



测绘工程专业学位论文

基于全球六边形网格的船舶轨迹预测
Ship Trajectory Prediction Based on Global
Hexagonal Grid

作 者：吴名荣
导 师：余接情 教授
校外导师：张少铖

中国矿业大学
二〇二四年六月

中图分类号 P237

学校代码 10290

UDC 528

密 级 公开

中国矿业大学
测绘工程专业学位论文

基于全球六边形网格的船舶轨迹预测
Ship Trajectory Prediction Based on Global
Hexagonal Grid

作 者 吴名荣 导 师 余接情

申请学位 资源与环境硕士专业学位 培养单位 环境与测绘学院

学科专业 测绘工程 研究方向 地理信息工程及应用

答辩委员会主席 张书毕 评 阅 人

二〇二四年六月

摘要

随着社会经济的快速发展，航运业迎来了巨大的发展契机，但同时也不可避免带来了堵塞、碰撞、偏航等公共问题。通过对船舶航行轨迹的预测能够及时发现船舶的轨迹变化，可以减少海上交通拥挤以及事故的发生概率，同时推动水上交通管理智能化和高效率运作。现有航迹预测方法大多关注于短距离的预测，在面向长距离预测时精度较差，且在历史轨迹不足时预测误差较大。本文紧紧围绕“船舶轨迹预测”这一主题，开展轨迹网格化、热点区域挖掘与航迹预测等内容研究。主要研究工作与成果如下：

(1) 设计了一种航迹快速网格化算法，实现轨迹的全球六边形网格表达，为后续的热点挖掘与航迹预测奠定了基础。首先，通过轨迹预处理技术解决原始 AIS 数据的噪声大、易缺失和数据不一致等问题。在数据预处理的基础上，设计实验充分验证了网格化算法的有效性。实验表明：本文算法能够准确地实现对任意航迹的快速网格化。

(2) 面向潜在航行目的地的计算需求设计了一种基于全球六边形网格的热点区域挖掘方法。该方法综合考虑了轨迹数据的时空特性，基于密度与时间两个维度来提取更为精细化的热点区域。以波罗的海 AIS 数据进行热点提取实验，结果表明：挖掘方法具有一定的准确性，提取结果基本为港口、海峡、交通枢纽等船舶活动频繁区域。

(3) 针对已有船舶轨迹预测方法存在的长远距离预测误差大、历史轨迹不足精度低等问题，提出了一种顾及目的地及航道权等多重约束的船舶轨迹预测方法，探讨了在历史轨迹充足与欠缺时的预测策略，开展可行性分析实验与预测精度分析实验。结果表明：本文预测方法具有一定的可行性和准确性。在历史轨迹充足的情况下，与 LSTM 预测方法相比，本文方法短距离预测误差较大，但在长距离预测精度更优；在历史轨迹不足时，本文方法相对于 LSTM 在长短距离的预测上都表现出较高的预测精度。

(4) 设计与开发了航迹预测原型系统，具备数据库管理、轨迹预处理、轨迹分析与轨迹预测等功能，实现了船舶轨迹的规律分析与实时预测。

关键词：船舶轨迹预测；热点区域挖掘；AIS 数据；全球六边形网格；轨迹网格化

Abstract

With the rapid development of social economy, the shipping industry has ushered in huge development opportunities, but at the same time it has inevitably brought about public problems such as blockage, collision, and yaw. By predicting the ship's navigation trajectory, changes in the ship's trajectory can be discovered in a timely manner, which can reduce maritime traffic congestion and the probability of accidents, while promoting intelligent and efficient operation of water traffic management. Most of the existing trajectory prediction methods focus on short-distance prediction, and have poor accuracy when facing long-distance prediction, and the prediction error is large when there are insufficient historical trajectories. This article closely focuses on the theme of "ship trajectory prediction" and conducts research on trajectory gridding, hotspot area mining and track prediction. The main research work and results are as follows:

(1) A fast gridding algorithm for tracks is designed to realize the global hexagonal grid expression of tracks, which lays the foundation for subsequent hotspot mining and track prediction. First, trajectory preprocessing technology is used to solve the problems of high noise, easy missing and data inconsistency in the original AIS data. Based on data preprocessing, designed experiments fully verified the effectiveness of the gridding algorithm. Experiments show that the algorithm proposed in this thesis can accurately achieve rapid gridding of any trajectory.

(2) A hotspot area mining method based on the global hexagonal grid is designed to meet the computing needs of potential navigation destinations. This method comprehensively considers the spatiotemporal characteristics of trajectory data and extracts more refined hotspot areas based on the two dimensions of density and time. Hotspot extraction experiment were conducted using Baltic Sea AIS data. The result showed that the mining method has a certain degree of accuracy, and the extraction results are basically areas with frequent ship activities such as ports, straits, and transportation hubs.

(3) Aiming at the problems such as long -distance prediction errors and low historical trajectory accuracy of existing long -distance predictions of ship trajectory prediction methods, a ship trajectory prediction method that takes into account multiple constraints such as destinations and channel rights is proposed to explore the relationship between sufficient historical trajectories and Forecasting strategies when there is a shortage, conduct feasibility analysis experiments and prediction accuracy

analysis experiments. The prediction method in this article has certain feasibility and accuracy. When there are sufficient historical trajectories, compared with the LSTM prediction method, the short-distance prediction error of this method is larger, but the long-distance prediction accuracy is better; when the historical trajectories are insufficient, , compared with LSTM, the method in this thesis shows higher prediction accuracy in both long and short distance predictions.

