

文章编号: 1674-5566(2020)01-0121-07

DOI:10.12024/jsou.20181202471

基于 DBSCAN 的 VMS 数据定置刺网渔船网次提取方法

原作辉^{1,2}, 张胜茂¹, 樊伟¹

(1. 中国水产科学研究院东海水产研究所 农业农村部远洋与极地渔业创新重点实验室, 上海 200090; 2. 上海海洋大学 海洋科学学院, 上海 201306)

摘要: 网次是捕捞努力量计算、渔业资源调查和渔业生产管理中的重要统计参数, 定置刺网作业过程中起网状态下的航速与航行、放网时相差较大, 可通过阈值进行提取。利用 2017 年浙临渔 12870 和浙三渔 66666 定置刺网渔船 VMS 数据, 首先对其航速及相邻作业点时间差进行统计分析, 得到合理的聚类参数, 然后采用 DBSCAN (Density-Based Spatial Clustering of Application With Noises) 算法对定置刺网周围作业点进行聚类, 提取作业网次, 继而统计各网次作业持续时间与距离。以 15 min 作为合理误差范围, 将各网次起止时间与日志起网时起止时间进行比较验证。结果表明: 该方法在网次识别上效果良好, 聚类提取的各个网次起网时的起止时间准确率达到 80% 以上。

关键词: DBSCAN; 渔船监控系统; 定置刺网; 网次
中图分类号: S 975; S 127 **文献标志码:** A

我国近海渔业主要有拖网、刺网、张网和笼壶等作业方式, 定置刺网也是传统捕捞作业方式之一, 在众多地区捕捞业中占有重要的地位, 其置网时利用插杆、打桩、锚、石或砂石袋等将刺网固定于水域, 渔获物多以马鲛、鳗鱼、梭子蟹、龙头鱼为主^[1]。目前随着渔船数量的增加及捕捞技术的提高, 渔业资源下降的速度加快, 使得渔业监管变得越来越重要。

渔船监控系统 (vessel monitoring system, VMS) 最早出现在 19 世纪 90 年代, 其初衷是为了监视渔船的位置, 确保渔船安全和及时救援, 随后由于其定位的准确性和近实时性, 因而在渔船捕捞努力量估算^[2-3] 和渔业资源评估^[4-5] 上得到了广泛应用。北斗卫星导航系统在渔业中的应用已经历时多年, 目前我国共计近 6 万多艘渔船配备北斗卫星船载终端设备。VMS 数据主要包括经纬度、航向和航速等信息, 单纯的点数据只能简单反映渔船的空间位置, 并不能够直接体现出渔船的捕捞状态、作业规律、捕捞强度等信息, 随着研究的逐步发展, 众多学者对大量的船位数

据进行分析挖掘, 在渔船状态判别^[6-8]、轨迹重塑^[9-10] 及渔场分析^[11-13] 等方面取得了一系列成果, 研究方法涉及样条插值、神经网络、贝叶斯、隐马尔可夫和空间聚类等, 其中空间聚类主要应用于作业状态划分及渔船模式识别方面。JOO 等^[14] 首先利用 VMS 数据对鳀鱼围网渔船进行轨迹重构, 根据航行时间、捕捞时间和离岸距离 3 个参数通过层次聚类方法将航次分为 4 类, 并分析了渔船不同捕捞模式与区域政策、船队管理及船长习惯的关系。SU 等^[15] 基于 DBSCAN (Density-based Spatial Clustering of Application with Noise) 聚类方法开发了 VMS 渔船实时监控系系统, 将 10 n mile 内数量在 3 艘以上的渔船信息发送给海岸警备队, 辅助监察和渔业管理部门。ZHANG 等^[16] 基于 DBSCAN 和 K-Mean 算法提出了一种渔船捕捞行为识别模型, 将航行轨迹划分为航行、捕捞和停泊 3 种模式, 继而建立无监督行为学习模型, 将测试结果与传统的高斯混合模型识别方法进行对比, 验证了该模型的准确性。

DBSCAN 是一种基于密度的空间数据聚类算

收稿日期: 2018-12-14 修回日期: 2019-02-28

基金项目: 国家自然科学基金(31772899); 上海市自然科学基金(17ZR1439800)

