基于支持向量机的浙江近海渔船捕捞方式识别

任迎春[1] 刘静[1] 翟振刚[2] 王薇[1] 林萱[31]

(1.嘉兴学院数理与信息工程学院，浙江嘉兴 314001；

2.中国电子科技集团公司第三十六研究所，浙江嘉兴 314033；

3.温岭师范附属小学，浙江台州 317500)

【摘 要】：以浙江近海拖网渔船、流动张网渔船和流动刺网渔船各 40 艘共 120 个样本为研究对象，以支持向量机为研究模型，根据 2018 年天通一号卫星监测系统中对应渔船的航速数据构建特征对其捕捞方式进行识别。实验结果表明，与 k 近邻、逻辑回归、决策树、随机森林、自适应提升树和神经网络等算法相比，基于支持向量机的浙江近海渔船捕捞方式的识别正确率最高，可以达到 90%，研究结果对监测浙江近海渔船捕捞行为、维护海洋生态平衡具有一定的参考价值。

【关键词】：近海渔船 捕捞方式 天通一号 支持向量机 识别模型

【中图分类号】：TP302.7**【文献标志码】**：A**【文章编号】**：1671-3079(2020)06-0082-08

浙江省拥有绵长的海岸线，渔业资源丰富。近海捕捞渔业在保障水产品供给、促进传统渔民及内陆农民就业、维护国家海洋权益等方面发挥了重要作用。[1-2]但近年来，随着渔船数量的日益增加和渔船机械化自动化程度的不断提高，捕捞强度越来越大，不合理的捕捞方式致使海洋渔业资源严重衰退，海洋渔业生产面临危机。[3-4]

为加强渔船作业的合法性和安全性，提高监控和管理效率，我国海洋渔业监管部门采用了北斗卫星或天通一号等渔船监控系统(Vessel Monitoring System,简称 VMS),[5-7]安装有 VMS 的渔船在出海捕捞过程中的时间、航向、位置和航速等信息都可以被记录下来。这些丰富的数据使基于数据挖掘技术识别渔船作业类型、判断渔船捕捞状态、分析渔船捕捞量、获取渔船捕捞行为特点等成为可能，进而为精细化的渔业资源保护和管理提供丰富的参考数据。

目前，基于 VMS 系统的渔船数据、海洋环境数据和渔业生产数据的分析研究受到人们的关注。[8-14]在国外，Witt 等人使用航速作为变量，提取航速阈值判断渔船的作业状态;[8]Deng 等人综合航速和航向阈值对渔船的作业状态进行建模分析，航向的使用进一步提高了预测精确性。[9]Joo 等人借助神经网络模型对渔船的航行和作业状态进行识别，[10]预测精度达到 76%。Russo 等人利用贝叶斯模型分析了捕捞量和捕捞行为之间的相关性。[11]Salcedo 等人基于 VMS 数据提取渔船轨迹，使用聚类分析识别出了渔船航行和捕捞状态。[12]而在国内，基于渔船监测系统的数据处理、渔船捕捞类型判别和复杂算法等方面的研究较少。张胜茂等人根据航速与航向的差值得出累计作业时长，分析捕捞强度。[13]郑巧玲等人借助 2014 年的北斗渔船数据构建基于 BP 神经网络的中国近海渔船作业方式识别模型，对捕捞类型进行分类和识别，取得了较好的效果。[14]然而，这些基于神经网络的识别模型

[1]**基金项目**：浙江省重点研发计划项目(2019C03099)；浙江省自然科学基金项目(LQ20F020027,LY18A010017)
作者简介：任迎春(1982-)，男，河南南阳人，嘉兴学院数理与信息工程学院教师，博士，研究方向为数据挖掘、计算机视觉等。

