

基于机器学习方法的太湖叶绿素 a 定量遥感研究

张玉超¹, 钱新^{1*}, 钱瑜¹, 刘建萍¹, 孔繁翔²

(1. 南京大学环境学院, 污染控制与资源化研究国家重点实验室, 南京 210093; 2. 中国科学院南京地理与湖泊研究所, 南京 210008)

摘要: 为了比较评价人工神经网络和支持向量机 2 种机器学习算法在水质遥感中的应用能力, 本研究首先从基础理论和学习目的入手, 对比分析了 2 种机器学习算法的理论体系; 其次, 以太湖为例, 基于 MODIS 遥感影像, 构建了反演太湖叶绿素 a 浓度的 2 种机器学习方法模型, 通过对模型的验证、稳定性和鲁棒性分析以及全湖反演结果对比 3 个方面评价了 2 种模型的泛化能力. 验证结果表明, 支持向量机模型对验证样本预测结果的均方差根和平均相对误差分别为 5.85 和 26.5%, 而人工神经网络模型的预测结果均方差和平均相对误差则高达 13.04 和 46.8%; 稳定性和鲁棒性评价亦说明, 以统计学习理论为基础的支持向量机模型具有更加良好的稳定性、鲁棒性, 空间泛化能力优于人工神经网络模型; 2 种机器学习算法对太湖叶绿素 a 的浓度分布反演结果基本一致, 但人工神经网络模型因其学习目标设定和网络构建中的“过学习”等缺陷, 造成了对东太湖以及湖心区叶绿素 a 的反演结果与实际监测结果差异较大.

关键词: 人工神经网络; 支持向量机; 水质遥感; 叶绿素 a; 太湖

中图分类号: X87 文献标识码: A 文章编号: 0250-3301(2009)05-1324-08

Quantitative Retrieval of Chlorophyll a Concentration in Taihu Lake Using Machine Learning Methods

ZHANG Yu-chao¹, QIAN Xin¹, QIAN Yu¹, LIU Jian-ping¹, KONG Fan-xiang²

(1. State Key Laboratory of Pollution Control and Resource Reuse, School of the Environment, Nanjing University, Nanjing 210093, China; 2. Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, China)

Abstract: We evaluated the performance of two machine learning methods, artificial neural net (ANN) and support vector machine (SVM), for estimation of chlorophyll a in Taihu Lake from remote sensing data. The theoretical analysis has been done from basic theory and learning target of these two methods first. Then two empirical algorithms have been developed to relate reflectance of MODIS to *in situ* concentrations of chlorophyll a. The performance of ANN and SVM is comparatively analyzed in terms of validation, stability and robustness assessment and chlorophyll a distribution of Taihu Lake from two algorithms. The root of mean square deviation (RMSE) and average relative error (ARE) of validation data is only 5.85 and 26.5% of SVM retrieval model, however, RMSE and ARE of ANN model is 13.04 and 46.8%. Stability and robustness assessment suggest that SVM provides the better performance than ANN. And the retrieval results show that the chlorophyll a distribution of the whole lake from two algorithms is similar, however, the chlorophyll a concentration in the eastern region and central region of Taihu Lake is distorted by ANN model because of the limitations, such as learning target setting and over-learning in net construction.

Key words: artificial neural net (ANN); support vector machine (SVM); water quality retrievals; chlorophyll a; Taihu Lake

湖泊水质遥感是定量遥感应用研究领域的新热点之一, 水质反演模型的构建是其应用的关键^[1,2]. 目前, 湖泊水质遥感的反演方法包括 3 种: 分析模型、经验模型和半经验模型^[3,4], 其中半经验模型是应用最为广泛的方法. 该方法根据光谱仪测量的水质参数光谱特征选择估算水质参数的最佳波段或波段组合^[5~10], 然后选用合适的数学方法^[8,11~17] 建立遥感数据和水质参数间的定量经验性算法.

然而, 这种半经验型反演方法对波段的筛选处理, 忽略了其他波段所包含的未知信息, 容易造成信息丢失; 其次, 由于内陆湖泊的光谱特征非常复杂, 水体组分间的相互影响^[1], 光谱反射率或辐亮度与水体组分间的关系存在着大量不确定性因素, 利用

传统精确数学方法构建的线性模型来反演叶绿素 a 浓度与反射光谱间的非线性关系, 影响了其预测能力. 随着人工智能技术的发展, 以人工神经网络 (artificial neural net, ANN) 为代表的传统机器学习方法以其自适应性、自组织性和容错性, 可以模拟极其复杂的关系, 成为一种非常有效的水质遥感反演策略^[1,18~24].

支持向量机 (support vector machine, SVM), 是一

收稿日期: 2008-05-05; 修订日期: 2008-12-06

基金项目: 国家重点基础研究发展规划 (973) 项目 (2008CB418003); 江苏省自然科学基金项目 (BK2007151); 教育部博士点基金项目 (20060284011)

作者简介: 张玉超 (1977~), 女, 博士, 讲师, 主要研究方向为水质遥感和水环境数值模拟, E-mail: zhangych@nju.edu.cn

* 通讯联系人, E-mail: xqian@nju.edu.cn

种新一代机器学习方法,它以小样本统计理论(统计学习理论)为基础,在解决小样本、非线性极高维模式识别问题中表现出许多特有的优势^[25-27].该方法在遥感领域内的应用,目前主要体现在遥感图像的分类和光谱影像融合方面^[28-30],在水色定量遥感研究领域应用较少^[31].对于湖泊监测而言,由于受到人力、物力和财力的限制,现场实测点的数量有限,因此,针对支持向量机在水质遥感领域的应用研究具有十分积极的意义.

