目录

[渔船航道识别算法 1](#_Toc507945954)

[研究的起点 1](#_Toc507945955)

[迭代计算的输入 2](#_Toc507945956)

[投票方案的设计 3](#_Toc507945957)

[一次完整的迭代流程 5](#_Toc507945958)

[渔船航道 7](#_Toc507945959)

[本章小结 9](#_Toc507945960)

# 渔船航道识别算法

在上一章，我们通过渔船捕捞行为识别算法比较了不同年份同一季度的捕捞区热度图，发现渔业资源有东移趋势。但是，这种趋势是如何变化的？区域之间又有怎样的联系？这些问题难以分析的原因在于，捕捞区域统计热度图是一种直观的、整体的分析方法，而渔业资源的真实变化是动态的、独立的、不均匀的。虽然已经可以统计捕捞区热度图，但局部可能随时间扩大或缩小，某些热点区域可能会东移，也可能保持不动。要想刻画这些变化，需要从更细粒度的角度分析捕捞区，对统计出的捕捞区热度图进行合理的分类。

## 研究的起点

VMS系统的部署推动了渔业资源的分析研究。Mill等通过识别捕捞行为，研究了渔业资源的分布情况[[1]](#endnote-1)。Fonseca等分析了捕捞区的动态变化[[2]](#endnote-2)。但是，这些研究往往只关注捕捞行为，而忽视了渔船在捕捞区之间的航行轨迹信息。我们认为这些轨迹信息描述了捕捞区域之间的联系，刻画了渔船的航道。而且，陆地上的道路能够把城市划分成区块，同理，海上的航道同样有助于捕捞区域的划分。所以，我们认为这部分信息应该予以重视，并提出问题：如何从VMS数据中识别渔船航道信息。

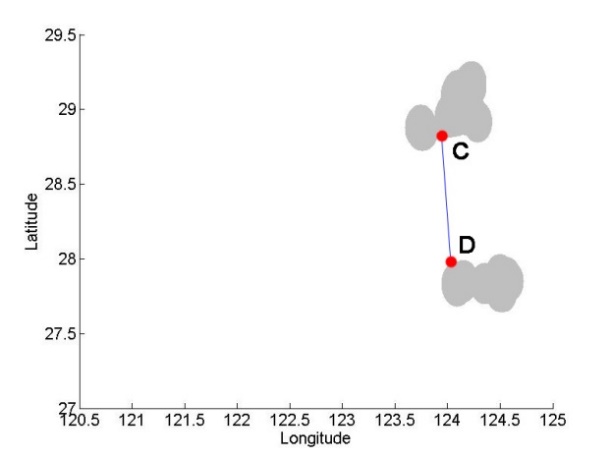
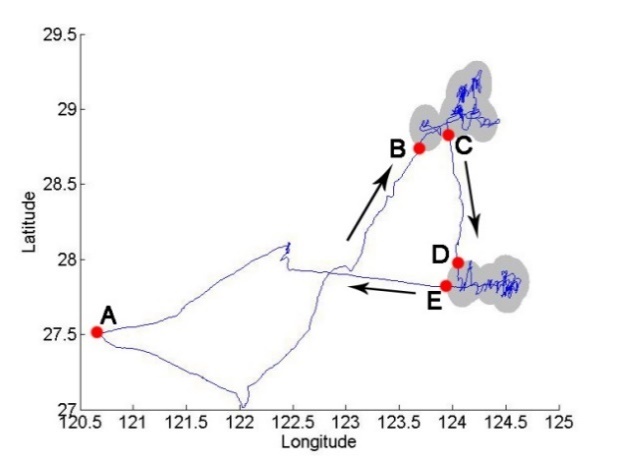
研究的难点在于，海上本身是没有航道和区块的。即使是目前相关研究得到的捕捞区，也只是通过粗略统计得到的，缺乏能够准确验证的手段。而捕捞区域识别的准确性又会影响到捕捞区域之间的轨迹，进而影响到渔船航道的识别。两者相互依存又相互矛盾。