(4) Designed and developed a trajectory prediction prototype system, which has functions such as database management, trajectory preprocessing, trajectory analysis and trajectory prediction, and realizes regular analysis and real-time prediction of ship trajectories.

Keywords: ship trajectory prediction; hot spot area mining; AIS data; global hexagonal grid; trajectory gridding

目 录

1 绪论.....	1
1.1 研究背景与意义.....	1
1.2 国内外研究现状.....	2
1.3 研究内容与技术路线图.....	7
1.4 论文章节安排.....	9
2 全球六边形网格与航迹预测技术	10
2.1 全球六边形网格.....	10
2.2 基于 LSTM 的航迹预测.....	14
2.3 本章小结.....	17
3 面向船舶轨迹预测的热点区域挖掘	18
3.1 总体思路.....	18
3.2 实验数据	19
3.3 轨迹预处理.....	21
3.4 航迹网格化.....	25
3.5 热点区域提取.....	33
3.6 本章小结.....	39
4 基于全球六边形网格的船舶航迹预测	40
4.1 总体思路.....	40
4.2 目标区域预测.....	41
4.3 路径预测.....	45
4.4 可行性分析实验.....	50
4.5 预测精度分析实验.....	56
4.6 本章小结.....	62
5 航迹预测原型系统开发与展示	63
5.1 系统开发环境.....	63
5.2 系统功能.....	63
5.3 数据库表设计.....	65
5.4 系统展示.....	67
5.5 本章小结.....	69
6 总结与展望.....	71
6.1 工作总结.....	71
6.2 未来展望.....	72
参考文献.....	73

图清单

图序号	图名称	页码
图 1-1	技术路线图	8
Figure 1-1	Technology roadmap	8
图 2-1	五种柏拉图式正多面体示意图	10
Figure 2-1	Schematic diagram of five Platonic regular polyhedrons	10
图 2-2	多面体定位	11
Figure 2-2	Polyhedral positioning	11
图 2-3	II 类孔径 4 六边形划分示意图	12
Figure 2-3	Schematic diagram of the hexagonal division of Class II aperture 4	12
图 2-4	球面与多面体投影映射图	12
Figure 2-4	Spherical and polyhedral projection mapping diagram	12
图 2-5	二十面体展开示意图	13
Figure 2-5	Schematic diagram of icosahedron expansion	13
图 2-6	三轴六边形坐标系统	13
Figure 2-6	Three-axis hexagonal coordinate system	13
图 2-7	菱形面坐标轴示意图	14
Figure 2-7	Schematic diagram of the coordinate axis of the rhombus surface	14
图 2-8	LSTM 网络结构示意图	15
Figure 2-8	Schematic diagram of LSTM network structure	15
图 2-9	LSTM 航迹预测模型结构示意图	16
Figure 2-9	Schematic diagram of the structure of the LSTM track prediction model	16
图 2-10	时间序列预测迭代过程示意图	17
Figure 2-10	Schematic diagram of the iterative process of time series forecasting	17
图 3-1	热点区域挖掘总体流程图	19
Figure 3-1	Overall flow chart of hotspot area mining	19
图 3-2	实验数据分布	20
Figure 3-2	Experimental data distribution	20
图 3-3	轨迹中的噪声点	21
Figure 3-3	Noise points in the trajectory	21
图 3-4	DP 算法轨迹压缩过程示意图	23
Figure 3-4	Schematic diagram of the trajectory compression process of DP algorithm	23
图 3-5	DP 算法轨迹压缩结果	24
Figure 3-5	DP algorithm trajectory compression result	24
图 3-6	轨迹数据缺失示意图	24
Figure 3-6	Schematic diagram of missing trajectory data	24
图 3-7	原始航迹由若干大圆弧段组成	26
Figure 3-7	The original track consists of several large arc segments	26
图 3-8	大圆弧网格化示意图	27
Figure 3-8	Schematic diagram of the large arc meshing algorithm	27

图 3-9	轨迹线 AB 采用大圆弧连接示意图	27
Figure 3-9	Schematic diagram of trajectory line AB connected by a large arc	27
图 3-10	计算 6 个大地方位角夹角示意图	28
Figure 3-10	Schematic diagram of calculating the included angles of six geodetic azimuth angles	28
图 3-11	三个连续互相相邻的网格	29
Figure 3-11	Three consecutive grids adjacent to each other	29
图 3-12	后续网格计算示意图	29
Figure 3-12	Schematic diagram of subsequent grid calculation	29
图 3-13	网格化结果	32
Figure 3-13	Gridding results	32
图 3-14	level=12 时网格生成时间	33
Figure 3-14	Grid generation time when level=12	33
图 3-15	GScan 热点挖掘示意图	33
Figure 3-15	Schematic diagram of GScan hotspot mining	33
图 3-16	两个轨迹点与各六边形网格相交示意图	34
Figure 3-16	Schematic diagram of two trajectory points intersecting each hexagonal grid	34
图 3-17	波罗的海热点区域提取结果	36
Figure 3-17	TBaltic Sea hotspot area extraction results	36
图 3-18	热点区域局部放大图	38
Figure 3-18	Partial enlargement of hotspot area	38
图 4-1	航迹预测总体流程图	41
Figure 4-1	Overall flow chart of track prediction	41
图 4-2	查找与当前航向最为接近的三个相邻网格示意图	42
Figure 4-2	Schematic diagram of finding the three adjacent grids closest to the current heading	42
图 4-3	选取目标区域示意图	42
Figure 4-3	Schematic diagram of selecting target area	42
图 4-4	奔向目标区域的预测模型示意图	43
Figure 4-4	Schematic diagram of the prediction model heading towards the target area	43
图 4-5	判断两条轨迹相似示意图	44
Figure 4-5	Schematic diagram of judging whether two trajectories are similar	44
图 4-6	历史轨迹充足的路径预测方法示意图	45
Figure 4-6	Schematic diagram of path prediction method with sufficient historical trajectories	45
图 4-7	海域障碍物网格化结果	46
Figure 4-7	Sea area obstacle gridding results	46
图 4-8	网格至分界线最短距离示意图	47
Figure 4-8	Schematic diagram of the shortest distance from the grid to the dividing	47