存在参数众多、计算复杂度过高、确定网络结构困难并且无法获取全局最优解等缺陷，制约着该模型的广泛使用。

本文利用渔船的航速数据构造特征并建立基于支持向量机(Support Vector Machine,SVM)的捕捞方式识别模型。SVM 以统计学理论为基础，[15]采用结构风险最小化准则设计学习机器，较好地解决了非线性、高维数、局部极小点等问题，更易于推广。研究表明，航速对渔船的捕捞方式较为敏感,不同的捕捞类型，其航速占比差异明显，而航速数据可通过天通一号卫星系统直接获取，便于操作和研究。因此，本文以 2018 年天通一号卫星系统监测的浙江省近海领域内 120 艘渔船为研究对象，构造其航速占比数据集，并采用支持向量机对渔船作业类型进行识别，所提出的渔船捕捞方式识别模型，可以为渔业监管部门提供辅助参考和决策依据，对保护海洋生态环境、打击非法捕捞行为具有重要意义。

1 数据准备

本文的实验数据均来自装有天通一号设备的浙江省近海渔船所采集、传输并存储后的真实数据,时间跨度为 2018 年 9 月 1 日至 2018 年 11 月 30 日。数据库存储了船号、船的所属人、功率、材质、规定的作业方式、作业海域等相关的静态信息以及渔船的航行时间、经纬度、航向、航速等实时动态信息。时间信息可以统计渔船出海至归来的时间跨度；经纬度可以捕获渔船的位置信息，便于安全救助、捕捞管理等；航速、航向可用于渔船捕捞状态、捕捞强度、捕捞方式的分析。由于数据库中渔船相关登记信息具有不完整性和不规范性，本文首先利用 SQL 语句统计具有完整信息的渔船数据，再基于完整信息进行数据处理和建模分析。

1.1 浙江省近海渔船捕捞方式分布

根据《渔业捕捞许可管理规定》，渔船捕捞的作业类型主要分为 9 种：刺网、张网、围网、拖网、钓具、耙刺、陷阱、笼壶和杂渔具(含地拉网、敷网、抄网、掩罩及其他杂渔具)。如表 1 所示,在统计 2017 及 2018 年的浙江省渔船数据时发现，刺网、拖网和张网这三种捕捞方式的使用范围占全部捕捞方式的 90%,如表 1 所示。因此，本文主要判别刺网、拖网和张网这三种主要捕捞方式。

表 1 2017 及 2018 年浙江省渔船捕捞作业类型统计表

类型		拖网	刺网	张网	围网	笼壶	钓具	耙刺	杂渔具	陷阱
2017	船数	898	599	188	38	51	23	1	70	0
	占比/%	48.07	32.07	10.06	2.03	2.73	1.23	0.05	3.75	0
2018	船数	809	599	233	28	40	17	10	70	0
	占比/%	44.79	33.17	12.90	1.55	2.21	0.94	0.55	3.88	0

1.2 3 种主要捕捞方式的原理分析

在近海作业时，拖网渔船首先以较大的航速航行到目标渔场，然后放慢航速并释放网具。网具完全释放后所经水域的生物基本都会被捕捞进网中，因此，拖网捕捞对海洋生物的破坏性较大，会严重破坏渔业的生态平衡。

刺网渔船装备一块长带形网具，网上带有若干挂刺。当渔船拖着刺网行进时，利用水流作用，刺网上的挂刺会钩住一部分鱼虾蟹等捕捞物。刺网渔船的结构简单，不受渔场环境限制，作业范围广。相对拖网而言，刺网的破坏性并不强，但对海洋生物链造成的影响依旧较大。

张网在我国渔业生产中使用较为广泛，是一种非常重要的传统定置工具。张网是将网固定在水中,然后利用水流迫使鱼类进入网囊的网具。张网是一种被动渔具，因此它的破坏性相比较而言会小一些。

1.3 3 种主要捕捞方式的航速分析

为研究不同捕捞方式具体作业状态时的航速特征，本文随机选取了两艘拖网渔船、两艘刺网渔船和两艘张网渔船 3 个月的航速数据，根据捕捞状态下的数据作航速占比图。

如图 1 所示，拖网渔船在航速占比图上容易形成单峰值图像，且一般在区间 1～3 之间达到峰值。如图 2 所示，刺网渔船在航速占比图上容易形成双峰值图像，其在区间 0～2 之间达到第一个峰值,在区间 6～8 之间达到第二个峰值且第二个峰值一般不会超过第一个峰值。如图 3 所示，张网渔船的航速占比图类似于半区间上的正态分布图，其在 0～1 之间达到最大值。

