

DOI: 10.19663/j.issn2095-9869.20200315002

http://www.yykxjz.cn/

郑芯瑜, 刘必林, 孔祥洪, 王雪辉. 基于 *K*-means 动态聚类的鸢乌贼角质颚模式识别. 渔业科学进展, 2021, 42(4): 64–72
ZHENG X Y, LIU B L, KONG X H, WANG X H. Pattern of recognition beaks in *Sthenoteuthis oualaniensis* based on *K*-means dynamic clustering. Progress in Fishery Sciences, 2021, 42(4): 64–72

基于 *K*-means 动态聚类的鸢乌贼 角质颚模式识别*

郑芯瑜¹ 刘必林^{1,2①} 孔祥洪¹ 王雪辉³

(1. 上海海洋大学海洋科学学院 上海 201306; 2. 大洋渔业资源可持续开发教育部重点实验室 上海海洋大学
国家远洋渔业工程技术研究中心 农业农村部大洋渔业开发重点实验室 农业农村部大洋渔业资源环境
科学观测实验站 上海 201306; 3. 中国水产科学研究院南海水产研究所 广东 广州 510300)

摘要 本研究采用 *K*-means 动态聚类算法, 对 2014—2019 年间采集于西北印度洋、热带东太平洋、中国南海的鸢乌贼(*Sthenoteuthis oualaniensis*)样本的角质颚进行识别。基于 *K*-means 动态聚类算法能够很好地区分来自 3 个海区的鸢乌贼, 对数据进行 *z*-score 标准化后, 任选 2 维角质颚形态学参数以曼哈顿距离和欧氏距离进行 *K*-means 动态聚类分析, 总正确区分率分别为 86.7% 和 88.7%。*K*-means 动态聚类算法对于鸢乌贼角质颚的识别有很大的参考价值, 后续改进优化 *K*-means 算法使其具有普适性, 将会提高鸢乌贼种群的识别能力。

关键词 鸢乌贼; 角质颚; 模式识别; 曼哈顿距离; 欧氏距离

中图分类号 S917.4 **文献标识码** A **文章编号** 2095-9869(2021)04-0064-09

头足类作为海洋食物链中重要的一环, 研究其分类状况可为其他捕食者的食性分析提供有力依据, 同时, 为进一步研究海洋食物网关系、生态系统能量流动情况以及资源评估提供重要基础(刘必林等, 2009; 范江涛等, 2015)。头足类的软组织(如胴体、腕足部等)与硬组织(如角质颚、内壳、耳石等)均可用作其种群和种类的判别, 一般认为硬组织比软组织更有效(Borges, 1990; 胡飞飞等, 2017)。作为头足类硬组织之一的角质颚因其结构稳定、耐腐蚀、特征信息丰富等而被广泛用于头足类的种类和种群判定(Jackson *et al.*,

1997)。Vega 等(2002)基于角质颚长度成功区分来自太平洋和大西洋的巴塔哥尼亚枪乌贼(*Loligo gahi*)的种群。方舟等(2014)研究发现, 北太平洋柔鱼的 2 个群体在耳石与角质颚形态学特征上存在显著差异。刘必林等(2015)根据角质颚长度的差异成功判别中国近海 5 种常见经济头足类。方舟等(2019)利用角质颚实现东海和黄海的头足类的科类判别。

鸢乌贼(*Sthenoteuthis oualaniensis*)广泛分布于印度洋—太平洋海域, 其地理种群结构复杂(刘必林等, 2010)。Liu 等(2019)采用传统的逐步判别法从胴体和

* 国家重点研发计划(2019YFD0901404)、国家自然科学基金面上项目(41876141)、上海市“浦江人才”计划项目(18PJ1404100)、上海市高校特聘教授“东方学者”岗位计划项目(0810000243)和上海市科技创新行动计划(19DZ1207502)共同资助 [This work was supported by National Key Research and Development Program of China (2019YFD0901404), National Natural Science Foundation of China (41876141), Shanghai Pujiang Program Under Contract (18PJ1404100), Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning under Contract (0810000243), and Shanghai Science and Technology Innovation Action Plan (19DZ1207502)]. 郑芯瑜, E-mail: 1187019761@qq.com

① 通讯作者: 刘必林, 教授, E-mail: bl-liu@shou.edu.cn

收稿日期: 2020-03-15, 收修改稿日期: 2020-04-08

角质颚形态差异对鸢乌贼的地理种群进行划分,但误判率较高。由于模式识别中的 K-means 动态聚类方法简单高效,为此,本研究引入该算法对鸢乌贼角质颚的形态学参数进行聚类分析,期待为头足类种群识别提供一种新的方法。

1 材料与方法

1.1 样品采集采样

样本于2014—2019年间分别由中国灯光围网渔船、鱿钓渔船、敷网渔船采集于西北印度洋(59°~64°E, 13°~20°N)、热带东太平洋(110°~119°E, 5°S~1°N)和中国南海(110°~118°W, 9°~21°N)(表1)。

样本在实验室解冻后,测量其胴长(mantle length,

ML, 精确到 1 mm)和体质量(body weight, BW, 精确到 1 g)。角质颚提取后,放入 75%酒精溶液中保存。分别从西北印度洋海区选取鸢乌贼大型群、热带东太平洋中型群及中国南海中型群(有发光器)的角质颚,选取胴长范围为 80~500 mm 的 150 对角质颚,使用精度为 0.01 mm 的游标卡尺对上颚和下颚各 5 个特征进行测量,包括上头盖长(upper hood length, UHL)、上脊突长(upper crest length, UCL)、上喙长(upper rostrum length, URL)、上侧壁长(upper lateral wall length, ULWL)、上翼长(upper wing length, UWL)、下头盖长(lower hood length, LHL)、下脊突长(lower crest length, LCL)、下喙长(lower rostrum length, LRL)、下侧壁长(lower lateral wall length, LLWL)和下翼长(lower wing length, LWL),测量标准参照文献(方舟等, 2014)(图 1)。