于是，我们提出了渔船航道识别算法来同时完成对于海上航道和区块的识别任务。这个算法包含两个部分：捕捞区粗粒度识别和迭代计算。前者利用上一章提到的渔船捕捞轨迹识别算法得到捕捞区。在此基础上，统计渔船在捕捞区之间的航行轨迹。后者将捕捞区和航行轨迹作为输入，利用一种投票方法迭代地尝试各种分类方案，并最终确定捕捞区的细粒度识别以及渔船航道。下面来介绍这种迭代方法的细节。

## 迭代计算的输入

上一节中，我们提到了迭代输入由两部分组成：捕捞区热度图和渔船在捕捞区之间的轨迹信息。本节中分别对两者进行介绍。

捕捞区热度图产生于渔船捕捞行为识别算法。通过数学形态学方法，从每个航次中提取对应的捕捞区，并按照季度进行统计，得到如图3-10的捕捞区热度图。



1. （b）

图3-1 对捕捞区之间的航行轨迹进行抽象

渔船在捕捞区之间的轨迹信息并不是直接使用真实的轨迹数据，而是将其抽象为一条线段。如图3-1（a）展示了某艘船一个航次的轨迹，灰色区域为识别出的捕捞区域，各个节点的访问顺序为ABCDEA，可以看出，CD间的轨迹就是渔船在两个捕捞区之间的航行轨迹。但是这段数据的数据量和变化多样，不利于接下来的处理。所以将其抽象为图3-1（b）所示的“航行线段”——只保留了离开捕捞区的C点和进入下一个捕捞区的D点。这样处理虽然忽略了大量的信息，但是能够得到统一的航行信息：线段两端应该落在不同的捕捞区内部。

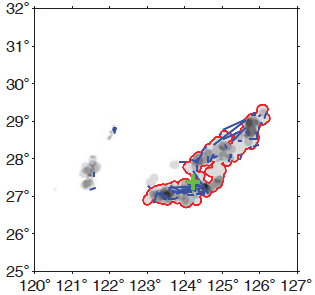
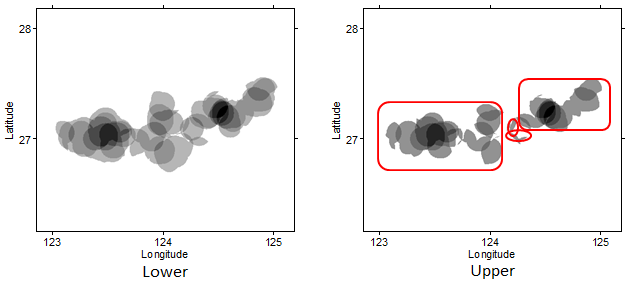


图3-2 2014年第二季度捕捞区统计热度图与航行线段

将迭代计算的两个输入叠加在一起进行分析，如图3-2所示。灰色区域为统计2014年第二季度得到的捕捞区热度图，红色曲线包络的是一个相互连通的捕捞区域，蓝色线段是对应时间段内统计的所有航行线段。首先，根据连通性，我们把捕捞区热度图粗略地分为七个区域，可以看出，航行线段不仅仅分布在不同区域之间，还分布于它的内部。以红色曲线勾勒的区域为例，作为一个面积较大的捕捞区域，内部包含了大量的航行线段。回顾我们对航行线段的定义可以提出问题：是否应该把一个连通的捕捞区划分为多个子区域？我们给出的答案是应该进行分类。对于较短的航行线段也许可以认为是渔船捕捞行为识别的误差导致的，但是长线段不能忽视。接下来的投票方案将介绍如何实现分类方法。

## 投票方案的设计

在上一节的最后，我们确定了要对捕捞区进行分类。但是如何分类呢？首先排出了形如K-means的分类方式——在一个捕捞区域内通过边界线的形式将其分割为多个小区域。因为这样做缺乏实际的指导作用。对于渔民而言，如果两个捕捞区之间的区分只是一条线，那么分类前与分类后就没有区别。所以，分类的目标是要将捕捞区划分为多个互不连通的小区域。

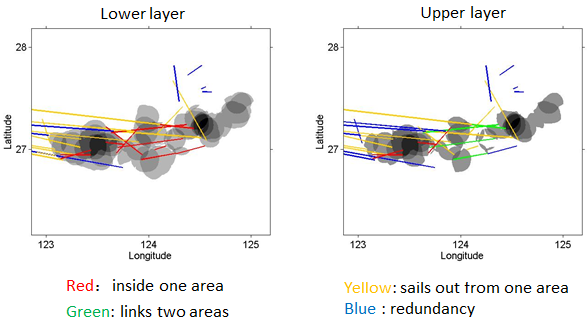


1. (b)