作者简介: 原作辉(1994—), 男, 硕士研究生, 研究方向为数据挖掘和渔业 GIS。E-mail: rsyuanzh@126.com

通信作者: 张胜茂, E-mail: ryshengmao@126.com

法,最早由 ESTER^[17]等提出,旨在找到由低密度分割的高密度区域,与基于距离的聚类算法相比,它的优点在于可以找到任意形状的聚类,能够较好识别出呈直线分布的定置刺网作业区域。利用浙江省定置刺网渔船 VMS 数据,基于 DBSCAN 算法识别渔船作业网次,计算出渔船作业网长,结果可作为沿海定置刺网渔船的捕捞努力量,并对其进行监督管理,为渔业资源保护和管理提供参考。

1 材料与方 法

1.1 VMS 数据及定置刺网作业特点

选用 2017 年浙江省定置刺网渔船浙临渔 12870 和浙三渔 66666 北斗 VMS 数据进行分析。船位数据来自北斗民用分离商,静态信息主要包括船名、船长和作业方式等,动态信息主要包括渔船作业时间、经纬度、航向和航速,数据时间分辨率为 3 min,空间分辨率为 10 m。

定置刺网渔船作业时通常以较大速度行至渔场做放网准备,航行速度为 4.5 m/s 以上;放网时航速较快,一般大于 3 m/s,约 1 h 左右,随后漂流等待或沿网查看。一段时间后开始起网,起网时速度较慢且持续时间较长,此时航速保持在 2.5 m/s 以下。在起网过程中若渔获量较少则采取边收边放网的作业方式,否则先起网,取鱼后放网。定置刺网网具一般在近海敷设,与水流方向垂直或呈一定夹角。

1.2 网次提取方法

1.2.1 VMS 数据预处理

浙临渔 12870 功率为 255 kW,2017 年全年船位点为 82 118 个;浙三渔 66666 功率为 396 kW,2017 年全年船位点为 103 135 个。由于渔船靠岸或海上停泊时同样发送船位数据,因此产生较多速度为零的点,在统计时首先删除该记录,确保渔船处于航行或捕捞状态。最后得到浙临渔 12 870 共 35 322 条有效数据,捕捞作业时间为 1—4 月、8—12 月;浙三渔 66666 共 43 722 条有效数据,作业时间为 1 月、8—12 月。定置刺网渔船作业主要包含航行、放网、漂流和起网等状态,在起网作业时航速较低,与其他状态差别较大,通过航速变化图可以清晰地进行区分,因此可以通过阈值分割方法得到起网状态下的点位数据,即根据刺网渔船航速统计船位点在各航速下出

现的频数,找到航速低值区域频数的峰值,获取峰值两侧的谷值作为起网的航速阈值,提取该航速范围内数据。

经提取后的船位点主要是各网次起网状态下的点数据,继而计算相邻点间时间间隔,单位为分钟。按照与航速统计时相同的方法得到时间间隔的变化图。此时时间间隔的取值主要集中为两部分:一部分是同一网次内时间间隔,其特征是时间间隔值小且频次高;另一部分是网次间时间间隔,其特征是时间间隔值大且频次低。由于两部分的频次相差较大,因此对所得结果做对数变换,在不改变数据性质的前提下压缩变量的尺度。随后观察得到两部分的交界点,并作为网次间时间间隔参数带入 DBSCAN 算法。

1.2.2 DBSCAN 算法

DBSCAN 作为一种基于密度的聚类算法已经得到广泛应用,它可以发现任意形状的聚类、自动确定聚类数以及有效剔除噪声点,对于定置刺网类似直线分布的模式能够较好识别,其中心思想是在聚类空间中某点一定范围内,若所包含对象的数目不小于某一给定阈值,则将当前对象与附近对象标记为一个类,对该类中未检索的对象进行相同判断,从而对该类进行扩展。常规 DBSCAN 中主要涉及两个参数:邻域(Eps)和邻域密度阈值(MinPts)。由于定置刺网的作业点聚集在网具周围,不同月份的数据往往叠置在一起,因此在判断网次时增添网次间时间间隔(T)进行约束。