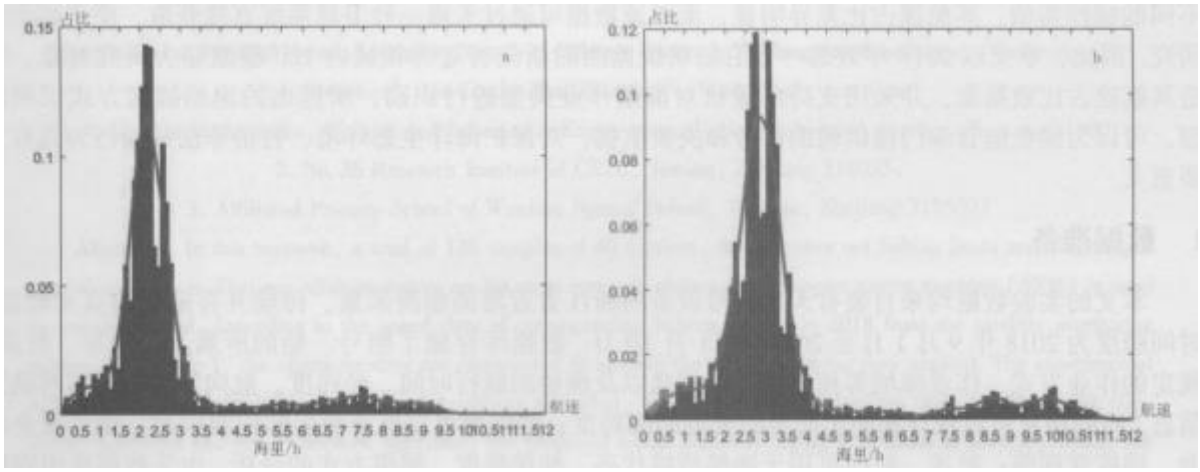

图 1 2 艘拖网渔船捕捞状态下航速占比图

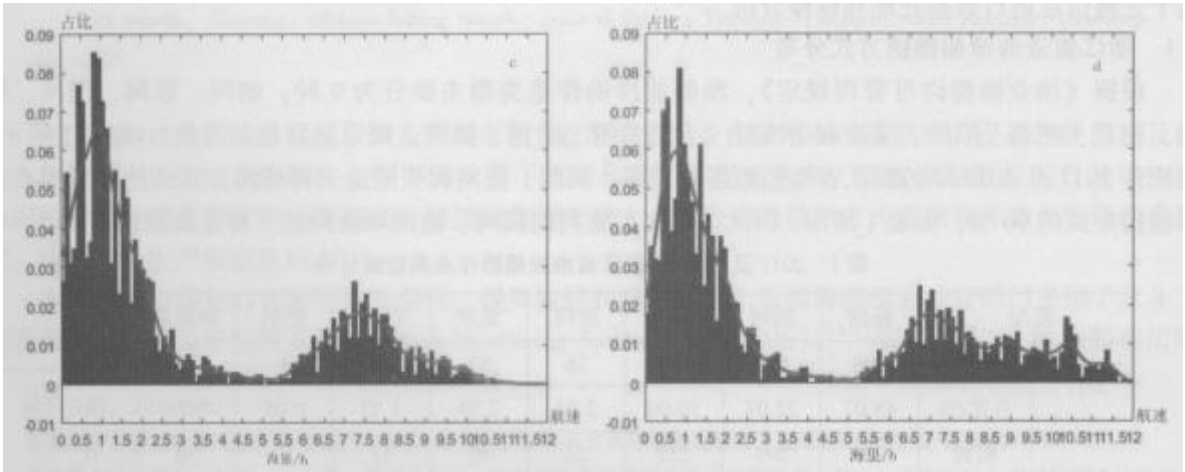

图 2 2 艘刺网渔船捕捞状态下航速占比图

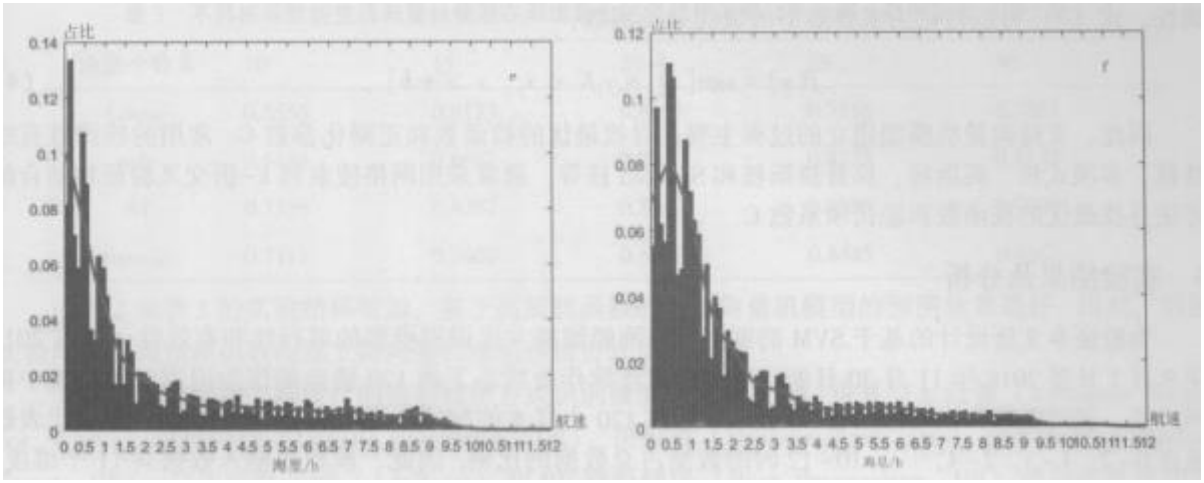

图 3 2 艘张网渔船捕捞状态下航速占比图

3 种主要捕捞方式的航速占比数据明显不同，具有较好的可区分性。因此，本文将挑选出 120 艘浙江近海渔船在 3 个月连续作业状态下的航速占比数据集，构造基于支持向量机的近海渔船捕捞方式识别模型。

2 支持向量机原理介绍