图3-3 相邻图层某个出现合并的区域

为了使分类后的捕捞区彼此分离，我们决定从捕捞区热度图的层次入手，保留热度数值大的区域，忽略热度值小的区域。设阈值θ=0，当热度值大于θ，可以保留图上全部的捕捞区信息；随着θ增大，保留的信息减少，且逐渐以高热度值为核心相互独立。通过分析对比，当θ取值为热度最大值的25%时，分类效果较好。如图3-3所示，是2014年第二季度的捕捞区的局部。将阈值θ=热度最大值\*25%作为热度图的切入点，观察到的相邻两个图层（a）下一层（b）上一层。在相邻图层间捕捞区存在三种变化方式：（Ⅰ）新增，上一层的空白位置对应到下一层是一个独立的捕捞区域；（Ⅱ）扩张，上一层中的某个捕捞区对应到下一层范围变大，但仍保持独立性；（Ⅲ）合并，上一层的两个或多个捕捞区域在下一层构成了一个连通的捕捞区。和前文一样，按照捕捞区的存在与否，把（Ⅰ）（Ⅱ）两种情况视为独立的捕捞区，对（Ⅲ）中捕捞区合并的情况单独分析。图3-3（a）中的捕捞区就符合方式（Ⅲ），是由图3-3（b）中的四个捕捞区合并组成的一个连通的整体。

利用这些“上一层”的小区域，我们来讨论如何进行分类。为此设计了一个投票方法：对于两种不同的分类方案，我们进行投票，得票数多的就是较优的方案。下面介绍具体的实现方法。



1. （b）

图3-4 四种选票。“红色”反对票，“绿色”支持票，

“黄色”部分支持票，“蓝色”弃权票

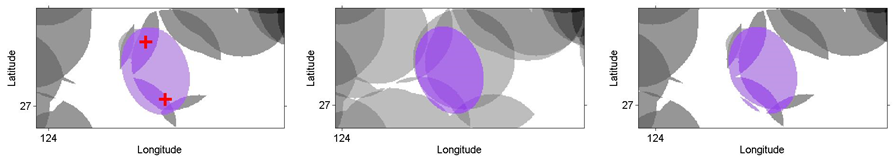
我们把航行线段作为“选票”。根据线段两端落点的不同，共有四种不同的选票，如图3-4所示：两端在同一个捕捞区内的选票用红色表示，称为反对票，得票数记作-2；两端在不同捕捞区内的用绿色表示，称为支持票，得票数记作+2；只有一端在捕捞区内的用黄色表示，是对其中一个捕捞区分类的支持，称为部分支持票，记作+1；两端均不在捕捞区内的用蓝色表示，称为弃权票，记作0。所以，对于一个分类方案进行投票，就是根据落点判断每一个选票的类型，将对应数字相加求和。图4-4是对图4-3的案例进行投票，图3-4（a）得票数为-11，图3-4（b）得票数为11。所以根据航行线段，图3-4（b）的分类方案要比图3-4（a）更佳。

## 一次完整的迭代流程

上一节介绍了如何根据投票结果获得更优的分类方案。在这一节中，将介绍如何利用迭代的方式筛选出最终的分类方案。

通过投票我们可以得出结论：相比于图3-4（a）中组合成一个整体，图3-4（b）中各个小区域保持独立会更合理一些。那么会不会存在更优的分类方案呢？我们尝试对多种组合进行投票。