为了比较分析 2 种机器学习算法在水质遥感中的应用能力,本研究从 ANN 和 SVM 的理论基础入手,详细探讨了 2 种机器学习方法的理论差异,并以太湖为对象,基于 MODIS 遥感影像,对 2 种反演模型的泛化能力进行了比较分析.

1 ANN 与 SVM 的基本原理

基于数据的机器学习是现代智能技术的重要方面. ANN 和 SVM 是 2 种典型的机器学习方法,在机器学习问题中发挥着极其重要的作用. 基于这 2 种方法的水质反演模型均反映了机器学习方法在回归函数估计研究领域内的应用. 但是, ANN 和 SVM 的理论体系却是完全不同的.

1.1 基础理论

ANN 是基于传统统计学理论的,其遵循的是样本数目趋于无穷大时的渐近理论. 而支持向量机是以小样本统计理论(统计学习理论)为基础的,统计学习理论(statistical learning theory, SLT)是一种专门研究小样本情况下的机器学习规律的理论,该理论针对小样本统计问题建立了一套新的理论体系,该体系下的统计推理规则不仅考虑了对渐近性能的要求,而且追求在现有有限样本的条件下得到最优结果.

1.2 应用原理

1.2.1 ANN 的原理

人工神经网络的模型很多,其中具有非线性连续转移函数的多层前馈网络的误差反向传播(error back propagation, BP)是现今应用最广泛的神经网络. BP 算法的学习过程通过对训练样本的反复学习,不断地调整各层间连接权值和阈值,并采用最快下降法使得权值沿着误差函数负梯度方向改变,收敛于最小点. 可见,该算法实际上是求误差函数的极小值^[32].

1.2.2 SVM 的原理

SVM 是从分类问题线性可分情况下的最优分

类面提出的,所谓最优分类线就是要求分类线不但能将两类正确分开(训练错误率为 0),而且使分类间隔最大,确保其推广能力. 在高维空间中,最优分类线就称为最优分类面,而离分类面最近的点且平行于最优分类面的超平面的训练样本称为支持向量. 其应用基本思想可以概括为:首先通过非线性变换将输入空间变换到一个高维空间,然后在这个新空间中求取最优线性分类面,而这种非线性变换是通过定义适当的内积函数(核函数)实现的. 上述支持向量机的基本原理是基于分类问题的,在回归或函数拟合问题中,只要把函数连续的输出通过一个阈值转化为二值函数,函数拟合问题就变成分类问题了^[25-27].

1.3 学习目标

机器学习的目的在于使期望风险最小,该目标的实现标志着机器学习方法具有良好的推广能力. BP 神经网络算法以训练样本信息为基础,利用样本定义经验风险来代替期望风险,即经验风险最小化理论(empirical risk minimization, ERM)准则. 理论上,只有在样本数量足够多的情况下,才能真正实现经验风险对期望风险的替代. 正是 ERM 准则的缺陷导致了神经网络的“过学习”问题:即训练样本的误差小,能够保证学习的高精度,但未必能够得到一个好的预测效果,反而可能导致真实风险的增加,推广能力的下降.

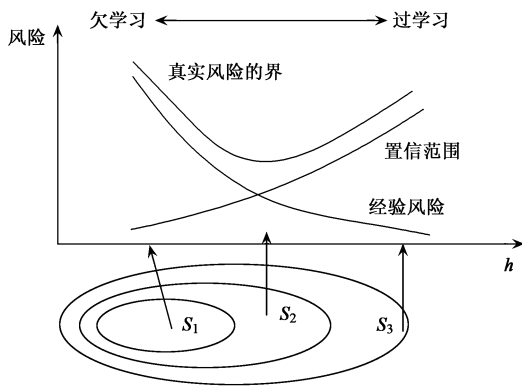
SVM 的统计学习理论用“推广性的界”来表示经验风险和真实风险之间的关系,将真实风险定义为经验风险的值和置信范围的值之和,并提出了结构风险最小化理论(structural risk minimization, SRM)准则:即通过设计函数集的某种结构使每个子集中都能取得最小的经验风险(如使训练误差为 0),然后选择适当的子集使置信范围最小的方法实现,这个子集中不仅使经验风险最小而且使其置信范围最小的函数就是最优函数(见图 1). 根据 SRM 准则, SVM 方法得到的最优函数是全局的最优,而不仅仅是训练样本的最优^[25-27].

上述理论分析表明,在研究样本有限问题时,理论上保证了 SVM 方法应该比 ANN 方法有着更好的适应性和推广性.