	line	
图 4-9	海域环境政治代价建模结果	48
Figure 4-9	Maritime environmental political cost modeling result	48
图 4-10	三条测试轨迹分布图	51
Figure 4-10	Distribution diagram of three test tracks	51
图 4-11	轨迹 L_1 时刻 A 预测结果	52
Figure 4-11	Prediction results at trajectory L_1 moment A	52
图 4-12	轨迹 L_1 时刻 B 预测结果	52
Figure 4-12	Prediction results at trajectory moment B	52
图 4-13	轨迹 L_1 时刻 C 预测结果	53
Figure 4-13	Prediction results at trajectory L_1 moment C	53
图 4-14	轨迹 L_1 时刻 D 预测结果	53
Figure 4-14	Prediction results at trajectory L_1 moment D	53
图 4-15	轨迹 L_2 某时刻预测结果	54
Figure 4-15	Prediction results at a certain moment in the trajectory L_2	54
图 4-16	图 4-17 轨迹 L_3 预测结果	55
Figure 4-16	Trajectory L_3 prediction results	55
图 4-17	历史轨迹充足时 RMSE 对比结果	58
Figure 4-17	RMSE comparison results when historical trajectories are sufficient	58
图 4-18	历史轨迹充足时 DIS_{max} 对比结果	59
Figure 4-18	DIS_{max} comparison results with sufficient historical trajectories	59
图 4-19	历史轨迹不足时 RMSE 对比结果	60
Figure 4-19	RMSE comparison results when historical trajectories are insufficient	60
图 4-20	历史轨迹不足时 DIS_{max} 对比结果	61
Figure 4-20	DIS_{max} comparison results when historical trajectories are insufficient	61
图 5-1	原型系统模块设计图	64
Figure 5-1	Prototype system module design diagram	64
图 5-2	系统主要功能结构图	65
Figure 5-2	Main functional structure diagram of the system	65
图 5-3	软件系统主界面图	67
Figure 5-3	Software system main interface diagram	67
图 5-4	轨迹预处理功能	68
Figure 5-4	Trajectory preprocessing function	68
图 5-5	姿态轨迹统计结果	68
Figure 5-5	Attitude trajectory statistical result	68
图 5-6	热点挖掘功能	69
Figure 5-6	Hotspot mining function	69
图 5-7	轨迹预测结果	69
Figure 5-7	Trajectory prediction results	69

表清单

表序号	表名称	页码
表 3-1	大圆弧网格化伪代码	29
Table 3-1	Pseudo code for large arc gridding	29
表 3-2	不同层级下网格生成时间	32
Table 3-2	Grid generation time at different levels	32
表 3-3	基于全球六边形网格的热点区域伪代码	35
Table 3-3	Pseudo code for hotspot areas based on global hexagonal grid	35
表 3-4	各热点区域类型数量统计表	38
Table 3-4	Statistics of the number of hotspot area types	38
表 3-5	区域挖掘结果热度系数表	38
Table 3-5	Heat coefficient table of regional mining results	38
表 4-1	基于六边形网格的 A*寻址算法伪代码	49
Table 4-1	Pseudo code of A* addressing algorithm based on hexagonal grid	49
表 4-2	历史轨迹充足时 RMSE 对比	58
Table 4-2	RMSE Comparison when historical trajectories are sufficient	58
表 4-3	历史轨迹充足时 DIS_{max} 对比	59
Table 4-3	DIS_{max} comparison under sufficient historical trajectories	59
表 4-4	历史轨迹不足时 RMSE 对比	60
Table 4-4	RMSE comparison when historical trajectories are insufficient	60
表 4-5	历史轨迹不足时 DIS_{max} 对比	61
Table 4-5	DIS_{max} comparison when historical trajectories are insufficient	61
表 5-1	软件环境表	63
Table 5-1	Software environment table	63
表 5-2	原始轨迹数据表	65
Table 5-2	Original trajectory data table	65
表 5-3	网格姿态统计表	66
Table 5-3	Grid posture statistics table	66
表 5-4	热点区域提取结果	66
Table 5-4	Hotspot area extraction results	66
表 5-5	目标区域预测结果	67
Table 5-5	Target area prediction results	67

1 绪论

1 Introduction

1.1 研究背景与意义 (Research Background and Significance)