如果对全部组合进行投票，组合数将随着区域数量上升呈指数型增长。实际上当三个区域排成一条直线时，基本上可以排除两端合并为一个整体而中间区域保持独立这种情况。根据这个剪枝条件，我们设计了一个由近及远的、不可逆的组合方案。具体步骤：取各个捕捞区的几何中心点两两连线，按线段长度排序；根据排序顺次进行合并、投票；若相比于合并前，合并后得票数较高，则将两个区域视为一个整体，不再分开讨论其他分组情况；当尝试合并的两个几何中心点对应的捕捞区归属于同一个整体，则不再进行尝试。对于n个捕捞区域，这种处理方案最少只需要进行（n-1）次比较，即n个捕捞区域合并为一个整体；最坏情况只进行(n-1)n/2次比较，对应着n个捕捞区域全部独立的情况。在算法实现过程中，利用并查集算法简化判断归属的时间开销，且并没有出现时间开销过大的情况。这是因为当n比较大时（最多为14个），其中包含了大量的形如图3-3中间部分的小捕捞区，这些区域往往合并前后得票数相同。我们倾向于将捕捞区尽量合并为一个整体，于是，对于得票数相同的情况，判断为合并而非独立。



1. （2） （3）

图3-5 两捕捞区的合并策略

合并的具体方法如图3-5所示。在对两捕捞区进行合并时，以两个捕捞区的几何中心点作为椭圆焦点并连线，延长线较短一侧作为长轴，作椭圆连接这两个捕捞区如图3-5（1）。将椭圆与下一层捕捞区取交集如图3-5（2），计算结果即为两捕捞区域的合并方式如图3-5（3）。这种合并方法虽然存在与其他捕捞区域产生交集的可能，但胜在计算简便，能够满足合并的需求。

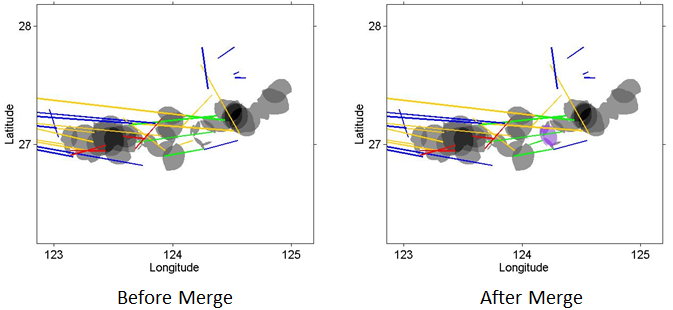


图3-6 比较合并前后的得票数

对于每次合并的尝试，比较合并前后的得票数大小。如图3-6所示，是对图3-5合并前后进行了投票，图3-6（1）的得票数为11，图3-6（2）的得票数为12，所以确定合并两个捕捞区域。接下来，继续按照线段长度从小到大的顺序进行迭代，依次比较捕捞区合并前后的得票情况，重复上述步骤。确定最终的分类方案。

## 渔船航道

通过上面介绍的算法，实现了对捕捞区的细粒度划分。在此基础上，通过统计各个捕捞区的出入度和转移关系，我们得到了对渔船在捕捞区之间航行轨迹的抽象结果——渔船航道。

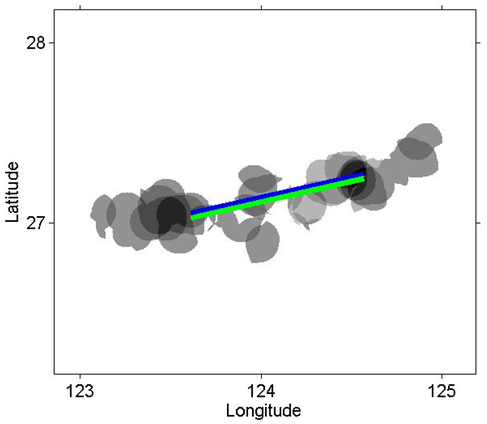
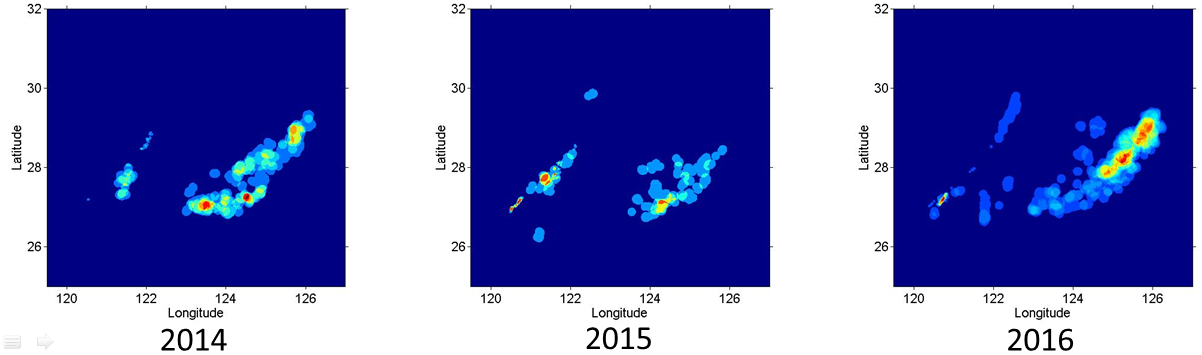