支持向量机是 Vapnik 在统计学理论基础上提出的一种机器学习算法,[15]其目标是将两类或多类不同的样本分离，而且需要使分类间隔达到最大。支持向量机在解决小样本、非线性和高维空间的实际问题中有较大优势，并被较多学者应用在海洋数据挖掘中。

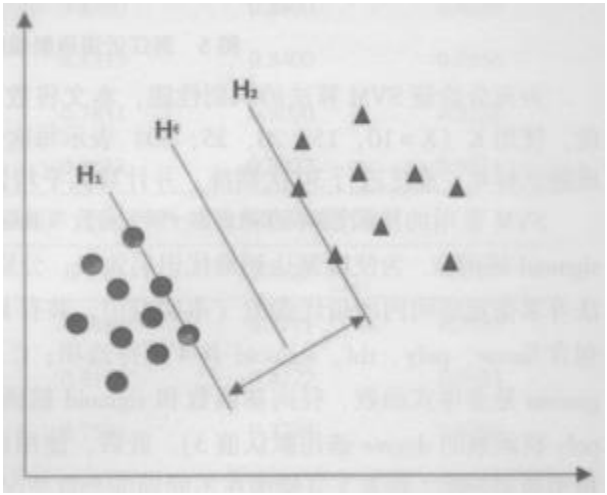

图 4 最优分类线

如图 4 所示,H1 和 H2 之间的间隔就是最大间隔，等于 2/||ω||,使间隔最大也就是||ω||最小。SVM 的目标函数为：

$$\min_{\omega,\,b}\ \frac{1}{2}\ \|\omega\|_2^2 \qquad (1)$$
$$\text{s.t. } y_i(\omega^T x_i + b) \geqslant 1,\ i = 1,\ 2,\ \cdots,\ n$$