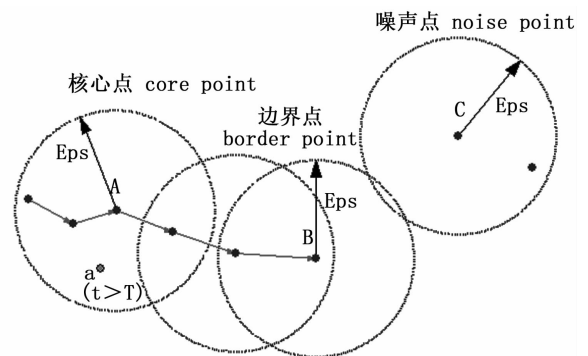


图 1 网次提取概念示意图

Fig. 1 Conceptual diagram of hauls extraction

假设 MinPts 为 3,算法执行中重要一步是判断检索点与 Eps 邻域内其他点的时间间隔,统计有效点数。核心点表示在该点的 Eps 邻域内存

在不少于 $MinPts$ 的样本点数,如点 A 所示,其邻域内包括 5 个点(含点 A),由于 a 点时间间隔大于 T ,实际有效点数为 4,因此判断为核心点。边界点表示该点在某核心点邻域内,但其本身不是核心点,如点 B 所示。噪声点为邻域内有效点数不满 3 个且未被包含在核心点邻域内的点,如点 C 所示。理论上核心点多为网次内部点,边界点为网次起止点。见图 1。

1.3 验证方法

2017 年浙临渔 12870 和浙三渔 66666 日志数据由浙江省海洋水产所提供。浙江省限额捕捞政策 9 月份开始实施,日志记录为 9—12 月。生产数据包括船名、捕捞状态(起网或放网)、起止时间、经纬度、排数和产量。通过将 DBSCAN 算法提取结果与日志中起网起止时间进行对比,

统计两者的时间差,若两者相差在 15 min 以内,则证明该网次提取结果准确。一般而言,定置刺网渔船在起网开始前及结束后通常会进行网具的整理工作,此时航速也较低,位移量较小,可能被视为网次的起止点,因此将 15 min 定义为合理误差范围。

2 结果与讨论

2.1 渔船状态分析

以 0.1 m/s 间隔统计 1 年内不同航速下船位点个数,可知渔船出海作业过程中航速主要分布在 0.3 ~ 7.0 m/s,航速主要集中在两段:0.5 ~ 2.0 m/s,为渔船起网时航速;3 ~ 7 m/s,为渔船航行或放网时航速(图 2)。

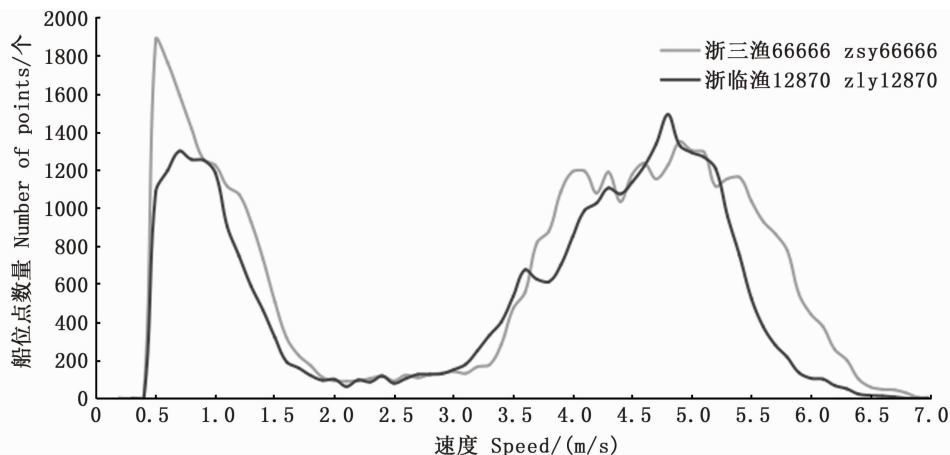


图 2 渔船各航速下船位点频数分布图

Fig. 2 Number of vessel points along with speed

根据定置刺网渔船作业特点,其放网和航行时速度难以区分,而起网时速度与其他状态相差较大,因此首先根据速度阈值提取起网状态下的作业点,以 0.5/2.0 m/s 为起网航速阈值对 VMS 数据进行提取,并计算相邻作业点的时间差,统计各时间间隔下船位点频数并取自然对数(图 3)。