2 数据的预处理

2.1 现场监测

现场监测时间为 2005 年 8 月 18 日和 9 月 15 日,监测点布设采取均匀分布(湖心区)与重点布设



函数集子集: $S_1 \subset S_2 \subset S_3$ VC维: $h_1 \leq h_2 \leq h_3$

图 1 结构风险最小化示意^[20]

Fig. 1 Sketch of SRM theory

(污染严重的竺山湖、梅梁湾、贡湖以及水质良好的胥湖和东太湖等)相结合的方法, 8月份在太湖全湖水域内共选取了 29 个监测点, 9月份共 13 个监测点(即 1~13 号)主要位于北太湖, 各监测点位置详见图 2. 监测因子主要包括 Chl a 和营养盐等, 并进行经纬度信息及相关气象条件的同步观测.

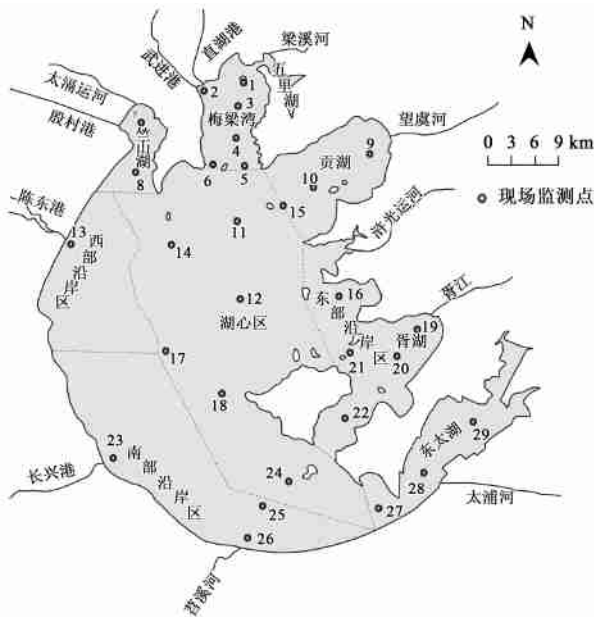


图 2 太湖现场监测布点示意

Fig. 2 Location of sampling stations of Taihu Lake

2.2 MODIS 遥感数据的预处理

本研究所选择分析的 MODIS 遥感影像的成像日期为 2005 年 8 月 15 日和 9 月 16 日. MODIS 数据的 B1 和 B2 波段原始空间分辨率为 250 m, 其余波段在几何校正过程中均由原来的 500 m 或 1 km 的分辨率重采样为 250 m. 利用 MODIS 数据本身带有

详细的经纬度波段信息, 选择 UTM 投影, WGS84 椭球体, 完成采样点 GPS 定位坐标和遥感图像坐标精确匹配. 几何纠正总误差控制在 0.5 个像元之内. 图像重采样采用最邻近点法. 根据 MODIS 头文件中辐射亮度缩放系数和缩放截距信息, 完成辐射定标, 再选用“6S”模型进行大气校正.

3 ANN 与 SVM 反演模型的构建

3.1 反演模型的构建流程

机器学习方法有着共同的特点, 都是从研究观测数据入手寻找规律, 利用这些规律对未来数据或无法观测的数据进行预测. 因此, 基于机器学习方法的反演模型的构建流程大体上是一致的: 第一步, 主要完成训练样本集和测试样本的构建; 第二步, 反演模型的训练与测试, 通过修改相关参数, 对反演模型进行反复训练与测试, 直至得到最优结果即训练好的反演模型; 第三步, 利用训练好的反演模型对研究对象进行全面反演, 并对其进行检验与分析.

3.2 反演模型的构建方式

3.2.1 研究对象

叶绿素 a 是影响湖泊水体光谱特征的重要参数, 也是湖泊水质遥感研究和应用比较成熟的水质遥感指标^[33]. 因此, 本研究以叶绿素 a 浓度作为水质反演指标.

3.2.2 输入信息

根据光谱特征分析, 太湖叶绿素在 405~496 nm、503~564 nm、672~767 nm 以及 703~709 nm 的反射率对 Chl a 的浓度变化较为敏感^[34], 为了让学习机器更加全面的掌握叶绿素 a 的反射特征, 本研究直接将 MODIS 位于可见光-近红外的波段反射率作为输入信息. MODIS 在 400~1000 nm 上有 18 个波段, 删除具有显著条状噪声带波段 B8、B9, 考虑到个别波段在波长范围上存在重复, 为了防止数据的冗余, 删除了存在波长范围重叠的波段, 最终将 B1~B4、B10、B11、B14b、B15、B17 以及 B19 共计 10 个波段上的信息作为反演模型的输入信息. 因此, 太湖水质反演问题就转化为 10 个输入信息、一个输出信息的神经网络或支持向量机的函数回归问题, 即:

$$X = (x_{B1}, x_{B2}, x_{B3}, x_{B4}, x_{B10}, x_{B11}, x_{B14b}, x_{B15}, x_{B17}, x_{B19})$$

$$Y = C_{Chl a}$$

3.2.3 训练、测试和验证样本的构建

2 次监测共计 42 个监测点, 其中 32 个监测点数据用来构建模型, 10 个监测点数据用来验证模型的

预测效果. 32 个建模样本中, 26 个作为训练样本, 其余 6 个作为测试样本. 训练、测试和验证样本为随机选取, 并同时考虑叶绿素 a 浓度的高低梯度分布.