利用拉格朗日因子法和对偶原理，上述问题又转化为如下的最优分类函数问题：

$$f(x) = \operatorname{sgn}\left[\sum_{i=0}^{n} \alpha_i y_i < x_i,\ x > + b\right] \qquad (2)$$

对于高维空间的非线性分类问题，其基本思想是通过核函数将低维的特征向量映射到高维特征空间，在新的特征空间中训练线性 SVM 模型来寻求最优分类器。另外，考虑到数据本身可能是具有噪声的非线性结构，outlier 点会严重影响支持向量机的分类性能，为使模型更加稳健，引入软间隔和惩罚项，改进后的支持向量机目标函数如下：

$$\min_{\omega,\,b}\ \frac{1}{2}\ \|\omega\|_2^2 + C\sum_i \xi_i$$
$$\text{s.t. } y_i(\omega^T x_i + b) \geqslant 1 - \xi_i,\ i = 1,\ 2,\ \cdots,\ n \qquad (3)$$
$$\xi_i \geqslant 0,\ i = 1,\ 2,\ \cdots,\ n$$

式（3）中，ξ_i是引入的松弛变量，其目的是保证在适当的惩罚项下，对错误分类的情况进行优化时依然能够收敛；C 为惩罚项系数，其目的是对错误分类的惩罚程度进行控制。根据 KKT 条件和对偶性，式（3）可以转化为求解如下的最优分类函数：

$$f(x) = \operatorname{sgn}\left[\sum_{i=0}^{n} \alpha_i y_i K < x_i,\ x > + b\right] \qquad (4)$$

因此，支持向量机模型建立的过程主要是寻找最优的核函数和正则化参数 C。常用的核函数有线性核、多项式核、高斯核、拉普拉斯核和 Sigmoid 核等，通常采用网格搜索和 k-折交叉验证相结合的方法寻找最优的核函数和惩罚项系数 C。

3 实验结果及分析

为验证本文所设计的基于 SVM 的浙江近海渔船捕捞方式识别模型的可行性和有效性，选取 2018 年 9 月 1 日至 2018 年 11 月 30 日浙江近海海域连续作业状态下的 120 艘渔船作为识别对象，其中拖网渔船、刺网渔船和张网渔船各 40 艘。计算该 120 个样本的航速占比数据集，具体数字分别代表航速在 0～2,1～3,2～4,…，10～12 时的数据占总数据的比例。因此，模型的输入数据共 11 个维度，部分样例见图 5（最后一列为样本标签，1、2、3 分别代表拖网、张网和刺网）。本文的实验环境为 Intel（R）Core（TM）i7-8550U,CPU@1.80GHz2.00GHz,RAM8.00GB,Windows10 操作系统,实现算法的语言为 Python.

x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	class
0.653949	0.090577	0.036381	0.022269	0.018620	0.018861	0.021464	0.046469	0.075553	0.060528	0.023235	2
0.581085	0.110298	0.044636	0.026221	0.021189	0.022495	0.033157	0.050511	0.068056	0.060304	0.029567	2
0.690441	0.184017	0.041927	0.015958	0.012323	0.012993	0.029333	0.046710	0.047794	0.041385	0.024885	3
0.456397	0.632705	0.381825	0.043829	0.028337	0.034834	0.046506	0.042836	0.023223	0.007791	0.002617	1
0.236458	0.416108	0.542100	0.193803	0.031388	0.040401	0.065910	0.065118	0.028589	0.006144	0.001584	1

图 5 浙江近海渔船航速占比数据集部分样本示例图

为充分论证 SVM 算法的识别性能，本文将数据集划分为各种大小的训练集和测试集。为表示方便，使用 K(K=10,15,20,25,30）表示每次选取的不同捕捞方式的渔船训练数目，其余 40-K 构成测试样本。重复进行 30 次训练，并计算其平均识别率。