由图 3 可知,两艘渔船呈现相似的分布,作业点数据时间间隔主要集中在 0 ~ 30 min,30 min 以内对应频数呈现下降的趋势,30 min 后时间间隔跨度较大,且频次主要在低值范围内波动。定置刺网作业往往是在起网结束后放网或继续航行至下一网次,由于提取的是起网状态下的作业点数据,因此时间间隔较大的值多表示前一网次起网结束点和后一网次起网开始点的时间差,且

网次间时间间隔数值不尽相同,任一时间间隔对应的频次不高,结合这一特点可知,相邻网次间时间间隔主要集中在 30 min 以后,因此以 30 min 为阈值区分网次。

2.2 渔船网次提取

结合北斗数据时间分辨率及速度提取阈值,选择以 3 min、2 m/s 为阈值得到 Eps 为 360 m;检索点 3 min 前后应各包含一点,因此 $MinPts$ 为 3;时间间隔阈值 T 为 30 min。使用 Java 编程实现 DBSCAN 算法,提取聚类后每一类的时间最小值与最大值作为该网次起网时开始点和结束点,计算各个网次的持续时间和累计网长,存入 MySQL 数据库。表 1 描述了浙临渔 12870 和浙三渔 66666 各月网次提取结果。

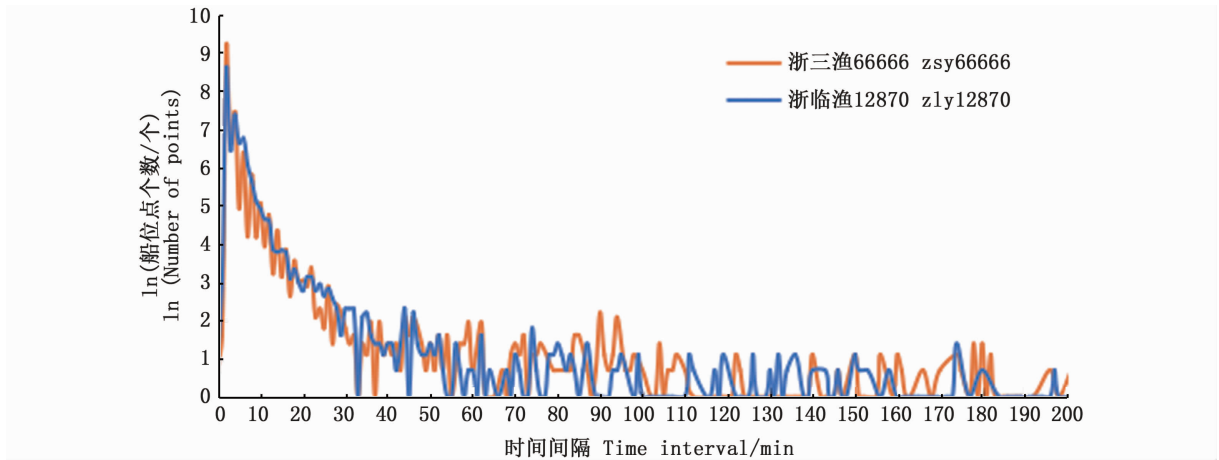


图3 渔船各时间间隔下船位点频数分布图

Fig.3 Number of vessel points along with time interval

表1 2017年定置刺网渔船各月网次统计

Tab.1 The monthly hauls statistics of set gillnet vessels in 2017

月份 Month	浙临渔 12870 zly12870			浙三渔 66666 zsy66666		
	网次 Number of hauls	累计起网时间 Accumulated time/h	累计网长 Accumulated length/km	网次 Number of hauls	累计起网时间 Accumulated time/h	累计网长 Accumulated length/km
1月 Jan.	14	29.81	100.6	39	78.27	277.3
2月 Feb.	18	60.52	127.9	-	-	-
3月 Mar.	6	20.84	44.6	-	-	-
4月 Apr.	41	169.10	323.8	-	-	-
8月 Aug.	44	120.11	290.3	81	185.96	345.1
9月 Sep.	25	69.36	157.1	41	133.47	251.8