3.2.4 模型参数的选取

本研究中所选用的神经网络模型, 是包含输入层、隐含层和输出层的 3 层 BP 神经网络, 输出层函数为线性函数, 隐含层函数为 sigmoid 函数, 即:

$$f(x) = \frac{1}{1 + e^{-x}}$$

SVM 选用的核函数是常用的径向基核函数, 即:

$$K(x, x_i) = \exp\left[-\frac{\|x - x_i\|^2}{\sigma^2}\right]$$

ANN 模型中隐含层节点数和 SVM 模型中的参数 σ 、惩罚系数 C 通过训练测试确定. 在训练过程中多次选择不同的参数值, 分别通过考察模型对训练样本以及测试样本的预测结果, 寻求模型的最优参数值. 经筛选, 最优 ANN 模型中隐含层节点数为 2, 计算步数最大值设置为 10 000; 最优 SVM 模型的参数 σ 为 0.90, 惩罚系数为 1 000.

4 ANN 和 SVM 模型的泛化性评价

在机器学习理论中, 通常将学习结果对新情况的预测能力作为评价一个学习算法优劣的主要标准, 而学习算法的预测能力也常称之为该算法的泛化能力 (generalization)^[35]. 本节利用验证样本检验了 2 种模型的预测能力, 并从模型的稳定性和鲁棒性 2 个方面, 对 ANN 模型和 SVM 模型的泛化能力进行比较分析. 评价中主要考察模型预测结果的均方差根 (RMSE) 和平均相对误差 (ARE), 即:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (C_{mi} - C_{si})^2}$$

$$\text{ARE} = \frac{1}{n} \sum_{i=1}^n \frac{|C_{mi} - C_{si}|}{C_{mi}}$$

式中, n 为样本数; C_{mi} 为监测值; C_{si} 为预测值.

4.1 模型的验证

图 3 为 10 个验证样本对 ANN 模型和 SVM 模型的验证结果. ANN 模型对训练、测试样本预测结果的 RMSE 分别为 3.79 和 8.60, SVM 模型对训练测试样本预测结果的 RMSE 也分别达到了 4.16 和 10.92, 在模型构建效果相近的情况下, SVM 模型对验证样本的预测结果 RMSE 依然有 5.85, ARE 为 26.5%, 而 ANN 模型的 RMSE 却高达 13.04, ARE 达到了 46.8%, 这是由于 ANN 模型对高浓度验证样本的预测结果明显高于实测值造成的, 而 SVM 模型对

高浓度值和低浓度值均有较好的预测结果.

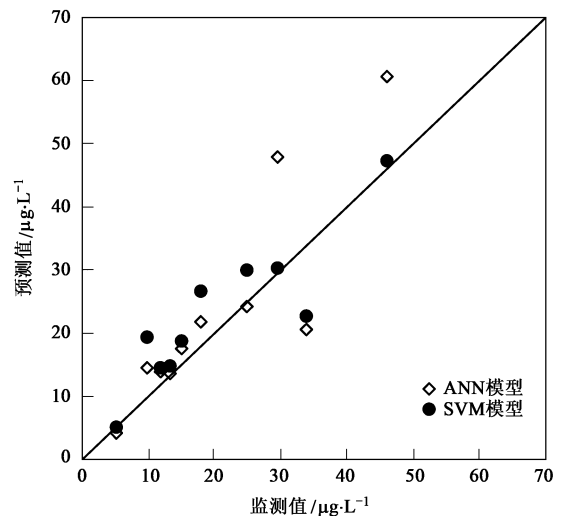


图 3 ANN 模型和 SVM 模型的验证结果

Fig. 3 Validation results of ANN model and SVM model

4.2 稳定性评价

在训练样本中抽取不同比例的样本数作为训练样本来训练模型, 通过分析模型对建模样本和验证样本的预测能力, 来实现对模型稳定性的评价. 本研究分别选取了 25%、50%、75%、100% 的训练样本, 即 7、13、20 和 26 个监测点的数据作为训练样本, 测试样本和验证样本保持不变. 2 种模型在反演叶绿素 a 的稳定性分析见表 1.

对建模数据的预测结果分析, ANN 模型和 SVM 模型均随着训练样本的增加, 对测试数据的预测能力逐渐提高. 对于验证样本的预测效果, SVM 模型呈现出预测精度逐渐稳定升高的趋势, 而 ANN 模型的预测结果并不稳定. 这是因为 ANN 模型以追求最低样本误差作为学习目标, 所构建的最优网络可能是局部最优, 出现了“过学习”的现象; 而 SVM 模型的学习目标在最低样本误差的基础上, 增加了一定的置信区间范围, 尽管对训练、测试样本的预测精度不如 ANN 模型, 但对验证样本效果明显优于 ANN 模型. 上述分析表明, 在训练样本数逐渐增多的情况下, SVM 模型能够快速地掌握研究对象的特征, 预测结果的稳定性要优于 ANN 模型.