SVM 常用的核函数有四种：线性核函数（linear）、多项式核函数（poly）、高斯核函数（rbf）和 sigmoid 核函数。为使模型达到最优识别效果，分别选取上述 4 种核方法训练模型，并利用网格搜索算法寻求指定空间内的最优参数（本实验中，共有 kernel、gamma、C 三个参数，其中 kernel 是核参数,包含 linear、poly、rbf、sigmoid 共 4 个待选项；C 是惩罚系数，其搜索范围从 0.1 到 2,步长为 0.1;gamma 是多项式函数、径向基函数和 sigmoid 核函数的参数,其搜索范围从 0.01 到 1,步长为 0.1;poly 核函数的 degree 选用默认值 3)。最后，使用训练集上表现最好的模型对测试集进行预测并评估模型效果。表 2 和表 3 分别为在不同训练个数情况下，不同核函数的支持向量机模型在对应训练集和测试集上的最优识别结果。

表 2 不同核函数的支持向量机模型在训练集上的最优识别率（K 为每一类渔船的训练个数）

渔船个数 K	10	15	20	25	30
Linear	0.5667	0.6222	0.7167	0.7733	0.7889
poly	0.7000	0.7111	0.7833	0.8267	0.8556
rbf	0.7667	0.8444	0.8833	0.8933	0.9111
sigmoid	0.7333	0.7778	0.8500	0.8667	0.8778

表 3 不同核函数的支持向量机模型在测试集上的最优识别率(K 为每一类渔船的训练个数)

渔船个数 K	10	15	20	25	30
Linear	0.5556	0.6133	0.6833	0.7556	0.7667
poly	0.6444	0.6667	0.7500	0.8222	0.8333
rbf	0.7556	0.8267	0.8667	0.8889	0.9000
sigmoid	0.7111	0.7600	0.8333	0.8445	0.8667

由表 2 和表 3 的实验结果可知，基于高斯核函数的支持向量机模型的预测效果最好。因此，后续实验均选择高斯核函数构建支持向量机渔船捕捞方式识别模型。

为进一步说明本文所设计的渔船捕捞方式识别模型的优越性能，将其与 k 近邻(k-Nearest Neighbor,kNN)、逻辑回归(Logistic Regression,LR)、决策树(Decision Tree,DT)、随机森林(Random Forest,RF)、自适应提升树(AdaBoost)和 BP 神经网络(Back Propagation,BP)等分类算法的实验结果作对比分析。

表 4 和表 5 为不同训练个数情况下，7 种算法分别在对应训练集和测试集上的最优识别结果。

表 4 7 种算法在训练集上的最优识别率(K 为每一类渔船的训练个数)

算法	K=10	K=15	K=20	K=25	K=30
kNN	0.6000	0.6667	0.7166	0.7333	0.7667
LR	0.7333	0.7778	0.8333	0.8400	0.8556
DT	0.6667	0.7333	0.8000	0.8133	0.8222
RF	0.7333	0.7778	0.8500	0.8800	0.8889
AdaBoost	0.7333	0.7556	0.8333	0.8400	0.8556
BP	0.6667	0.7111	0.7833	0.8000	0.8222
SVM	0.7667	0.8444	0.8833	0.8933	0.9111

表 5 7 种算法在相应测试集上的最优识别率(K 为每一类渔船的训练个数)

算法	K=10	K=15	K=20	K=25	K=30
kNN	0.5667	0.6400	0.6833	0.7111	0.7333
LR	0.7000	0.7733	0.8167	0.8222	0.8333
DT	0.6444	0.7200	0.7500	0.7778	0.8000
RF	0.7000	0.7733	0.8333	0.8667	0.8667
AdaBoost	0.7111	0.7467	0.7833	0.8000	0.8333
BP	0.6556	0.7067	0.7667	0.7778	0.8000
SVM	0.7556	0.8267	0.8667	0.8889	0.9000

从表 4 和表 5 的实验结果可知，随着训练数据规模不断增大，K 近邻、逻辑回归、决策树、随机森林、自适应提升树、BP 神经网络和 SVM7 种算法在训练集和测试集的识别性能均不断提升。

图 6 为 7 种算法在不同测试集上的最优识别率曲线图(7 种算法在训练集上的最优识别率曲线图可以类似绘出)，图 7 为在 K=30 情形下，7 种算法在近海渔船数据集上的识别准确率对比图(单位为 100%)。