10月 Oct.	28	85.00	195.0	56	93.49	261.8
11月 Nov.	11	41.13	105.1	39	79.70	267.3
12月 Dec.	14	36.03	126.1	59	86.77	322.4
总计 Total	201	631.9	1 470.5	315	657.66	1 725.7

2.3 渔船网次验证

对2017年浙临渔12870和浙三渔66666两艘渔船12月份提取的网次进行验证,网次提取结果如图4所示,12月份两艘渔船集中在舟山渔

场进行作业,浙临渔12870作业网次较为稀疏,浙三渔66666作业网次较为密集,绿色标记表示起网开始位置,红色标记表示起网结束位置。

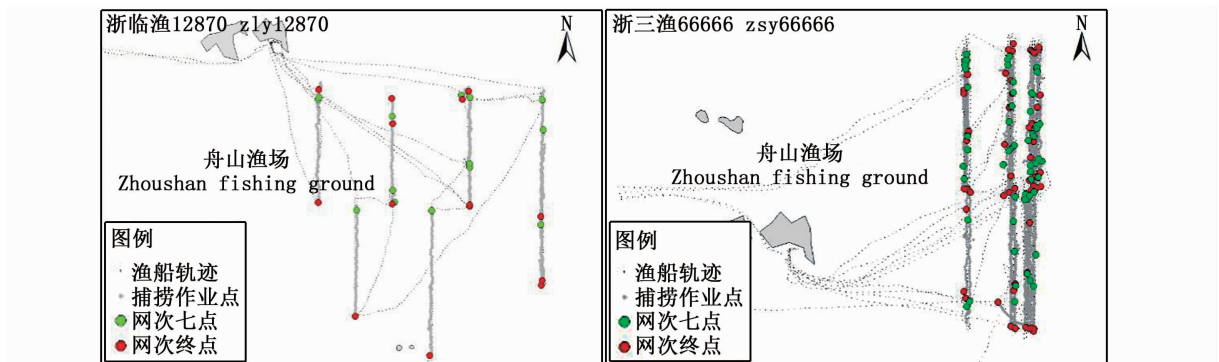


图4 12月份网次提取结果

Fig.4 Results of haul extractions in December

在对数据进行网次聚类时,由于非作业状态下渔船可能也存在连续的低速航行,其速度与渔船起网速度位于同一范围,因此在聚类时也被识别为 1 个网次,由于定置刺网敷设长度较大,网次内部包含点数多为数十个,目前的做法是删掉数据点小于 10 的网次,提高网次准确率;此外由于数据存在丢失现象,同一网次的点中若前后两点间隔距离过大,则容易被判断为两个网次,目前的解决办法是针对网次提取结果,依次计算相邻网次中前一网次结束和后一网次开始的作业点的速度均值、时间间隔及距离,若速度均值与时间的乘积小于距离,则进行网次合并,确保了定置刺网渔船网次提取方法的可行性。

表 2 将 DBSCAN 提取的各网次起止点时间与日志记录中起网时起止点时间进行对比验证。

算法提取的起网开始时间与实际起网开始时间差在 5 min 以内的占 73.9%,10 min 以内的占 78.0%,15 min 以内的占 82.1%,15 min 以上的占 17.9%;算法提取的起网结束时间与实际起网结束时间差在 5 min 以内的占 80.8%,10 min 以内的占 83.6%,15 min 以内的占 84.9%,15 min 以上的占 15.1%。差值在 15 min 以上主要由于船位数据的缺失:一方面起网开始或结束时船位点的缺失会直接影响提取结果;另一方面起网过程中船位点的缺失会对聚类结果中网次数量和起网起止时间点的准确性带来影响,若丢失点中航速和航向较为稳定,可以通过网次合并方法进行解决,若发生较大波动,出现前后两点距离大于 360 m,则仍不能被合并为同一网次,使得提取结果产生较大误差。