表 1 采样信息

Tab.1 Sampling information

海区 Sea area	经纬度 Latitude and longitude	采样日期 Sampling date	样本量 Sample number/ind.	胴长 Mantle length/mm	体质量 Weight/g
西北印度洋 Northwest Indian Ocean	59°~64°E 13°~20°N	2019.06	50	120~443	105~5564
热带东太平洋 Tropical eastern Pacific	110°~119°E 5°S~1°N	2014.04—2014.05	50	119~351	45~1975
中国南海 South China Sea	110°~118°W 9°~21°N	2017.01—2017.05	50	80~213	16~460

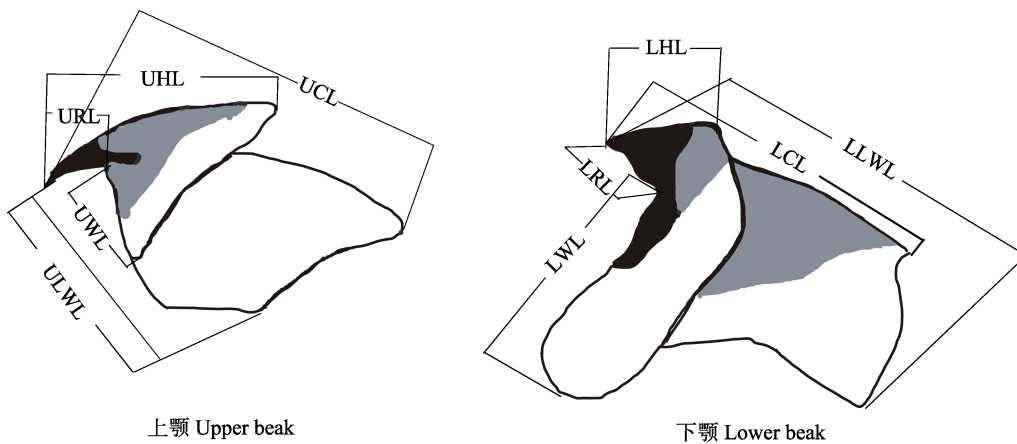


图 1 角质颚外部形态测量示意图

Fig.1 Scheme of morphometric measurement for beak

UHL: 上头盖长; UCL: 上脊突长; URL: 上喙长; ULWL: 上侧壁长; UWL: 上翼长; LHL: 下头盖长;
LCL: 下脊突长; LRL: 下喙长; LLWL: 下侧壁长; LWL: 下翼长

UHL: Upper hood length; UCL: Upper crest length; URL: Upper rostrum length; ULWL: Upper lateral wall length;
UWL: Upper wing length; LHL: Lower hood length; LCL: Lower crest length; LRL: Lower rostrum length;
LLWL: Lower lateral wall length; LWL: Lower wing length

1.2 角质颚数据标准化

为排除个体大小对角质颚尺寸的影响,采用

z-score 标准化将原始角质颚数据值映射到[0,1]之间,该标准化方法保留了原来数据中存在的关系,消除量纲和数据取值范围影响,其计算公式如下:

$$X_{ij} = \frac{x_{ij} - u}{\delta} \quad (1)$$

式中, u 为各角质颚形态学数据的均值; δ 为各角质颚形态学数据的标准差; x_{ij} 为第 i 行 j 列的角质颚数据, i 为样本总量, $i=1,2,\dots,150$, j 为角质颚上头盖长、上脊突长等 10 个形态学参数, $j=1,2,\dots,10$; X_{ij} 为标准化后的角质颚数据。

1.3 K-means 聚类建模

第 1 步, 以标准化后的 150 组角质颚样本总量为行, 以 10 个形态学参数为列, 形成 150×10 的总矩阵。第 2 步, 根据 K-means 聚类原则, 将预处理完毕的角质颚数据按照采样的海区来源分为 3 类, 即 $K=3$ 。第 3 步, K-means 算法在数据集中随机选择 1 个点作为第 1 个初始类的中心点, 接着选择距离该点最远的点作为第 2 个初始类簇中心点, 最后选择距离前 2 个点的最近距离最大的点作为第 3 个初始类簇的中心点, 最终选定 3 个初始聚类的中心即 C_1, C_2, C_3 (曾俊, 2020)。第 4 步, 选择任意 2 组角质颚形态学参数 x_{ij} ($i=1,2,\dots,150; j=1,2,\dots,10$) 按曼哈顿距离 (或欧氏距离、马氏距离等) 归给某一个距离最近的聚类中心 C_k ($0 < k < 3$) (曾俊, 2020; 王同兴等, 2010), 与聚类中心最小距离为:

$$\|x_{ij} - C_k\| = \min_{1 \leq l < 3} \|x_{ij} - C_l\| \quad (2)$$

其中, 曼哈顿距离为:

$$L_1(x_a, x_b) = \sum_{l=1}^n |x_a^{(l)} - x_b^{(l)}| \quad (3)$$

式中, $n=10$, 即 10 维角质颚形态学数据; x_a 和 x_b 分别为 10 维角质颚形态学参数矩阵的任一维 (150×1)。

其中, 欧氏距离为:

$$L_2(x_a, x_b) = \left(\sum_{l=1}^n |x_a^{(l)} - x_b^{(l)}|^2 \right)^{\frac{1}{2}} \quad (4)$$