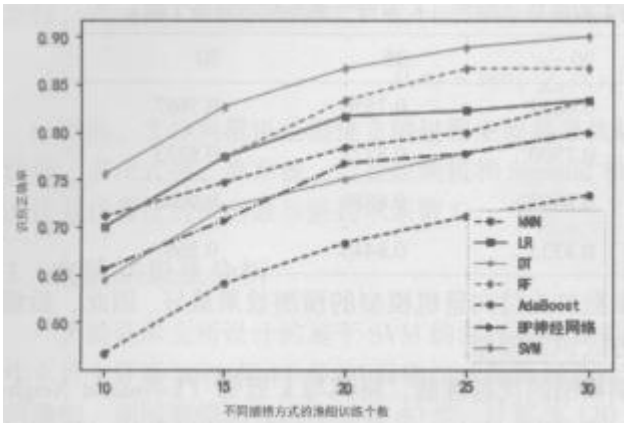

图 6 7 种算法的最优识别率对比曲线图

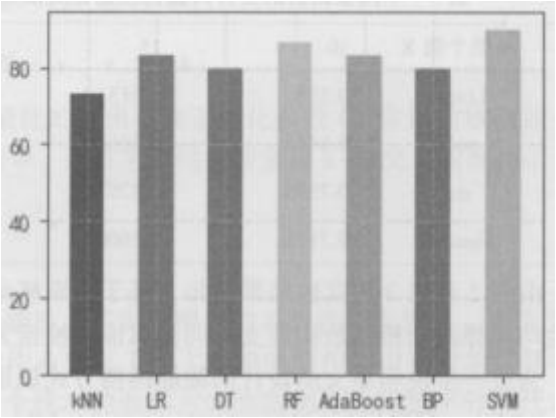

图 7 7 种算法的最优识别率对比柱状图（K=30）

从图 6 和图 7 的实验结果可以看出，在同一 K 值即固定训练规模的情况下，本文所提出的基于 SVM 的近海渔船捕捞方式识别算法均优于其他 6 种经典算法。

表 6 是在 K=30 时结合 SVM 输出结果整合的渔船捕捞方式分类精度矩阵图，矩阵的横向表示分类的结果和比例，第一行表示 30 个测试样本识别出 10 个拖网渔船，且刺网渔船和张网渔船并没有被错误地分类到拖网样本中，因此拖网的识别率为 100%;第二行表示 30 个测试样本识别出 8 只刺网渔船，其中有 2 个刺网渔船被错分为张网渔船，因此刺网渔船的识别率为 80%；第三行表示 30 个测试样本识别出 9 只张网渔船，其中有 1 只张网渔船被错分为刺网渔船，因此张网渔船的识别率为 90%。最终的分类精度由对角线上正确分类的样本个数与测试样本总数的比值决定，因此，基于 SVM 的渔船捕捞方式模型的整体识别精度为 90%。值得指出的是，拖网和张网的识别率均高于流刺网,这与 Russo 等人的分析具有一致性，主要因为刺网作业方式在空间移动上较为复杂，从而削弱了航速的规律性。

表 6 K=30 时基于 SVM 的浙江近海渔船捕捞方式分类矩阵

实际分类	判别分类			
	拖网/只	刺网/只	张网/只	识别正确率/%
10 只拖网	10	0	0	100
10 只刺网	0	8	2	80
10 只张网	0	1	9	90
总识别正确率/%	100	88.89	81.82	90.00

出现刺网和张网之间判别错误的可能原因有两个：一是本文设计的基于渔船航速占比的特征并未涵盖所有的渔船作业类型信息，如刺网和张网在捕捞过程中出现航速相似的情形；二是考察的渔船在作业时存在违规行为，导致传回的数据与捕捞许可证上登记的捕捞方式存在不符合的现象，进而影响了本文模型的识别精度。