表 2 定置刺网渔船网次验证

Tab.2 Hauls verification of set gillnet vessels

渔船 Vessel	网次数 Number of hauls	网次起点时间差 Time difference of haul beginning points				网次终点时间差 Time difference of haul ending points			
		5 min	10 min	15 min	15 min 以上	5 min	10 min	15 min	15min 以上
		浙临渔 12870 zly12870	14	7	2	1	4	12	0
浙三渔 66666 zsy66666	59	47	1	2	9	47	2	1	9
合计 Total	73	54	3	3	13	59	2	1	11
百分比 Percent/%	100	73.9	78.0	82.1	17.9	80.8	83.6	84.9	15.1

3 结论

定置刺网渔船捕捞过程中起网时航速与其他状态下相差较大,通过分析航速变化可以直接提取起网作业点,基于密度的 DBSCAN 算法可以识别出满足条件的起网区域,继而根据时间间隔来区分并提取网次。根据 DBSCAN 算法和 VMS 船位数据提取了 2017 年浙江省两艘定置刺网渔船全年网次,统计网次起网时起止时间与实际值的时间差,在允许的误差范围内准确率达到 80% 以上,验证了该方法的可行性,所得结果可用于计算定置刺网渔船的捕捞持续时间及放网距离,并作为捕捞努力量分析捕捞经济效益,也可作为统计量参与限额捕捞和渔业资源调查评估。

参考文献:

[1] 石建高. 捕捞渔具准入配套标准体系研究[M]. 北京: 中国农业出版社, 2017.
SHI J G. Study on access standard system of fishing gear

[M]. Beijing: China Agriculture Press, 2017.
[2] VERMARD Y, RIVOT E, MAHÉVAS S, et al. Identifying fishing trip behaviour and estimating fishing effort from VMS data using Bayesian Hidden Markov Models[J]. Ecological Modelling, 2010, 221(15): 1757-1769.
[3] 张胜茂, 杨胜龙, 戴阳, 等. 北斗船位数据提取拖网捕捞努力量算法研究[J]. 水产学报, 2014, 38(8): 1190-1199.
ZHANG S M, YANG S L, DAI Y, et al. Algorithm of fishing effort extraction in trawling based on Beidou vessel monitoring system data[J]. Journal of Fisheries of China, 2014, 38(8): 1190-1199.
[4] MULLOWNEY D R, DAWE E G. Development of performance indices for the Newfoundland and Labrador snow crab (*Chionoecetes opilio*) fishery using data from a vessel monitoring system[J]. Fisheries Research, 2009, 100(3): 248-254.
[5] STELZENMÜLLER V, MAYNOU F, BERNARD G, et al. Spatial assessment of fishing effort around European marine reserves: implications for successful fisheries management [J]. Marine Pollution Bulletin, 2008, 56(12): 2018-2026.
[6] WANG Y, WANG Y B, ZHENG J. Analyses of trawling