第 5 步, 重新计算各个新类的聚类中心值 $C_{k'}$:

$$C_{k'} = \frac{1}{N_k} \sum_{x \in S_k} x \quad (k=1,2,3) \quad (5)$$

式中, N_k 是第 k 个聚类集合 S_k 中所包含的角质颚样本个数; 迭代逐次修正更新聚类的中心点, 当没有改变任何点的分类情况及聚类中心, 则算法收敛; 否则继续修正。算法收敛的条件为:

$$C_k^{n+1} = C_k^n \quad (\text{对所有 } k=1,2,3 \text{ 成立}) \quad (6)$$

若上式成立, 则算法收敛, 计算结束, 利用 MATLAB 软件编程实现。

2 结果与分析

2.1 角质颚大小

未经标准化的原始角质颚形态参数分析显示, 西北印度洋海区整体角质颚最大, 其次为热带东太平洋、中国南海鸢乌贼角质颚最小 (表 2)。经过标准化后的角质颚数据, 由曼哈顿距离法和欧氏距离分析显示, 西北印度洋海区鸢乌贼角质颚最大, 热带东太平洋次之, 中国南海最小 (图 2 和图 3)。

2.2 聚类中心

欧氏距离与曼哈顿距离法分析显示, 对于 10 组角质颚形态学参数, 任选 2 组, 各海区的聚类中心位置均会相应变化。以曼哈顿距离为例, 西北印度洋海区、

表 2 原始鸢乌贼地理种群形态参数/mm
Tab.2 Morphologic indices of primitive *S. oualaniensis*/mm

形态参数 Morphologic indices	西北印度洋 Northwest Indian Ocean			热带东太平洋 Tropical eastern Pacific			中国南海 South China Sea		
	最小值	最大值	平均值±标准差	最小值	最大值	平均值±标准差	最小值	最大值	平均值±标准差
	Min	Max	Mean±SD	Min	Max	Mean±SD	Min	Max	Mean±SD
上头盖长 UHL	13.0	17.2	14.1±0.9	9.2	13.0	11.6±0.8	8.3	10.9	9.8±0.6
上脊突长 UCL	15.9	20.6	17.5±0.9	12.9	19.0	15.2±1.2	10.3	13.7	12.3±0.8
上喙长 URL	3.8	6.6	5.0±0.7	3.1	4.9	3.8±0.5	2.2	4.3	3.4±0.4
上侧壁长 ULWL	10.0	14.6	12.3±1.1	8.7	12.3	11.0±0.8	7.9	11.0	9.7±0.7
上翼长 UWL	3.2	5.3	4.5±0.5	4.0	5.7	4.7±0.6	2.4	3.8	3.3±0.4
下头盖长 LHL	3.4	5.4	4.2±0.4	3.0	4.2	3.6±0.3	2.6	4.0	3.3±0.3
下脊突长 LCL	6.8	10.3	8.6±0.8	5.6	8.0	6.9±0.6	2.6	6.5	5.8±0.6
下喙长 LRL	3.7	5.7	4.6±0.5	2.6	4.5	3.5±0.4	1.6	3.6	2.9±0.4
下侧壁长 LLWL	11.1	15.7	12.7±1.1	7.8	11.8	13.4±1.4	7.2	9.9	8.7±0.6
下翼长 LWL	3.7	8.8	7.4±1.1	4.2	6.6	5.4±0.5	4.0	6.2	5.0±0.5

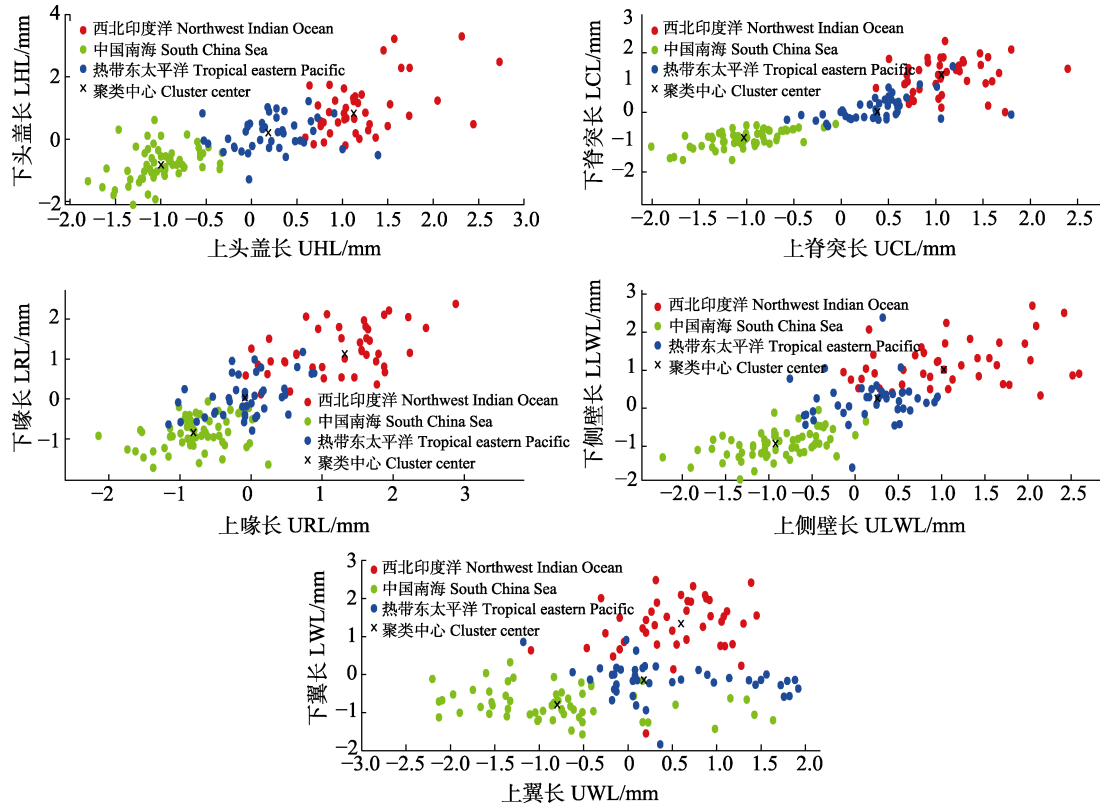


图 2 基于曼哈顿距离的鸢乌贼地理种群聚类效果
 Fig.2 Clustering effect of *S. oualaniensis* based on Manhattan