4.3 鲁棒性评价

鲁棒性 (robustness) 来自于现代控制理论, 主要用于评价控制系统在其特性或参数发生摄动时仍可使质量保持不变的性能. ANN 模型和 SVM 模型在模型参数确定的情况下, 其预测性能不仅与训练样本

表 1 对 ANN 模型和 SVM 模型的稳定性分析

Table 1 Stability analysis of ANN and SVM models

模型	训练样本数/个	建模样本				验证样本	
		训练样本		测试样本		均方差根 RMSE	平均相对误差 ARE/ %
		均方差根 RMSE	平均相对误差 ARE/ %	均方差根 RMSE	平均相对误差 ARE/ %		
ANN	7	14.52	28.2	12.38	30.9	8.45	34.0
	13	11.40	22.4	10.58	39.8	7.78	31.7
	20	3.74	15.5	9.67	38.4	11.20	37.0
	26	3.79	14.9	8.60	55.0	13.04	46.8
SVM	7	0.00	3.5	22.72	91.4	20.14	103.2
	13	2.05	4.9	12.34	36.5	8.81	58.6
	20	4.96	12.8	12.13	40.1	7.03	41.0
	26	4.16	15.0	10.92	20.0	5.85	26.5

的数量有关, 还与其质量密切相关. 因此, 本研究中的鲁棒性评价, 是通过分析训练样本中假定特异样本对模型预测能力的影响程度来实现的, 即: 在同一训练样本和测试样本条件下, 将训练样本的某一目标值做人为改动, 进而分析模型对验证样本的影响程度. 为了考察特异训练样本对 2 种模型的影响, 本研究假设了如下 2 个情景: ①训练样本中叶绿素 a 由原始值 $5.36 \mu\text{g/L}$ 分别变为 10、20、30、40、50 和 $60 \mu\text{g/L}$; ②训练样本中叶绿素 a 由原始值 $42.41 \mu\text{g/L}$ 分别变为 40、30、20 和 $10 \mu\text{g/L}$.

结果表明, 2 种模型对上述 2 种情景的反应不

尽相同, 图 4 为浓度值发生改变时, 预测值的变动范围. 对于机器学习算法, 训练样本的个别特异样本会造成预测结果发生不同程度的变化, ANN 模型对于特异训练样本比 SVM 模型要更加敏感, 造成个别验证样本在 2 种情景下预测值变动范围分别高达 $47 \mu\text{g/L}$ 和 $28 \mu\text{g/L}$, 而 SVM 同样情况下, 预测值最大变动范围仅为 $7.9 \mu\text{g/L}$ 和 $6.6 \mu\text{g/L}$. 但总体来看, 个别训练数据的特异, 并不会造成模型预测能力的严重失真, 这说明了 2 种模型均具有良好的鲁棒性, 相对而言, SVM 模型不会因个别训练数据失真而造成预测结果发生较大尺度变动, 具有更优秀的鲁棒性.

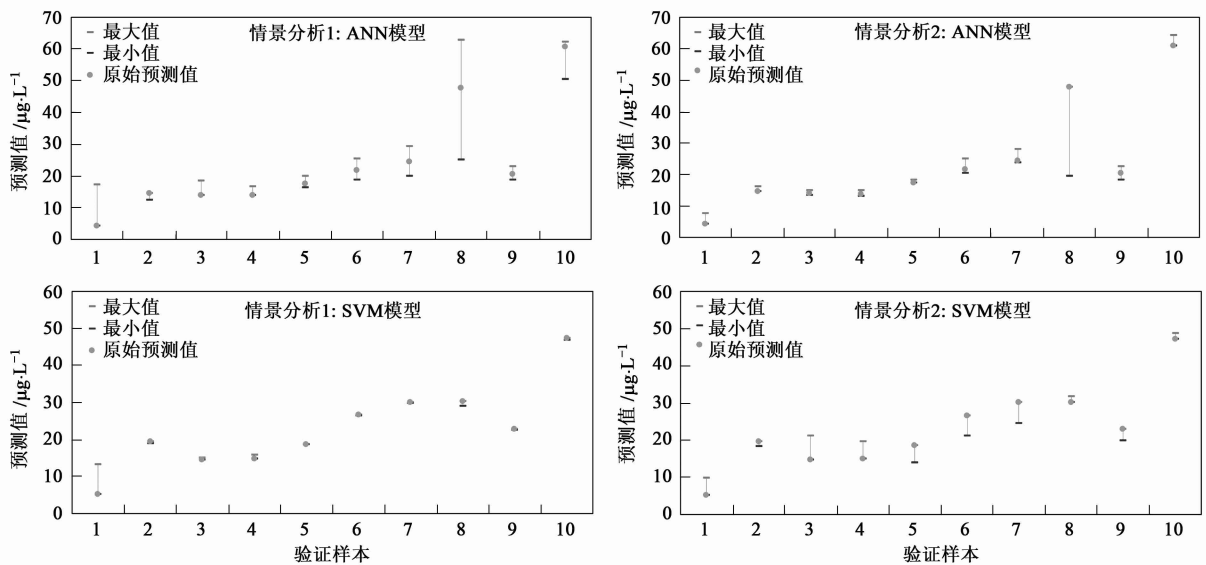


图 4 ANN 模型和 SVM 模型鲁棒性分析

Fig. 4 Robustness analysis on ANN and SVM models

4.4 泛化性评价

通过对 2 种模型预测能力的验证以及稳定性和

鲁棒性分析可知, 相对而言, SVM 模型的泛化能力要优于 ANN 模型. 正是 2 种机器学习方法在理论基

础上的不同导致了应用中的差异.