4 结语

不合理的作业方式会破坏海洋生态环境、导致海洋渔业资源衰退。实验结果表明，本文所提模型可用于拖网、刺网和张网

渔船的捕捞方式识别，其中拖网识别正确率为100%,刺网识别正确率为80%,张网识别正确率为90%，总体识别正确率达到90%。本文的研究结果对监测浙江近海渔船捕捞行为、维护海洋生态平衡具有一定的参考价值。

由于受数据来源限制，本文讨论的对象和范围局限在浙江省近海海域的拖网、刺网和张网 3 种作业方式的渔船。后续的研究工作可以考虑增大数据样本，将其余 6 种捕捞方式纳入识别范围。另外，本文仅使用航速占比作为特征构建渔船捕捞方式识别模型，后期可考虑融入渔船的材质、长度、宽度和深度、航向、轨迹及注册登记时的文本信息等数据优化特征，进一步提升模型的识别精度，以便为保护渔业资源安全和海洋生态环境提供长远的技术支持。

致谢：本文用以建模的天通一号浙江近海渔船的船位数据，由中电科(宁波)海洋电子研究院有限公司提供，谨此致谢！

参考文献：

[1]郭媛媛.“浙”潮奔涌逐浪高——改革开放40年浙江省海洋与渔业事业发展回眸[J].浙江国土资源，2019,189(1)：23-25.

[2]王美青，徐萍，孙永朋，等.浙江省渔业资源开发利用评价及对策建议[J].浙江农业科学，2020,61(2)：368-370.

[3]冯波，陈新军，朱国平.补充型过度捕捞的确认及其对渔业管理的启示[J].资源开发与市场，2010,26(1)：20-23.

[4]闫文彦，蒋杨徽，蔡立煌，等.舟山市不同作业类型渔船捕捞产量的灰色分析[J].浙江海洋大学学报(自然科学版)，2018,37(5):96-102.

[5]张胜茂，裴凯洋，吴祖立，等.基于VMS的近海捕捞渔船出海时间与航程量化分析[J].上海海洋大学学报，2019(4)：1-13.

[6]冯艳.北斗卫星导航定位系统的船位精度计算方法LJ].现代制造技术与装备，2016(8)：20-21.

[7]张胜茂，张衡，唐峰华，等.基于船位监控系统的拖网捕捞努力量提取方法研究[J].海洋科学，2016,040(003)：146-153.

[8]WITT M J, GODLEY B J, ROSS T. A Step Towards Seascape Scale Conservation: Using Vessel Monitoring Systems (VMS) to Map Fishing Activity [J].Pios One,2007,2(10):ellll-ell19.

[9]DENG R, DICHMONT C, MILTON D, et al. Can Vessel Monitoring Dystem Data also be used to Study Trawling Intensity and Population Depletion? The Example of Australia, s Northern Prawn Fishery [J].Canadian Journal of Fisheries & Aquatic Sciences,2005,62(3)：611-622.

[10]JOO R, BERTRAND S, CHAIGNEAU A, et al. Optimization of an Artificial Neural Network for Identifying Fishing Set Positions from VMS Data： An Example from the Peruvian Anchovy Purse Seine Fishery [J]. Ecological Modelling, 2011,222(4):1048-1059.

[11]RUSSO T, PARISI A, PRORGI M, et al. When Behaviour Reveals Activity: Assigning Fishing Effort to Metiers Based on VMS Data Using Artificial Neural Networks [J].Fisheries Research,2011,111(1)：53-64.

[12]JOO R, SALCEDO O, GUTIERREZ M, et al. Defining Fishing Spatial Strategies from VMS Fata： Insights from the World’s largest Monospecific Fishery [J].Fisheries Research,2015,164：223-230.

[13]张胜茂，程田飞，王晓璇，等.基于北斗卫星船位数据提取拖网航次方法研究[J].上海海洋大学学报,2016,25(1)：135-141.

[14]郑巧玲.基于北斗卫星渔船监测系统的浙江省近海渔船捕捞方式和作业渔场分析[D].上海：上海海洋大学，2016：1-77.

[15]张学工.关于统计学习理论与支持向量机[J].自动化学报，2000(1)：36-46.