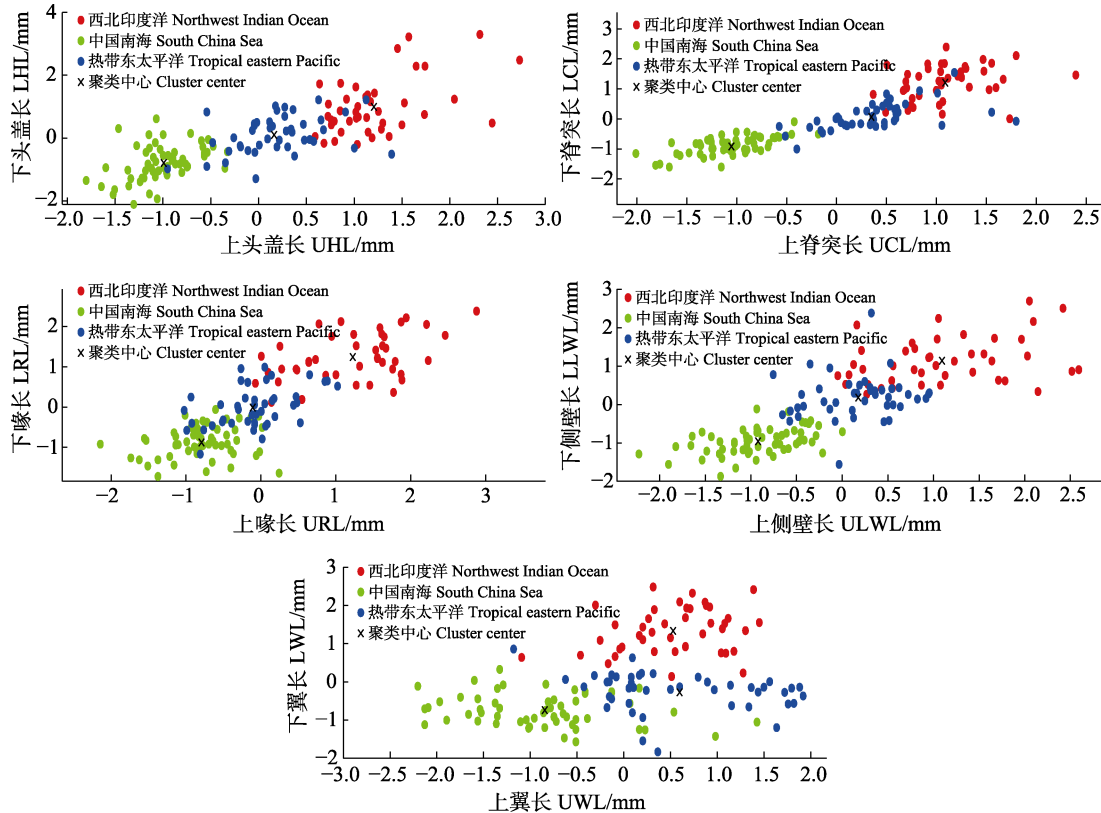


图 3 基于欧氏距离的鸢乌贼地理种群聚类效果
 Fig.3 Clustering effect of *S. oualaniensis* based on Euclidean

热带东太平洋海区、中国南海鸢乌贼上、下头盖长的聚类中心分别为(1.12, 0.83)、(0.18, 0.21)和(-1.00, -0.82), 上、下脊突长聚类中心分别为(1.06, 1.26)、(0.38, 0.02)、(-1.03, -0.85), 上、下喙长的聚类中心

分别为(1.32, 1.13)、(-0.08, 0.03)、(-0.81, -0.84), 上、下侧壁长的为聚类中心分别为(1.03, 1.01)、(0.26, 0.26)、(-0.92, -0.93), 上、下翼长为聚类中心分别为(0.60, 1.34)、(0.18, -0.14)、(-0.80, -0.79)(表3)。

表3 聚类中心的存储位置
Tab.3 Storage locations for cluster centers

距离 Distance	聚类中心 Cluster centers	上头盖长/ 下头盖长 UHL/LHL	上脊突长/ 下脊突长 UCL/LCL	上喙长/ 下喙长 URL/LRL	长侧壁长/ 下侧壁长 ULWL/LLWL	上翼长/ 下翼长 UWL/LWL
曼哈顿距离 Manhattan	西北印度洋 Northwest Indian Ocean	1.12/0.83	1.06/1.26	1.32/1.13	1.03/1.01	0.60/1.34
	热带东太平洋 Tropical eastern Pacific	0.18/0.21	0.38/0.02	-0.08/0.03	0.26/0.26	0.18/-0.14
	中国南海 South China Sea	-1.00/-0.82	-1.03/-0.85	-0.81/-0.84	-0.92/-0.93	-0.80/-0.79
欧氏距离 Euclidean	西北印度洋 Northwest Indian Ocean	1.21/0.99	1.09/1.21	1.23/1.24	1.09/1.14	0.53/1.33
	热带东太平洋 Tropical eastern Pacific	0.16/0.10	0.35/0.07	-0.11/-0.01	0.18/0.18	0.60/-0.27
	中国南海 South China Sea	-0.99/-0.79	-1.05/-0.92	-0.79/-0.88	-0.92/-0.96	-0.85/-0.74