在经济全球化的环境下，各国联系日趋紧密，海运成为了各个国家之间贸易往来的最重要的交通运输方式，海运贸易量约占世界总贸易量的 90%，并且仍然保持着高速增长。为了提高船舶运输能力和运输效率，船舶数量在持续增加，船舶类型显现出专业化、大型化和高速化的发展趋势，随之带来海上贸易繁荣的同时也造成许多隐患。重要通道水域、港口和航道水上交通流密度持续增加，航道通航环境日益复杂，导致水上交通事故风险不断提高，造成水上交通事故频繁发生，这严重危及到社会的经济发展与人的生命安全。因此，有必要加强对船舶的监控，及时察觉船舶异常的轨迹，减少水上交通事故的潜在风险。发现船舶异常轨迹的基础工作是对船舶航行动态的预知，通过对船舶航行轨迹的预测能够及时发现船舶的轨迹变化，对船舶可能遭遇的风险提前进行规避，避免了船舶可能面临的偏航、碰撞、搁浅等风险，可以为相关决策部门提供可靠的决策依据，实现对船舶的有效监控，对海上交通管制具有重要的意义。同时，利用对船舶航行轨迹的预知，可实现船舶的调度优化、海域管理、防灾减灾等应用，从而改善航运服务，这是实现船舶智能交通服务不可或缺的关键技术之一。

AIS（船舶自动识别系统，Automatic Identification System）技术的快速发展让船舶海量轨迹数据的获取成为现实，海事部门与民用的航运数据中心都已积累了海量 AIS 历史记录，为船舶轨迹数据挖掘特别是船舶轨迹预测的相关研究提供了基础性条件。依托于船舶当前轨迹，通过对历史数据的深度挖掘以发现其背后的关键信息，实现船舶航行轨迹的实时预测，对于提高海上安全、优化航线规划以及增强国际贸易效率等具有至关重要的作用。目前，国内外学者提出了各种不同的船舶轨迹预测模型，对船舶航行轨迹的预测进行了深入的研究，在多源数据融合、模型通用性、预测精度等方面取得了重大突破，但现有预测方法仍然具有一定的局限性，例如提取部分样本数据进行模型分析、局限短距离的轨迹预测以及历史轨迹不足导致预测精度差等问题，未能充分挖掘海量 AIS 数据所蕴含的潜在价值来进一步解决船舶安全航行等问题。网格技术在数据挖掘和轨迹预测中提供了精确的空间划分，优化数据处理效率，增强模式识别准确性，有助于揭示复杂空间关系，从而提高预测和分析的质量与速度。全球离散网格是一种强大的网格分析框架，通过将连续的地理空间转换为一系

列离散的单元，极大地简化了数据的复杂性，使得我们能够在全球尺度上模拟和分析各种复杂的地理现象和过程，易于扩展至全球及区域范围的热点区域提取与轨迹预测。这种离散化的处理不仅便于实现对船舶动态轨迹的实时监测和分析，而且还能有效地支持复杂的轨迹预测模型的构建和验证。船舶航行是一种大范围或全球性的活动，涉及到复杂而广阔的海域，为更好地模拟和分析船舶航行的区域性和全球性特征，全球离散网格成为了一种有效的工具。通过在全球离散网格上进行船舶轨迹预测研究，可以更加有效地仿真和研究船舶航行的局部与全球特性，探索船舶运动规律、交通流动和风险评估等问题。

本文基于中国电子科技集团第 54 所的委托项目进行选题，以民用船舶代替敏感目标进行海上移动目标的热点挖掘与轨迹预测，为海域安全、国土安全及其态姿研判提供技术支持与方法借鉴。

1.2 国内外研究现状（Reviews on Related Researches）

1.2.1 全球离散网格

全球离散格网系统（Discrete Global Grid System, DGGS）采用特定的方法将地球离散化，形成无缝无叠的多尺度格网结构，用编码等效替代传统地理坐标参与数据操作^[1]。DGGS 的发展始于 20 世纪 80 年代，作为一个处理全球地形数据的分析框架，该框架已发展成为一个空间参考系统，并已被集成成为一个数据结构，用于全球一致地存储、使用和分析空间信息，以及相应的辅助信息。发展至今，多种全球离散网格被提出，按照网格剖分的方式主要分为：等经纬度格网^[2, 3]、变间隔经纬度格网^[4, 5]、自适应格网^[6, 7]与多面体格网^[8, 9]。其中，多面体全球离散网格是借助于正多面体对参考椭球面进行递归剖分，并采用投影方式将正多面体上的点通过逆投影至球面而形成的格网系统。相对于经纬度网格和自适应网格，多面体离散网格单元图形在地球表面各区域具有相似性与层次性，能够避免高纬度的数据冗余，同时更易构建全球数据管理模型。三角形、四边形与六边形是三种常用的构建多面体全球离散网格的几何图形，相较于三角形与四边形，六边形全球离散格网系统展现出优异的几何特性，具有覆盖效率高、邻近关系最简单、各向同性等优势，更有利于空间数据的组织、处理和分析，不仅是封闭球面上的最佳数据采样几何结构，而且也是支持多尺度计算的拓扑结构，同时也是位置相关事件模拟与推演的首选方案^[9, 10]。构建及编码构成了全球六边形离散网格技术的根基，对此，学界已提出众多多元化的解决方案。Sahr^[9]采用中心投影和施耐德投影相互转换的方法实现多面体与球面的映射，建立正二十面体全球离散格网。童晓冲等人^[11]利用数值投影变换理论，通过多项式逼近方法，基于地面控制点的变形指标，建立了多面体到球面的映射