图3-7 渔船航道

如图3-7所示，灰色部分是图3-6中样例通过迭代，最终确定的分类方案，从最初的四个区域变为两个区域。统计渔船在先后在两个区域进行捕捞作业的轨迹，即选票中的“支持票”，作图上蓝绿色线段，线段端点定位在对应区域的几何中心点上。蓝色表示自西向东，绿色表示自东向西，线段粗细表示船次，也可以称为流量。图3-7上两个区域之间自西向东航行有三个船次，自东向西有三个船次。

我们称这样的连线为渔船航道。通过构建渔船航道，就可以构筑与陆地上的路网相似的道路网络，并以此为基础分析渔业资源的分布变化情况。



1. （2） （3）

图3-8 连续三年第二季度捕捞区热度图

比如图3-8是2014年至2016年第二季度的捕捞热度图，通过观察可以发现捕捞区有向东北移动的趋势，还可以通过捕捞区的几何中心位置、轮廓等特征描述这种变化。但是这些都是笼统的分析，无法进行局部捕捞区域的刻画。

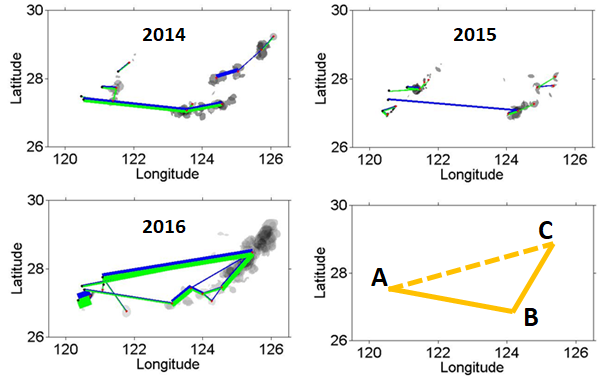


图3-9 连续三年海上路径变化

通过分类算法对捕捞区进行划分并构筑路网，有助于对渔业资源及渔民行为进行分析。如图3-9所示，对于连续三年第二季度的海上路径，我们把它抽象为一个三角形进行分析：A点表示港口，BC两点表示两个捕捞区域。在2014年第二季度，大部分的渔船出海沿着AB边航行，2015年对应季度船次整体明显减少，而到了2016年，只有少量船只沿AB航线捕鱼，大部分的船次都选择直接沿AC边捕捞。相比于B点，选择更远的C点进行捕捞，这充分说明B点渔业资源减少。同时，2014年便开始在C点捕捞的渔民冒险和开拓意识较强，而2016年仍在B点捕捞的渔民相对而言比较保守。

## 本章小结

为了将捕捞区热度图进行细粒度划分，同时找出渔船航道，本章提出了渔船航道识别算法。它包含两部分，一个是第二章介绍的渔船捕捞行为识别算法，得到两组参数：捕捞区之间的航行轨迹和捕捞区热度图。另一个是通过迭代的方式对两两分类方案进行投票，保留得票数高的方案，得到最终的分类结果。并根据出入度统计出渔船航道。根据渔船航道的流量变化，我们发现了渔民捕捞策略的变化。

1. Mills C M, Townsend S E, Jennings S, et al. Estimating high resolution trawl fishing effort from satellite-based vessel monitoring system data[J]. Ices Journal of Marine Science, 2006, 64(2):248-255. [↑](#endnote-ref-1)
2. Fonseca T, Campos A, Afonso-Dias M, et al. Trawling for cephalopods off the Portuguese coast—Fleet dynamics and landings composition[J]. Fisheries Research, 2008, 92(2):180-188. [↑](#endnote-ref-2)