从聚类中心的位置可初步得出,不同的距离算法和远离样本均值中心的点,均会对聚类中心产生一定的偏移。

2.3 判别成功率

选择10组上、下角质颚形态学参数中的任意2组,例如上、下头盖长,上、下脊突长等,尽管各海区的聚类中心位置会相应变化,但聚类识别的结果不变。

曼哈顿距离分析显示,西北印度洋鸢乌贼43尾被成功识别,7尾漏分至热带东太平洋,识别成功率为86.0%;热带东太平洋鸢乌贼37尾被成功识别,13尾漏分至中国南海,识别成功率为74.0%;中国南海鸢乌贼50尾均被成功识别,识别成功率为100%;总体识别成功率为86.7%(表4、表5)。

欧氏距离分析显示,西北印度洋鸢乌贼43尾被成功识别,7尾漏分至热带东太平洋,识别成功率为86.0%;热带东太平洋鸢乌贼40尾被成功识别,10尾漏分至中国南海,识别成功率为80.0%;中国南海鸢乌贼50尾均被成功识别,识别成功率为100%;总体识别成功率为88.7%(表4、表5)。

从三大海区鸢乌贼的整体区分情况来看,当角质颚数据经过z-score标准化,采用欧氏距离进行聚类分析的总正确区分率为88.7%,略优于曼哈顿距离(表5)。本研究初步得出,曼哈顿距离和欧氏距离均适

用于鸢乌贼角质颚的识别,二者对于西北印度洋和中国南海的鸢乌贼角质颚识别结果一致,但无论采用以上哪种距离以及哪种角质颚形态学参数的组合方法,对于中国南海鸢乌贼的判别效果最好,而对于热带东太平洋鸢乌贼的漏分率以及中国南海鸢乌贼的错分率均最高,但其中欧氏距离识别热带东印度洋鸢乌贼结果更为准确,成功区分率为80.0%(表4、表5)。

3 讨论

3.1 角质颚在头足类地理种群判别的优势

头足类广泛分布于大西洋、太平洋、印度洋和南极等海域,是海洋中重要的食物资源之一,研究其种群分类情况可为更高消费者的食性分析提供依据,更好地了解海洋食物网关系,为资源评估以及合理开发奠定基础(范江涛等,2015)。形态学法通过度量头足类外形长度特征、解剖学特征等,实现头足类分类鉴定最常用的传统方法,而头足类的硬组织要比软组织更适用于种类与种群的鉴定识别(胡飞飞等,2017;韩青鹏等,2017)。其中,硬组织中的耳石和角质颚均可作为种类和种群划分的主要依据,但与耳石相比,角质颚更易提取、生长纹更易观察、耐腐蚀且结构稳定,其广泛存在于更高营养级消费者的胃内(刘必林等,2009),应用前景远大于耳石。

表 4 三大海区鸢乌贼地理种群聚类统计结果/ind.
Tab.4 Cluster result statistics of *S. oualaniensis*/ind.

距离 Distance	实测/聚类 Measured/Cluster	西北印度洋 Northwest Indian Ocean	热带东太平洋 Tropical eastern Pacific	中国南海 South China Sea	实测 Measured
曼哈顿距离 Manhattan	西北印度洋 Northwest Indian Ocean	43	7	0	50
	热带东太平洋 Tropical eastern Pacific	0	37	13	50
	中国南海 South China Sea	0	0	50	50
	聚类总和 Sum of clusters	43	44	63	150
欧氏距离 Euclidean	西北印度洋 Northwest Indian Ocean	43	7	0	50
	热带东太平洋 Tropical eastern Pacific	0	40	10	50
	中国南海 South China Sea	0	0	50	50
	聚类总和 Sum of clusters	43	47	60	150

表 5 不同方法下对比三大海区鸢乌贼地理种群区分效果
Tab.5 Distinction effect of different algorithms on *S.oualaniensis*

距离 Distance	海区 Sea area	漏分率 Leakage/ %	错分率 Error rate/%	总正确 区分率 Accuracy/%
曼哈顿距离 Manhattan	西北印度洋 Northwest Indian Ocean	14.0	0	86.7
	热带东太平洋 Tropical eastern Pacific	26.0	15.9	
	中国南海 South China Sea	0	20.6	
欧氏距离 Euclidean	西北印度洋 Northwest Indian Ocean	14.0	0	88.7
	热带东太平洋 Tropical eastern Pacific	14.0	0	
	中国南海 South China Sea	20.0	14.9	

3.2 鸢乌贼角质颚尺寸差异

研究表明, 鸢乌贼的角质颚作为摄食器官, 喙是咬食过程中首先接触的部分(刘必林等, 2009)。鸢乌贼的生长栖息环境、饵料丰富程度、种内变异等因素会造成种群间喙生长变化的不同(方舟等, 2014)。角质颚尺寸大小因受摄食、食性和环境等不同的影响而

有所差异(瞿俊跃等, 2018), 不同部位在摄食过程中作用不同, 导致角质颚不同部位的生长也有所差异(胡贯宇等, 2016)。

本研究中, 鸢乌贼角质颚形态参数数据标准化前后的结果均显示, 西北印度洋海区鸢乌贼角质颚最大, 热带东太平洋次之, 而南海海区的角质颚最小, 进一步说明, 角质颚的相对大小与鸢乌贼的胴长无关, 角质颚形态的地理区域差异是由于抽样偏差导致的可能性不大, 与刘必林等(2019)结论一致。西北印度洋主要受到海表风引起的上升流影响, 使含氧不足但营养丰富的深层水涌至表层, 带来较高的初级生产力, 而热带东太平洋主要受较暖赤道逆流和较冷南赤道流的影响, 初级生产力较低, 所以, 判断印度洋西北部的鸢乌贼比热带东太平洋的鸢乌贼具有更大的角质颚尺寸(Burkill *et al*, 1993; 陆化杰等, 2014)。热带东印度洋的角质颚大小分布范围与本研究其他 2 个海域存在较多重叠, 故其漏判率和误判率相应高。相比之下, 南海海区的鸢乌贼受黑潮支流、季节环流等影响, 渔场的水温偏低, 其角质颚具有较小形态(Hu *et al*, 2000; 冯波等, 2014; Liu *et al*, 2019)。此外, 本研究根据 1 个海域鸢乌贼渔业生产进行采样, 较难实现同一时间内进行捕获样本, 另外, 鸢乌贼样本均为成体样本, 不存在个体发育期不同而影响角质颚形态, 标准化处理的角质颚数据排除了网具选择性对角质颚大小的影响(李思亮等, 2010)。