关系，从而实现了全球六边形离散格网系统的搭建。针对正二十面体的三角面不利于建立格网单元行列索引和生成的球面格网单元存在较大形状差异等问题，梁晓宇等^[12]提出了菱形三十面体六边形全球离散格网系统的构建方法。Sahr^[13]提出了多分辨率三孔等积六边形格网，在改进广义平衡三进制基础上建立了二十面体上的唯一编码方案。王蕊等人^[14]根据平面四孔六边形格网系统结构特点，设计格点四叉树层次编码结构，实现了二维直角坐标与单元编码的相互转换，实验表明编码运算效率高于同类方法。贲进等^[15]研究的三孔六边形全球离散格网系统代数编码方案提高了格网系统的运算效率。

经过数十年的沉淀，全球六边形离散网格理论与技术已日益成熟，为地理空间数据编码、寻址、离散化等研究开辟了新路径，广泛应用于数据可视化、空间数据索引、制图综合以及军事模拟等领域^[16]。为解决室内外数据模型兼容性低、室内模型缺乏统一空间参考框架的问题，邬群勇等^[17]在室内模型通过引入全球六边形离散格网系统作为坐标参考系统提供位置编码信息，并证实该模型具有在统一框架下空间规划计算的可行性和有效性。加拿大劳里埃大学的Robertson^[18]基于六边形格网开发了数据集成与分析系统 IDEAS，适用于大规模数据建模和分析。Li MK^[19]将异构地形数据集成到离散全球网格系统中，在此基础上并在此框架上进行了量化、积分和聚合等操作，完成多分辨率高程数据的建模。周建斌^[20]利用六边形全球离散格网系统构建了广域多尺度的兵棋地图。通过推演案例展示，该方法具有推演范围广、灵活性高等优点，弥补了传统兵棋地图在空间范围受限、尺度变换不灵活等缺陷。梁启爽等^[21]基于菱形三十面体六边形全球离散格网系统建立遥感图像六边形像素数学模型，保证了六边形像素遥感图像数据与标准文件格式兼容，打破六边形单元与矩形像素遥感图像的数据组织壁垒，为遥感图像数据组织提供了可行的技术途径。

1.2.2 热点区域挖掘

根据韦氏词典中的定义，热点区域是指一个比其他区域具有更多的兴趣点、人类频繁活动的地理区域^[22-24]，本文定义为船只频繁经过且其达到一定停留时长的区域。国内外的研究方法主要是通过对大量历史轨迹进行聚类，提取轨迹点分布较为密集的聚类簇，进而挖掘出热点区域^[25-27]。聚类是一种非监督式学习方法，基于相似性将对象聚集成不同的类簇或子集，使同一个类簇中的对象都具有相似的属性^[28-30]。

经过了数十年的发展，轨迹聚类涌现了大量相关研究，大致可划分为：基于划分的聚类算法、基于网格的聚类算法、基于密度的聚类算法与基于层次的聚类算法。MacQueen^[31]于 1967 年首次提出 k-means 聚类算法，该算法的基本思想为：从数据集中根据一定策略选择 k 个点作为每个簇的初始中心，并依

照最邻近原则，基于距离划分簇并进行调整，再将各数据聚集到它最近的聚类中心所代表的簇中，k-means 算法开启了聚类算法的历史先河。在 1998 年，Agrawal 等人^[32]提出 CLIQUE 聚类算法，该算法有效结合了基于密度和网格的优点，可以自动识别数据子空间中的簇，同时整个聚类过程相比于其他算法更为高效。一年后，基于高密度的 DBSCAN 聚类算法被 Ester M 提出^[33]，DBSCAN 是一种无监督的 ML 聚类算法，不要求我们指定集群的数量，避免了异常值，并且在任意形状和大小的集群中取得较好的效果。2000 年，一个重要的子空间聚类方法 PROCLUS 面世，该算法通过迭代地识别数据中的相关子空间，并将数据点分配到这些子空间中以实现聚类，适合发现形状类似超球面的数据簇，其聚类过程相对于 CLIQUE 算法更为高效^[34]。