5 基于 2 种机器学习算法的太湖 Chl a 浓度反演结果

为了进一步对比 2 种机器学习算法的应用能力,本研究利用构建好的 ANN 模型和 SVM 模型分别反演了太湖 2005 年 8 月 15 日和 9 月 16 日叶绿素 a 全湖分布情况(见图 5).

从总体上看, 2 种机器学习算法对太湖 2 个时刻叶绿素 a 全湖分布的反演结果基本上是一致的:

①太湖叶绿素 a 浓度呈现西部高于东部、北部高于南部的特征; ②梅梁湖、贡湖的北部沿岸区以及太湖

的西南沿岸区叶绿素 a 浓度较高; ③太湖的东部沿岸区以及东太湖地区生长了大量的水生植物, 水质较好, 叶绿素 a 浓度较低. 上述结论与太湖叶绿素 a 常规分布大体相同, 太湖地区夏季盛行东南风, 漂浮于水体表面的藻类在风生流的作用下, 容易在下风向积聚, 形成叶绿素 a 高浓度区域. 2005 年 8 月 15 日卫星过境之前, 太湖地区盛行东南偏东风(详见图 5 中风速风向矢量图), 造成了藻类在太湖西部地区的堆积; 而 9 月 16 日卫星过境之前盛行东南风, 太湖西北地区的叶绿素 a 浓度明显增高. 此外, 东太湖地区由于水面覆盖着水生植物, 造成该区域各波段反射率明显低于其他水域, 机器学习算法通过对该