研究表明, 同一种群的雌雄鸢乌贼在角质颚形态

学参数上存在差异(朱凯等, 2016、2020; 李建华等, 2018), 本研究中, 雌性角质颚形态学参数的均值大于雄性, 三大海区雌雄角质颚的形态学参数差异显著($P<0.05$), 后续将引入鸢乌贼雌雄差异进一步分析。

3.3 K-means 算法优越性

K-means 算法属于非监督分类, 无需先验经验和训练数据集、方法简单、计算快速、处理大数据集时保持可伸缩性和高效性, 是聚类分析最常用的算法之一(林涛等, 2019)。目前, 利用角质颚判别种类与种群的方法有很多, 其中, 刘必林等(2019)在研究角质颚的地理变异中采用逐步判别分析, 得出鸢乌贼的正确识别率为 63.5%。范江涛等(2015)基于南海鸢乌贼性别, 建立角质颚判别函数, 正确判别率达到 65%以上。马迪等(2019)运用傅里叶法基于角质颚上颚的判别总正确率为 83.3%。陈芑等(2015)结合主成分分析判别成功率为 76.9%。而本研究基于模式识别的 K-means 动态聚类法对于识别鸢乌贼种群快速有效, 总正确区分率为 88.7%。

4 展望

传统形态学度量角质颚形态参数方法耗力费时, 数据的读取与记录存在误差, 期待后续基于计算机模式识别自动提取角质颚形态学长度, 将大大提高度量角质颚形态参数的效率。K-means 动态聚类法采用迭代方法, 得到的结果通常陷入局部最优解、且该算法对于离群点比较敏感, 会导致聚类中心产生一定偏移(杨俊闯等, 2019)。此外, 对于鸢乌贼在不同海域的不同时期生长阶段、洄游变化等因素会导致群体形态存在差异, 而这些差异会导致 K-means 聚类结果出现偏差。

针对上述不足, 期望后续尽可能在同一时间段内采样, 考虑雌雄个体以及季节差异等, 进一步的研究有必要对 K-means 算法进行改进, 使用剪枝后的数据集, 并采用局部离群因子检测算法得到离群点集(许艳静等, 2019), 或利用加权密度法选取初始聚类中心点(马克勤等, 2020), 降低 K-means 算法对离群点的敏感性; 基于样本方差结合其他算法, 例如, 遗传算法、人工蜂群算法等具有强鲁棒性和全局搜索能力的算法, 避免 K-means 算法过早陷入全局最优解(王丝丝等, 2018), 后续还需改良构造出普遍适用于识别各海区鸢乌贼的算法, 能具体区分同一海域雌雄鸢乌贼的不同种群, 期待进一步为头足类资源评估提供科学依据。

参 考 文 献

- BORGES T C. Discriminate analysis of geographic variation in hard structures of *Todarodes saguttatus* (Lamarck 1798) from North Atlantic Ocean. ICES Marine Science Symposium, 1990(199): 433-440
- BURKILL P H, MANTOURA R F C, OWENS N J P. Biogeochemical cycling in the northwestern Indian Ocean: A brief overview. Deepsea Research Part II: Topical Studies in Oceanography, 1993, 40(3): 643-649
- CHEN P, FANG Z, CHEN X J. Population identification of *Ommastrephes bartramii* based on beak's external morphological characters. Marine Fisheries, 2015, 37(1): 1-9 [陈芑, 方舟, 陈新军. 基于角质颚外部形态学的柔鱼种群判别. 海洋渔业, 2015, 37(1): 1-9]
- FAN J T, QIU Y S, CHEN Z Z, et al. Morphological difference of the beak between two stocks of *Sthenoteuthis oualaniensis* inhabiting South China Sea. Periodical of Ocean University of China (Natural Science), 2015, 45(10): 42-49 [范江涛, 邱永松, 陈作志, 等. 南海鸢乌贼两个群体间角质颚形态差异分析. 中国海洋大学学报(自然科学版), 2015, 45(10): 42-49]
- FANG Z, CHEN X J, JIN Y. Preliminary study of species identification of cephalopod in East China Sea and Yellow Sea based on beak morphology. Journal of Shanghai Ocean University, 2019, 28(3): 365-372 [方舟, 陈新军, 金岳. 基于角质颚的东、黄海头足类科类判别. 上海海洋大学学报, 2019, 28(3): 365-372]
- FANG Z, CHEN X J, LU H J, et al. Morphology and growth of beaks in two cohorts for neon flying squid (*Ommastrephes bartramii*) in the North Pacific Ocean. Acta Ecologica Sinica, 2014, 34(19): 5405-5415 [方舟, 陈新军, 陆化杰, 等. 北太平洋两个柔鱼群体角质颚形态及生长特征研究. 生态学报, 2014, 34(19): 5405-5415]
- FENG B, YAN Y R, ZHANG Y M, et al. A new method to assess the population of *Sthenoteuthis oualaniensis* in South China Sea. Progress in Fishery Science, 2014, 35(4): 1-6 [冯波, 颜云榕, 张宇美, 等. 南海鸢乌贼资源评估的新方法. 渔业科学进展, 2014, 35(4): 1-6]
- HAN Q P, LU H J, CHEN X J, et al. Morphological analysis of beaks of *Uroteuthis chinensis* in the northern South China Sea. South China. Fisheries Science, 2017, 13(4): 122-130 [韩青鹏, 陆化杰, 陈新军, 等. 南海北部海域中国枪乌贼角质颚的形态学分析. 南方水产科学, 2017, 13(4): 122-130]
- HU F F, CHEN X J, LIU B L, et al. Review on identification of cephalopods classification. Marine Fisheries, 2017, 39(1): 110-120 [胡飞飞, 陈新军, 刘必林, 等. 头足类分类鉴定的研究进展. 海洋渔业, 2017, 39(1): 110-120]
- HU G Y, CHEN X J, FANG Z. Effect of individual growth on beak morphometry of jumbo flying squid *Dosidicus gigas* off the Peruvian exclusive economic zone. Journal of