各国学者对经典聚类方法提出了各种改进策略，成功应用到不同的复杂场景，并取得不错的聚类效果。Li 等^[35]基于层次聚类的思想设计了一种移动微聚类算法(moving micro-clustering, MMC)，并应用在移动轨迹数据的聚类中。Tu^[36]提供了一种基于 BIRCH 算法的在线分隔时间序列数据的方法，该方法具有出色的时间序列模式检测能力，同时时间复杂度也相对较低，能够实现轨迹的精确聚类。Deng 等人^[37]提出一种名为 GRIDEN 的空间算法，基于密度和网格实现聚类，同时支持多密度聚类和并行计算。该算法通过超方格单元构建网格，并允许用户通过参数 k 平衡算法的效率和准确度，从而提升了算法的灵活性。Izakian 等学者^[38]提供了一种自动化模糊聚类粒子群优化技术，用于发现移动轨迹数据结构特征。该技术利用粒子群优化方法寻找最优的聚类簇数量和最佳的聚类中心，以提升算法性能。Agrawal 等^[39]在 OPTICS 聚类算法的基础上提出了 ST-OPTICS 聚类算法，在簇排序过程中增加对数据对象的时间戳考虑，为获得高质量、任意形状与嵌套的聚类提供可行技术途径。Cai M 等^[40]研究出一种基于改进的 CURE 算法的大规模网络文本数据聚类算法，该算法采用了与统计思想相结合的样本提取算法，根据不同的数据密度合理选择样本点，实验表明运算效率快且聚类结果准确，因此在大型数据挖掘中应用广泛。针对现有热点区域挖掘方法面对大规模轨迹数据时伸缩性差，计算效率低等问题，赵林江等^[41]设计了一种基于网格密度的 GScan 聚类算法，该算法首先将轨迹空间划分成网格单元，设定网格单元的密度阈值提取热点网格，最后完成热点区域的提取。王培晓等^[42]在基于密度峰值的快速聚类算法中引入了时间约束，并改进了原始算法的样本属性计算策略，使得新算法 ST-CFSFDP 能够区分同一区域不同时间下的簇集，同时应对单个簇集出现多个密度峰值的情况有着良好表现。Wang^[43]在热点挖掘领域引入了两级细分概念，并对 K-Means++算法进行改进，以实现出租车乘客热点的精细聚类，优化了研究中存在的聚类精度低和算法时

间复杂度高等问题。Wang L 等人^[44]研究了一种利用 Hausdorff 距离和层次密度进行带噪声应用空间聚类的船舶 AIS 轨迹聚类方法，该方法能够自适应地聚类具有形状特征的船舶轨迹，并具备良好的聚类可扩展性。传统的聚类算法在面对大量的 AIS 数据样本时通常表现出很低的执行效率，胡昕源^[45]提出一种改进 QuickBundles 算法，并对船舶轨迹采样方法和距离度量方式进行改进，最终实现船舶轨迹的有效聚类。

1.2.3 活动轨迹预测

活动轨迹预测是指根据移动对象当前的轨迹预测出该对象未来的移动轨迹。现有的对于航迹预测技术的研究主要分为三种：基于统计概率的轨迹预测^[46, 47]、基于机器学习的轨迹预测^[48, 49]与基于滤波与仿生的轨迹预测^[50, 51]。

(1) 基于概率统计的轨迹预测

基于概率统计的航迹预测方法假设历史轨迹数据与预测轨迹数据之间存在一定的相关性，通过移动目标的历史轨迹数据来构建相应的数学模型。目前其经典的轨迹预测方法有基于马尔科夫链（Markov Chain, MC）方法、基于隐式马尔科夫模型（Hidden Markov Model, HMM）方法、高斯混合模型（Gaussian mixture model, GMM）与贝叶斯网络等。MC^[52]假设目标的运动状态是一个随时间变化的马尔科夫过程，即目标在每个时刻的运动状态仅与前一个时刻的状态相关，而与之前的状态无关，因此可以通过概率矩阵描述目标各个状态之间的转移概率。HMM 与 MC 类似，通过目标运动过程中的隐蔽状态计算目标下一时刻最有可能的位置^[53]。但是马尔科夫类模型在轨迹预测中需要构建目标的转移概率矩阵，随着目标轨迹数量的增加，构建的矩阵规模也相应增加，造成计算效率的降低。GMM 由多个高斯概率分布的密度函数组合而成，通过期望最大化（EM）算法迭代地估计高斯分布的参数，以最佳方式拟合历史轨迹数据进行建模，该模型通过计算移动目标路径的综合分布概率，并将现有轨迹信息输入，从而实现对目标未来路径的预测^[54]。贝叶斯网络本质上是一个用于表达变量间概率依赖关系的图形化模型，它通过有向无环图描述变量之间的条件依存性，从而能够进行复杂概率推理和预测^[55]。乔少杰^[56]依据贝叶斯网络原理，研究出一种基于轨迹时间连续贝叶斯网络的移动目标不确定轨迹预测方法，深入分析了移动对象的速度和方向对其未来路径预测的显著影响。基于概率统计的轨迹预测方法具有简单易实现、适用性广泛等优点，但在处理复杂轨迹预测问题和提高预测精度方面存在一定的局限性。