- track and fishing activity based on the data of vessel monitoring system (VMS): a case study of the single otter trawl vessels in the Zhoushan fishing ground[J]. *Journal of Ocean University of China*, 2015, 14(1): 89-96.
- [7] 张胜茂, 程田飞, 化成君, 等. 北斗卫星数据提取耙刺渔船作业状态方法研究[J]. *渔业信息与战略*, 2015, 30(4): 293-300.
- ZHANG S M, CHENG T F, HUA C J, et al. Extraction method of rake thorn status based on data from Beidou vessel monitoring system [J]. *Fishery Information & Strategy*, 2015, 30(4): 293-300.
- [8] 张胜茂, 唐峰华, 靳少非, 等. 基于北斗卫星数据的拖网渔船状态与网次提取[J]. *渔业信息与战略*, 2015, 30(3): 205-211.
- ZHANG S M, TANG F H, JIN S F, et al. Trawler state and net times extraction based on data from Beidou vessel monitoring system [J]. *Fishery Information & Strategy*, 2015, 30(3): 205-211.
- [9] HINTZEN N T, PIET G J, BRUNEL T. Improved estimation of trawling tracks using cubic Hermite spline interpolation of position registration data[J]. *Fisheries Research*, 2010, 101(1/2): 108-115.
- [10] RUSSO T, PARISI A, CATAUDELLA S. New insights in interpolating fishing tracks from VMS data for different métiers[J]. *Fisheries Research*, 2011, 108(1): 184-194.
- [11] JENNINGS S, LEE J. Defining fishing grounds with vessel monitoring system data[J]. *ICES Journal of Marine Science*, 2012, 69(1): 51-63.
- [12] RUSSO T, PARISI A, CATAUDELLA S. Spatial indicators of fishing pressure: preliminary analyses and possible developments[J]. *Ecological Indicators*, 2013, 26: 141-153.
- [13] 原作辉, 杨东海, 樊伟, 等. 基于卫星 AIS 的中西太平洋金枪鱼延绳钓渔场分布研究[J]. *海洋渔业*, 2018, 40(6): 649-659.
- YUAN Z H, YANG D H, FAN W, et al. On fishing grounds distribution of tuna longline based on satellite automatic identification system in the western and central Pacific[J]. *Marine Fisheries*, 2018, 40(6): 649-659.
- [14] JOO R, SALCEDO O, GUTIERREZ M, et al. Defining fishing spatial strategies from VMS data: insights from the world's largest monospecific fishery[J]. *Fisheries Research*, 2015, 164: 223-230.
- [15] SU Y Y, CHANG S J. Spatial cluster detection for the fishing vessel monitoring systems [C]//*Proceedings of OCEANS 2008 - MTS/IEEE Kobe Techno-Ocean*. Kobe, Japan: IEEE, 2008: 1-4.
- [16] ZHANG J L, GENG J L, WAN J, et al. An automatically learning and discovering human fishing behaviors scheme for CPSCN[J]. *IEEE Access*, 2018, 6: 19844-19858.
- [17] ESTER M, KRIEGEL H P, SANDER J, et al. A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise [C]//*Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*. Portland, Oregon: ACM, 1996: 226-231.

Method of set gillnet hauls extraction based on DBSCAN and VMS data

YUAN Zuohui^{1,2}, ZHANG Shengmao¹, FAN Wei¹

(1. Key Laboratory of Oceanic and Polar Fisheries, Ministry of Agriculture and Rural Affairs, East China Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Shanghai 200090, China; 2. College of Marine Sciences, Shanghai Ocean University, Shanghai 201306, China)

Abstract: Vessel monitoring system (VMS), as a comprehensive application system integrating global satellite positioning, electronic map, network communication and database technology, aims to acquire the ship's position and operating status information in near-real time, and transmit them to the onshore monitoring center to realize the information interaction. More than sixty thousand fishing vessels have installed VMS so far. The VMS data mainly contains time, position, speed, direction and rate of turn with a temporal resolution of 3 minutes and a spatial resolution of 10 meters which can be analyzed deeply using data mining technology to identify the status of vessels, calculate fishing effort, search fishing ground and so on. DBSCAN is a density-based spatial clustering algorithm designed to find high-density regions segmented by low-density. Compared with distance-based clustering algorithms, it can identify noise data and find clusters of arbitrary shapes. In this paper, we used VMS data of Zhelinyu 12870 and Zhesanyu 66666 in 2017 and DBSCAN to extract the hauls of set gillnet. The speed of hauling gillnet is quite different from that during tracking or setting the net, so they could be extracted them by threshold. Firstly, we obtained the speed threshold of vessel's fishing condition by the statistics of navigational speed and extracted the points whose speed is in the range of threshold value. Secondly, the time difference of adjacent points was carried out to obtain time interval between hauls. Thirdly, the DBSCAN algorithm was used to cluster the fishing points around the set gillnet, which determines the starting time and the ending time of each fishing haul. The difference of the starting time and the ending time of each haul between clustering result and fishing log were calculated, and 15 minutes were taken as a reasonable error-tolerant range. The comparisons showed that the accuracy of the hauls was above 80%.

Key words: DBSCAN; vessel monitoring system; set gillnet; haul