- Fisheries of China, 2016, 40(1): 36–44 [胡贯宇, 陈新军, 方舟. 个体生长对秘鲁外海茎柔鱼角质颚形态变化的影响. 水产学报, 2016, 40(1): 36–44]
- HU J, KAWAMURAM H, HONG H, *et al.* A review on the currents in the South China Sea: Seasonal circulation, South China Sea warm current and Kuroshio intrusion. *Journal of Oceanography*, 2000, 56: 607–624
- JACKSON G D, BUXTON N G, GEORGE M J A. Beak length analysis of *Moroteuthis ingens* (Cephalopoda: Onychoteuthidae) from the Falkland Islands region of the Patagonian shelf. *Journal of the Marine Biological Association of the United Kingdom*, 1997, 77(4): 1235–1238
- LI J H, CHEN X J, CHEN P, *et al.* Morphological and growth characteristics of *Dosidicus gigas* beak in the equatorial waters. *Marine Fisheries*, 2018, 40(4): 385–393 [李建华, 陈新军, 陈芑, 等. 赤道海域茎柔鱼角质颚形态及生长特征分析. 海洋渔业, 2018, 40(4): 385–393]
- LI S L, CHEN X J, LIU B L, *et al.* Morphological analysis on population structure of Neon flying squid (*Ommastrephes bartramii*) in the Northwest Pacific Ocean. *Periodical of Ocean University of China (Natural Science)*, 2010, 40(3): 43–48 [李思亮, 陈新军, 刘必林, 等. 基于形态法的西北太平洋柔鱼种群结构研究. 中国海洋大学学报(自然科学版), 2010, 40(3): 43–48]
- LIN T, ZHAO C. Nearest neighbor optimization K-means clustering algorithm. *Computer Science*, 2019, 46(11A): 216–219 [林涛, 赵臻. 最近邻优化的 K-means 聚类算法. 计算机科学, 2019, 46(11A): 216–219]
- LIU B L, CHEN X J, FANG Z, *et al.* Species identification of cephalopods based on beak morphometric variables. *Oceanologia et Limnologia Sinica*, 2015, 46(6): 1365–1372 [刘必林, 陈新军, 方舟, 等. 基于角质颚长度的头足类物种判别. 海洋与湖沼, 2015, 46(6): 1365–1372]
- LIU B L, CHEN X J, WANG X H, *et al.* Geographic, intraspecific and sexual variation in beak morphology of purple back flying squid (*Sthenoteuthis oualaniensis*) throughout its distribution range. *Marine and Freshwater Research*, 2019, 70(3): 417–425
- LIU B L, CHEN X J. Beak length analysis of the purpleback flying squid *Sthenoteuthis oualaniensis* in northwest Indian Ocean. *Progress in Fishery Sciences*, 2010, 31(1): 8–14 [刘必林, 陈新军. 印度洋西北海域鸢乌贼角质颚长度分析. 渔业科学进展, 2010, 31(1): 8–14]
- LIU B L, CHEN X J. Review on the research development of beaks in Cephalopoda. *Journal of Fisheries of China*, 2009, 33(1): 157–164 [刘必林, 陈新军. 头足类角质颚的研究进展. 水产学报, 2009, 33(1): 157–164]
- LIU H J, WANG C J, CHEN X J. Preliminary study on the biological characteristics of *Sthenoteuthis oualaniensis* in the high seas nearby the equator of eastern Pacific during April to June. *Journal of Shanghai Ocean University*, 2014, 23(3): 441–447 [陆化杰, 王从军, 陈新军. 4–6月东太平洋赤道公海鸢乌贼生物学特性初步研究. 上海海洋大学学报, 2014, 23(3): 441–447]
- MA D. Species identification of common inshore cephalopods in the East China Sea and the Yellow Sea based on beak. Master's Thesis of Shanghai Ocean University, 2017 [马迪. 基于角质颚的东黄海常见头足类鉴定与分类. 上海海洋大学硕士研究生学位论文, 2017]
- MA K Q, YANG Y J, QIN H W, *et al.* K-means clustering algorithm combining max-min distance and weighted density. *Computer Engineering and Applications*, 2020, 56(16): 50–54 [马克勤, 杨延娇, 秦红武, 等. 结合最大最小距离和加权密度的 K-means 聚类算法. 计算机工程与应用, 2020, 56(16): 50–54]
- QU J Y, LI J H, CHEN P, *et al.* The morphology and growth characteristics of beaks of European common cuttlefish (*Sepia officinalis*) in the coast of West Africa. *Progress in Fishery Sciences*, 2018, 39(2): 164–170 [瞿俊跃, 李建华, 陈芑, 等. 西非沿岸乌贼角质颚形态及生长特征. 渔业科学进展, 2018, 39(2): 164–170]
- VEGA M A, ROCHA F J, GUERRA A, *et al.* Morphological differences between the Patagonian squid *Loligo gahi* populations from the Pacific and Atlantic Oceans. *Bulletin of Marine Science*, 2002, 71(2): 903–913
- WANG S S, ZHANG J L, CHEN C, *et al.* Optimization of K-means based on variance statistics and improved swarm intelligent algorithm. *Journal of Systems Science and Mathematical Sciences*, 2018, 38(10): 1117–1127 [王丝丝, 张敬磊, 陈慈, 等. 基于方差与改进群智能算法的 K-means 聚类优化. 系统科学与数学, 2018, 38(10): 1117–1127]
- WANG T X, GUO J J, WANG Q. The recognition of soil sample based on the K-means dynamic clustering analysis. *Building Science*, 2010, 26(7): 52–56, 71 [王同兴, 郭骏杰, 王强. 基于 K 均值动态聚类分析的土壤识别. 建筑科学, 2010, 26(7): 52–56, 71]
- XU Y J, ZHU J M, DING Q Y, *et al.* Outlier fraud user detection algorithm for crowdsourcing competition. *Statistics and Information Forum*, 2019, 34(10): 20–26 [许艳静, 朱建明, 丁庆洋, 等. 众包竞赛的离群点欺诈用户检测算法研究. 统计与信息论坛, 2019, 34(10): 20–26]
- YANG J C, ZHAO C. Survey on K-means clustering algorithm. *Computer Engineering and Applications*, 2019, 55(23): 7–14, 63 [杨俊闯, 赵超. K-Means 聚类算法研究综述. 计算机工程与应用, 2019, 55(23): 7–14, 63]
- ZENG J. Analysis of partition-based data mining K-means clustering algorithm. *Modern Electronics Technique*, 2020, 43(3): 14–17 [曾俊. 基于划分的数据挖掘 K-means 聚类算法分析. 现代电子技术, 2020, 43(3): 14–17]
- ZHU K, WANG X H, ZHANG P, *et al.* A study on morphological variations and discrimination of medium and dwarf forms of purple flying squid *Sthenoteuthis oualaniensis* in the southern South China Sea. *Journal of Tropical Oceanography*, 2016, 35(6): 82–88 [朱凯, 王雪辉, 张鹏,