(2) 基于机器学习的轨迹预测

在轨迹预测领域，借助 GPU 等硬件资源的升级以及信息化时代数据量的爆炸性增长，机器学习和深度学习技术得到了迅速的发展和广泛应用。通过大规

模历史轨迹数据的训练，这些方法能够从海量数据中挖掘出隐藏的模式和规律，并将其应用于未来目标轨迹的预测中。文献^[57]提出基于 BP 神经网络的船舶轨迹预测算法，根据船位差预测船舶的轨迹，设计以预测船位差实现航迹预测的思路，证实了 BP 神经网络船舶轨迹预测算法不仅用时短，而且预测的准确性高。循环神经网络（recurrent neural network，RNN）能够捕捉时间特征且具有短时记忆功能，Min K^[58]首先基于 RNN 对车辆的障碍物路径进行预测，然后通过预测路径的不确定性预测车辆的轨迹。然而，在使用 RNN 进行长期轨迹预测时，不恰当的参数设置可能会导致梯度消失或爆炸问题，进而使预测结果的准确度大幅下降。长短期记忆（long short-term memory，LSTM）网络在 RNN 结构的基础上增加了输入门、遗忘门和输出门，实现了对状态信息的筛选，解决了神经网络中长序列依赖等问题^[59-61]。针对地面移动目标，利用 LSTM 算法能准确预测高速公路上车辆的未来纵向和横向轨迹，同时可对周边车辆进行远视轨迹预测，有效提高交互式驾驶环境中的预测精度^[62, 63]。针对海上移动目标特征复杂的特性，LSTM 算法可将船舶的特征作为网络的输入，通过实验表明基于 LSTM 相对于传统的航迹推算法与 BP 神经网络预测法具有更优的性能^[64]。文献^[65]基于深度学习技术提出了一种新型的预测模型，该模型采用循环神经网络-长短期记忆（RNN-LSTM）框架，并利用 AIS 数据进行航迹特征分析的实证研究。实验结果表明，与传统预测算法相比，该模型不仅实现简便、预测准确性高，而且在处理时间序列数据方面更为高效。针对现有船舶轨迹预测模型预测准确度低的问题，郭逸婕等^[66]构建了一种基于注意力机制的时域卷积网络和双向长短时记忆网络结合的船舶轨迹预测模型，该模型首先搭建 TCN 网络（temporal convolutional network）提取船舶轨迹的序列特征，后搭建 Bi-LSTM 网络学习轨迹序列的前后状况来提取序列中更多的信息，实现对船舶未来轨迹的预测。基于机器学习的轨迹预测对历史轨迹数据特征的学习更加准确，弥补了传统概率统计方法预测精度不高的缺点，适用于大数据集的预测任务，但该方法存在训练速度慢、内存消耗大与模型参数选取困难等问题。

（3）基于滤波与仿生的轨迹预测

在基于滤波与仿生的轨迹预测方法中，目前比较经典的有遗传算法、粒子群算法与卡尔曼滤波算法。遗传算法(Genetic Algorithms, GA)通过模仿达尔文进化论中的遗传选择和自然淘汰机制，实现了一种全局概率搜索策略。该算法拥有自我适应、全局寻优和内在并行处理等特性，体现出卓越的全局搜索能力。Baklacioglu T 等^[67]采用遗传算法预测飞机飞行轨迹，首先对飞机飞行轨迹数据进行归一化处理，然后建立了合理的函数集和拟合评价标准最后利用遗传规划算法对飞行轨迹进行预测，通过仿真结果表明该算法预测精度较高，拟合轨迹

与实际飞行轨迹吻合。谭伟等^[68]提出了基于遗传算法的 BP 神经网络的航迹预测模型，结合遗传算法有效克服了 BP 神经网络易陷入局部最小的缺点，得到了更高的学习精度和更快的收敛速度。在文献^[69, 70]提到，粒子群算法是一种模拟生物行为的群智能优化算法，具有结构简单且易于实现、容易收敛且鲁棒性较强的优点，实验表明其预测精度高且算法收敛快。卡尔曼滤波器（Kalman filter, KF）能够对移动目标的状态行为进行估计，通过不断更新状态估计值，结合观测数据和系统动态模型，实现对轨迹的动态预测。Prevost C G 等人^[71]首先将 KF 用于轨迹预测，随后 KF 被应用于短距离的车辆轨迹预测^[72]，基于卡尔曼滤波的动态轨迹预测算法^[73]经过真实移动目标数据集验证获得了较好的预测效果。文献^[74]研究了无迹卡尔曼滤波算法在目标跟踪的应用情况，该算法首先选定了特定时间下的历史轨迹，并依据遗传算法来对航行经纬度、速度及航向进行动态分析，结果表明这种预测方法具有一定的准确性。文献^[75]提出了基于卡尔曼滤波的船舶轨迹预测算法，通过引入系统噪声和测量噪声对卡尔曼滤波算法进行修正，再以 AIS 船舶轨迹数据为基础，使用最小二乘法对船舶轨迹预测进行建模，实现对船舶轨迹的准确预测。由上可知，基于滤波与仿生的轨迹预测方法在融合多传感器信息、适应复杂环境和实时性方面具有优势，但在复杂性和参数调节上存在一定的挑战。

综上可知，目前现有的活动轨迹预测方法在预测精度、模型适用性、预测性能等方面取得了一系列的进展，但主要集中在对单步或者段短距离的预测上，通常限制于较小的空间范围进行轨迹的预测与分析。然而，当需要进行长距离或者大范围的预测时，预测结果与真实轨迹偏离较大。船舶航行往往是大范围乃至全球性的活动，因此迫切需要一种能够在大范围视角下进行分析且具备较高预测精度的方法对其进行预测分析。此外，大多预测方法较为依赖于历史轨迹的数量与质量，当历史轨迹不足时，往往预测效果不佳。

1.3 研究内容与技术路线图（Research Content and Technical Roadmap）

1.3.1 本文研究内容

针对现有航迹轨迹预测方法长远距离预测误差大、历史轨迹不足精度低等问题，本文借助全球六边形离散网格技术，围绕轨迹网格化、热点区域提取与船舶轨迹预测展开研究，且完成相应原型系统的设计与开发。主要研究内容如下：

(1) 航迹网格化。研究一种直接基于球面的航迹网格化算法，能够准确地实现对任意航迹的快速网格化，并开展实验验证算法的性能，为后续的热点区域提取与轨迹预测提供技术基础。

