEN L, ZHENG Q, XIONG X, et al. Dynamic and statistical features of internal solitary waves on the continental slope in the northern south china sea derived from mooring observations[J]. Journal of Geophysical Research: Oceans, 2019, 124(6): 4078-4097.
- [6] DJORDJEVIC V D, REDEKOPP L G. The fission and disintegration of internal solitary waves moving over two-dimensional topography[J]. Journal of Physical Oceanography, 1978, 8(6): 1016-1024.
- [7] KOOP C G, BUTLER G. An investigation of internal solitary waves in a two-fluid system[J]. Journal of Fluid Mechanics, 1981, 112: 225.
- [8] GRUE J, JENSEN A, RUSS P O, et al. Properties of large-amplitude internal waves[J]. Journal of Fluid Mechanics, 1999, 380: 257-278.
- [9] WALKER S A, MARTIN A J, EASSON W J. Comparison of laboratory and theoretical internal solitary wave kinematics[J]. Journal of Waterway, Port, Coastal, and Ocean Engineering, 2003, 129: 210-218.
- [10] HELFRICH K R, MELVILLE W K. On interfacial solitary waves over slowly varying topography[J]. Journal of Fluid Mechanics, 1984, 149: 305-317.
- [11] KAKUTANI T, YAMASAKI N. Solitary waves on a two-layer fluid[J]. Journal of the Physical Society of Japan, 2007, 45(2): 674-679.
- [12] MIYATA M. An internal solitary wave of large amplitude[J]. Deep Sea Research Part B Oceanographic Literature Review, 1985, 23(2): 43-48.
- [13] CHOI W, CAMASSA R. Fully nonlinear internal waves in a two-fluid system[J]. Journal of Fluid Mechanics, 1999, 396: 1-36.
- [14] MICHALLET H, BARTHÉLEMY E. Experimental study of interfacial solitary waves[J]. Journal of Fluid Mechanics, 1998, 366: 159-177.
- [15] DU H, WEI G, WANG S D, et al. Experimental study of elevation- and depression-type internal solitary waves generated by gravity collapse[J]. Physics of Fluids, 2019, 31(10): 102-104.
- [16] OSTROVSKY L A, STEPANYANTS Y A. Do internal solitons exist in the ocean?[J]. Reviews of Geophysics, 1989, 27(3): 293-310.
- [17] WU J. Mixed region collapse with internal wave generation in a density-stratified medium[J]. Journal of Fluid Mechanics, 2006, 35(3): 531-544.
- [18] QIAN M A, CHUNXIN Y, XIAOPEI L, et al. The investigation of internal solitary waves over a continental shelf-slope[J]. Journal of Oceanology and Limnology, 2020, 38(3): 695-706.
- [19] MCCULLOCH W S, PITTS W H. A logical calculus of the ideas immanent in nervous activity[J]. The Bulletin of Mathematical Biophysics, 1988, 5: 115-133.
- [20] 祁亨年. 支持向量机及其应用研究综述[J]. 计算机工程, 2004, 30(10): 6-9.
QI Hengnian. Support vector machines and application research overview[J]. Computer Engineering, 2004, 30(10): 6-9.
- [21] 张润, 王永滨. 机器学习及其算法和发展研究[J]. 中国传媒大学学报: 自然科学版, 2016, 23(2): 10-18.
ZHANG Run, WANG Yongbin. Research on machine learning with algorithm and development[J]. Journal of Communication University of China(Science and Technology), 2016, 23(2): 10-18.
- [22] CORTES C, VAPNIK V N. Support-vector networks[J]. Machine Learning, 1995, 20(3): 273-297.
- [23] 陈永义, 俞小鼎, 高学浩, 等. 处理非线性分类和回归问题的一种新方法 (I)——支持向量机方法简介[J].

- 应用气象学报, 2004, 15(3): 345-354.
CHEN Yongyi, YU Xiaoding, GAO Xuehao, et al. A new method to deal with nonlinear classification and regression problems (I) ——Introduction to support vector machine method[J]. Journal of Applied Meteorological Science, 2004, 15(3): 345-354.
- [24] BREIMAN L. Random Forests[J]. Machine Learning, 2001, 45(1): 5-32.
- [25] 方匡南, 吴见彬, 朱建平, 等. 随机森林方法研究综述[J]. 统计与信息论坛, 2011, 26(3): 32-38.
FANG Kuangnan, WU Jianbin, ZHU Jianping, et al. A review of technologies on random forests[J]. Stats & Information Forum, 2011, 26(3): 32-38.
- [26] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016: 24-26.
ZHOU Zhihua. Machine learning[M]. Beijing: Tsinghua University Press, 2016: 24-26.
- [27] 牛兴霞, 杨奎河. 基于支持向量机的多类分类研究[J]. 信息技术, 2006(11): 27-31.
NIU Xingxia, YANG Kuihe. Research on multi-classification based on support vector machine[J]. Information Technology, 2006(11): 27-31.
- [28] WANG J, ZHANG M, MEI Y, et al. Study on inversion amplitude of internal solitary waves applied to shallow sea in the laboratory[J]. IEEE Geoscience and Remote Sensing Letters, 2021, 18 (4): 577-581.

Relationship between wave elements of internal solitary waves based on machine learning

LI Zhi-xin, WANG Jing, ZHANG Meng

(Physics and Optoelectronic Engineering, Ocean University of China, Qingdao 266100, China)

Received: Nov. 5, 2020

Key words: internal solitary wave; amplitude; flume experiment; support vector machine; random forest

Abstract: The internal solitary waves that propagate in the ocean carry enormous energy and momentum. Internal solitary waves of varying amplitudes have a different impact on energy exchange and offshore engineering in the ocean. Therefore, it is essential to study the relationship between the amplitude of internal solitary waves, half-wave width, and hydrological characteristic parameters such as depth, stratification, and density. Previously, the relationship between the amplitude of an internal solitary wave and these parameters was constrained by multiple theories. In this paper, a series of comprehensive experiments under different depth, stratification, and density were designed using flume in the laboratory. The relationship between the amplitude of internal solitary waves, half-wave width, depth, stratification, and density is found to be nonlinear. Thus, the machine learning method can be used to establish a nonlinear relationship between the above parameters. We developed a sample database of 1 266 sets, including 970 training sets and 296 test sets using two models, support vector machine (SVM) and random forest (RF). The parameters of the model have been optimized. Finally, the average relative error of the SVM model is 17.3%, whereas that of the RF model is 15.5%. The results show that the machine learning method is effective and feasible. This method can be applied to various hydrological conditions, which effectively solve applicability issues in the previous theory.

(本文编辑: 杨 悦)