等. 南海南部鸢乌贼中型群与微型群形态学差异及其判别分析. 热带海洋学报, 2016, 35(6): 82–88]
 ZHU K, ZHANG L C, XIAO C Y, *et al.* Characterizing fecundity of dwarf form of female purple flying squid (*Sthenoteuthis*

oualaniensis) in the South China Sea. Progress in Fishery Sciences, 2020, 41(6): 140–148 [朱凯, 张立川, 肖楚源, 等. 南海鸢乌贼微型群雌性个体繁殖力研究. 渔业科学进展, 2020, 41(6): 140–148]

(编辑 马瑾艳)

Pattern of Recognition Beaks in *Sthenoteuthis oualaniensis* Based on *K*-Means Dynamic Clustering

ZHENG Xinyu¹, LIU Bilin^{1,2①}, KONG Xianghong¹, WANG Xuehui³

(1. College of Marine Sciences, Shanghai Ocean University, Shanghai 201306, China;

2. Key Laboratory of Sustainable Exploitation of Oceanic Fisheries Resources, Ministry of Education, Shanghai Ocean University, National Distant-Water Fisheries Engineering Research Center, Key Laboratory of Oceanic Fisheries Exploration, Ministry of Agriculture and Rural Affairs, Scientific Observing and Experimental Station of Oceanic Fishery Resources, Ministry of Agriculture and Rural Affairs, Shanghai 201306, China;

3. South China Sea Fisheries Research Institute, China Academy of Fishery Sciences, Guangzhou, Guangdong 510300, China)

Abstract Cluster analysis has been widely used for pattern recognition, machine learning, and in other fields. The *K*-means dynamic clustering algorithm is simple and efficient, which is why it is one of the most commonly used methods of cluster analysis. The beak of cephalopods, comprising hard tissue, has been widely used to determine species and identify populations owing to its stable structure, corrosion resistance, easily observed growth lines, and abundant characteristic information, causing it to have great application prospects. In this study, the *K*-means dynamic clustering algorithm was used on 150 pairs of *Sthenoteuthis oualaniensis* beaks within the mantle length range of 120~200 mm. Samples were collected from the northwest Indian Ocean, the tropical eastern Pacific Ocean and the South China Sea from 2014 to 2019. The results showed that *S. oualaniensis* from the northwest Indian Ocean had the largest beaks, followed by the tropical eastern Pacific Ocean, and those in the South China Sea. The *K*-means dynamic clustering algorithm showed that *S. oualaniensis* from the three areas can be well distinguished. We used z-scores to normalize the data the created a 2D beak morphological parameter matrix to randomize the data before we conducted a *K*-means dynamic clustering analysis with Manhattan distance and Euclidean distance. The total correct discrimination rate was 86.7% and 88.7%, respectively. This study also identified that the geographic regional differences in beak morphology are unlikely to be due to sampling bias. From the location of the clustering center, we concluded that the Manhattan and Euclidean distance algorithms and outlying points will generate deviations from the clustering center. The *K*-means dynamic clustering algorithm for beaks of the *S. oualaniensis* has great reference value. We identified improvements that optimize the *K*-means algorithm to expand capability for universal use. These improvements and a retrieval system will improve our capabilities to identify *S. oualaniensis* species.

Key words *Sthenoteuthis oualaniensis*; Beak; Pattern of recognition; Manhattan distance; Euclidean distance

① Corresponding author: LIU Bilin, E-mail: bl-liu@shou.edu.cn