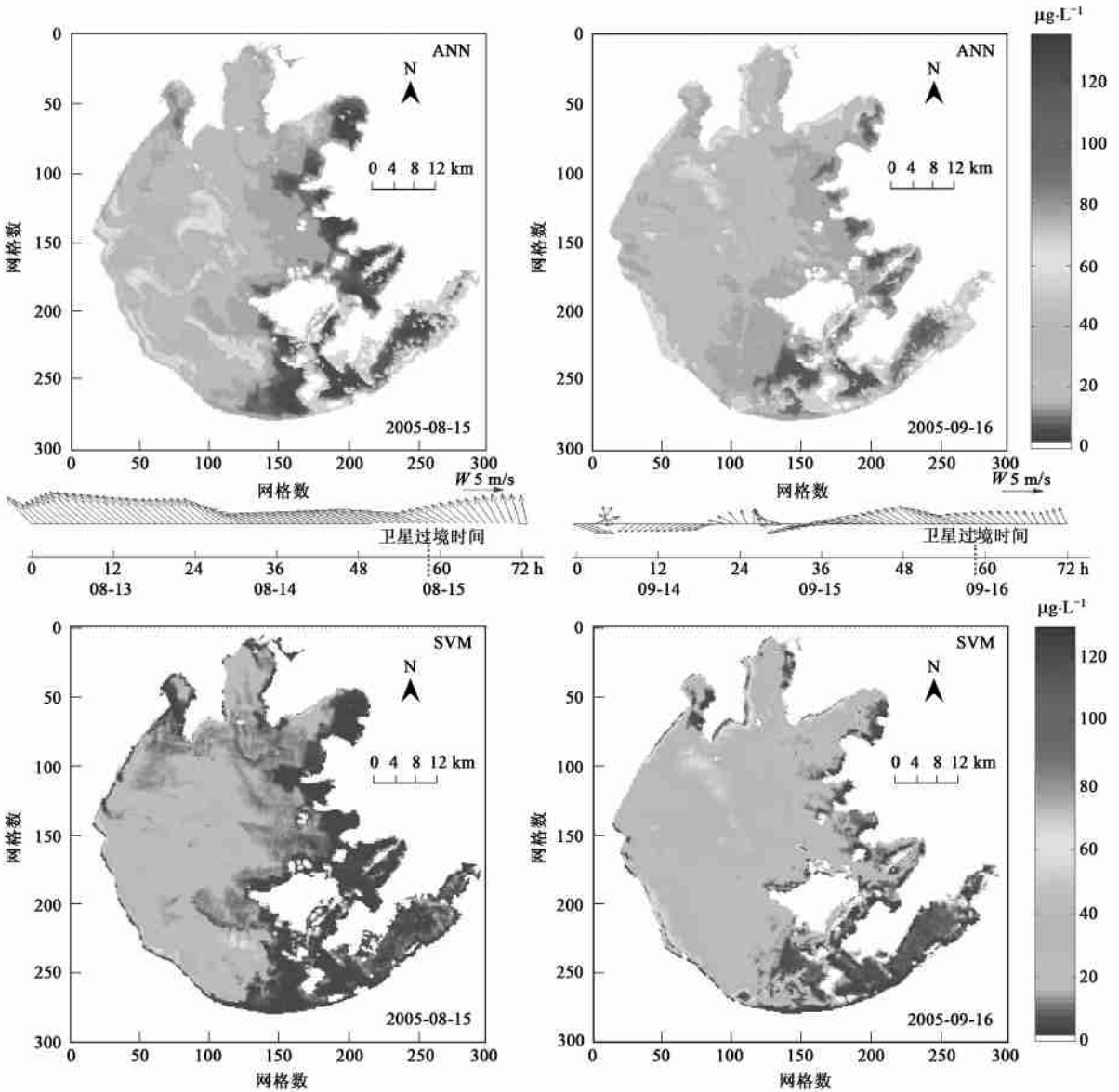


图 5 太湖叶绿素 a 浓度分布图

Fig. 5 Chlorophyll a distribution of Taihu Lake

水域训练样本的学习, 掌握了其光谱特征, 因此, 能够准确的反演出低浓度的特点。

但是, 2 种方法的反演结果中均出现了竺山湖和贡湖叶绿素 a 浓度 8 月 15 日偏低的现象。据太湖流域水资源保护局相关资料^[36,37]: 2005 年, 太湖流域梅雨期降水偏少, 太湖水位较常年偏低, 但是 8 月上旬的台风“海棠”、“麦莎”造成太湖流域普遍出现暴雨, 作为太湖主要入湖河流的太 运河(占 24.9%) 入湖流量迅速增加, 对河口附近叶绿素 a 浓度存在一定冲刷作用, 造成叶绿素 a 浓度偏低; 对于贡湖, 2005 年 6 月引江济太调水工程通过望虞河引水入贡湖达 10.8 亿 m³, 贡湖蓝藻基本消失, 水质通报显示^[38,39], 至 8 月份, 尽管调水已经结束, 但贡湖依然维持着 III 类水质。上述分析表明, 8 月 15 日的太湖竺山湖和贡湖的反演结果具有一定的合理性。

2 种方法在东太湖近岸区, 以及湖心区叶绿素 a 高浓度区域出现了预测结果的差异。东太湖地区水生植物繁盛, 水质保持为 III 类, 中营养水平, 叶绿素 a 浓度较低, ANN 模型反演出个别地区叶绿素 a 浓度高达 60 μg/L, 明显与事实不符, 而 SVM 模型在东太湖的反演结果与实际情况相符较好; 文中 4.1 节对 2 种模型的验证结果表明, ANN 模型对于高浓度验证样本的预测结果偏离监测值 31.9% 左右, 而 SVM 模型仅为 2.6%, 因而会出现 2 种模型在高浓度值预测结果上的显著差异。

6 结论

(1) 在研究有限样本问题时, 基于统计学习理论的 SVM 方法在理论上比 ANN 方法有着更好的适应性和推广性。

(2) SVM 对验证样本预测结果的 RMSE 和 ARE 分别为: 5.85 和 26.5%, 而 ANN 模型的 RMSE 和 ARE 却分别高达 13.04 和 46.8%。

(3) 稳定性和鲁棒性评价说明, SVM 模型能够快速掌握研究对象的特征, 预测结果的稳定性要优于 ANN 模型; 而个别训练数据的特异, 均不会造成 2 种模型预测能力的严重失真, 说明 ANN 和 SVM 均具有良好的鲁棒性。

(4) 2 种机器学习算法对太湖 2 个时刻叶绿素 a 全湖分布的反演结果基本上是一致的, 但在东太湖近岸区和湖心区 ANN 模型的学习目标设定和网络构建中的“过学习”等缺陷造成了反演结果与监测结果差异较大, 而 SVM 模型反演结果的相符性较好。

致谢: 本研究中的太湖叶绿素 a 浓度监测数据

由中国科学院太湖生态系统研究站提供; MODIS 遥感数据由中国科学院地理与资源研究所地球变化信息与研究中心提供。

参考文献:

- [1] 吕恒, 江南, 李新国. 内陆湖泊的水质遥感监测研究[J]. 地球科学进展, 2005, 20(2): 185-192.
- [2] 乐成峰, 李云梅, 孙德勇, 等. 太湖叶绿素 a 浓度时空分异及其定量反演[J]. 环境科学, 2008, 29(3): 619-626.
- [3] 刘灿德, 何报寅. 水质遥感监测研究进展[J]. 世界科技研究与发展, 2005, 27(5): 40-44.
- [4] 周艺, 周伟奇, 王世新, 等. 遥感技术在内陆水体水质监测中的应用[J]. 水科学进展, 2004, 15(3): 312-317.
- [5] Dekker A G, Peters S W. The use of the thematic mapper for the analysis of eutrophic lakes: a case study in the Netherland [J]. Int J Remote Sens, 1993, 14(5): 799-821.
- [6] Baban S J. Detecting water quality parameters in the Norfolk Broads, U K, using landsat imagery [J]. Int J Remote Sens, 1993, 14(7): 1247-1267.
- [7] Tassan S. A numerical model for the detection of sediment concentration in stratified river plumes using thematic mapper data [J]. Int J Remote Sens, 1997, 18(12): 2699-2705.
- [8] Chen Z M, Hanson J D, Curran P J. The form of the relationship between suspended sediment concentration and spectral reflectance: its implications for the use of Daedalus 1268 data [J]. Int J Remote Sens, 1991, 12(1): 215-222.
- [9] Lathrop R G. Landsat thematic mapper monitoring of turbid inland water quality [J]. Photogramm Eng Remote Sens, 1992, 58(4): 4465-470.
- [10] Kallio K, Kuster T, Koponen S, et al. Retrieval of water quality from airborne imaging spectrometry of various lake types in different seasons [J]. Sci Total Environ, 2001, 268(1): 56-77.
- [11] 赵碧云, 贺彬, 朱云燕, 等. 滇池水体中叶绿素 a 含量的遥感定量模型[J]. 云南环境科学, 2001, 20(3): 1-3.
- [12] 焦红波, 查勇, 李云梅, 等. 基于高光谱遥感反射比的太湖水体叶绿素 a 含量的估算模型[J]. 遥感学报, 2006, 10(2): 242-248.
- [13] 雷坤, 郑丙辉, 王桥. 基于中巴地球资源 1 号卫星的太湖表面水体水质遥感[J]. 环境科学学报, 2004, 24(3): 376-380.
- [14] 闻建光, 肖青, 杨一鹏, 等. 基于 Hyperion 数据的太湖水体叶绿素 a 浓度遥感估算[J]. 湖泊科学, 2006, 18(4): 327-336.
- [15] 祝令亚, 王世新, 周艺, 等. 应用 MODIS 监测太湖水体叶绿素 a 浓度的研究[J]. 应用技术, 2006, 2: 25-28.
- [16] 光洁, 韦玉春, 黄家柱, 等. 分季节的太湖悬浮物遥感估测模型研究[J]. 湖泊科学, 2007, 19(3): 241-249.
- [17] 祝令亚, 王世新, 周艺, 等. 应用 MODIS 监测太湖水体悬浮物浓度的研究[J]. 水科学进展, 2007, 18(3): 444-450.
- [18] 石爱业, 徐立中, 杨先一, 等. 基于知识和遥感图像的神经网络水质反演模型[J]. 中国图像图形学报, 2006, 11(4): 521-528.
- [19] Keiner L E, Yan X H. A neural network model for estimating sea surface chlorophyll and sediments from thematic mapper imagery[J]. Remote Sens Environ, 1998, 66(2): 153-165.
- [20] Buckton D, Omorgain E. The use of neural networks for the

- estimation of oceanic constituents load on the MERIS instrument [J]. *Int J Remote Sens*, 1999, **20**(9): 1841-1851.
- [21] Gross L, Thiria S. Applying artificial neural network methodology to ocean color remote sensing [J]. *Ecol Model*, 1999, **120**(2-3): 237-246.
- [22] 詹海刚, 施平, 陈楚群. 利用神经网络反演海水叶绿素浓度 [J]. *科学通报*, 2000, **45**(17): 1879-1884.
- [23] 张亭祿, 贺明霞. 基于人工神经网络的一类水域叶绿素 a 浓度反演方法 [J]. *遥感学报*, 2002, **6**(1): 40-44.
- [24] 黄海清, 何贤强, 王迪峰. 神经网络反演海水叶绿素 a 浓度的分析 [J]. *地球信息科学*, 2004, **6**(2): 31-36.
- [25] 边肇祺, 张学工编著. 模式识别 [M]. (第二版). 北京: 清华大学出版社, 2000. 284-303.
- [26] Vapnik V N 著, 张学工译. 统计学习理论的本质 [M]. 北京: 清华大学出版社, 2000. 96-100.
- [27] Cristianini N, Shawe-Taylor J 著. 李国正, 王猛, 曾华军译. 支持向量机导论 [M]. 北京: 电子工业出版社, 2004. 82-106.
- [28] Camacho H, Gonalves P, Caetano M. Contribution of multispectral and multitemporal information from MODIS images to land cover classification [J]. *Remote Sens Environ*, 2008, **112**(3): 986-997.
- [29] Foody G M, Mathur A. The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM [J]. *Remote Sens Environ*, 2006, **103**(2): 179-189.
- [30] Zhu G, Blumberg D G. Classification using ASTER data and SVM algorithms: The case study of Beer Sheva, Israel [J]. *Remote Sens Environ*, 2002, **80**(2): 233-240.
- [31] Gustavo C, Luis G, Jordi M, *et al.* Retrieval of oceanic chlorophyll concentration with relevance vector machines [J]. *Remote Sens Environ*, 2006, **105**(1): 23-33.
- [32] 周伽. 非经典数学方法在非线性时间序列预测中的应用研究 [D]. 南京: 南京航空航天大学, 2005. 18-27.
- [33] 祝令亚. 湖泊水质遥感监测与评价方法研究 [D]. 北京: 中国科学院遥感应用研究所, 2006. 15-21.
- [34] 王艳红, 马荣华, 邓正栋. 基于分区的太湖叶绿素 a 遥感估测模型 [J]. *河海大学学报*, 2007, **35**(1): 86-91.
- [35] 张海, 徐宗本. 学习理论综述 (I): 稳定性与泛化性 [J]. *工程数学学报*, 2008, **25**(1): 1-8.
- [36] 太湖流域管理局水文处. 太湖流域片水情年报 (2005) [R]. 2007. 2-22.
- [37] 太湖流域管理局. 太湖流域及东南诸河水资源公报 (2005) [R]. 2006. 22-24.
- [38] 太湖流域水资源保护局. 太湖流域及东南诸河省界水体水资源质量状况通报 (99 期) [R]. 2005. 4-6.
- [39] 太湖流域水资源保护局. 太湖流域及东南诸河省界水体水资源质量状况通报 (100 期) [R]. 2005. 4-6.