(2) 热点区域挖掘。在航迹网格化的基础之上，面向潜在航行目的地的计算需求，研究并设计一种基于全球六边形网格的热点区域挖掘方案，并设计相关实验验证方法的可行性与准确性。

(3) 船舶轨迹预测。针对已有船舶轨迹预测方法长远距离预测误差大、历史轨迹不足精度低等问题，研究一种顾及目的地及航道权等多重约束的船舶轨迹预测方法，探讨在历史轨迹充足与欠缺的情况下的预测策略，并设计定性与定量实验对预测方法进行验证评估。

(4) 航迹预测原型系统开发。在完成理论研究与实验验证的基础上，结合项目需求与研究成果完成航迹预测原型系统的开发，原型系统能够充分挖掘海量历史轨迹数据的运动规律，实现对船舶轨迹的实时预测。

1.3.2 技术路线图

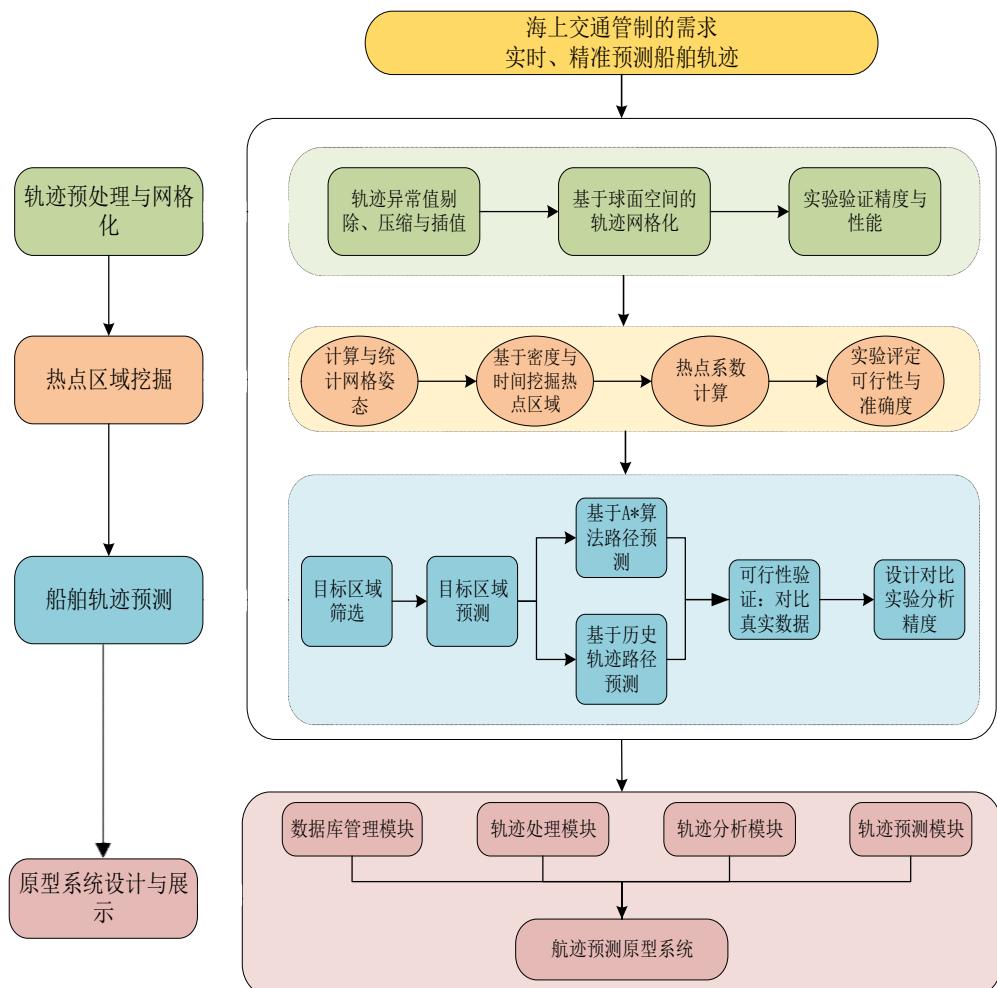


图 1-1 技术路线图

Figure 1-1 Technology roadmap

1.4 论文章节安排 (Chapter Arrangement of the Paper)

本文以探索船舶轨迹预测方法为主要内容，结合上述研究内容与技术路线，本文共分为六个章节，具体安排如下：

第一章：绪论。首先，叙述在经济全球化时代下的船舶轨迹预测的研究背景与意义；其次，归纳与分析全球离散网格、热点区域挖掘以及航迹预测技术的研究现状，并总结分析现有方法存在的问题。最后，阐述本文的研究内容与技术路线。

第二章：全球六边形网格与航迹预测技术。首先，从网格构建与网格编码介绍全球六边形离散网格技术；其次，详细介绍基于 LSTM 航迹预测技术的原理与模型构建过程。此章作为后续研究的理论基础。

第三章：基于全球六边形网格的热点区域挖掘。首先，介绍本文热点区域挖掘的总体思路；然后，逐一介绍本文热点区域挖掘的几大步骤：数据预处理、轨迹网格化及热点区域计算等；最后，设计相关实验来验证评估热点提取方法的可靠性与准确性。

第四章：基于全球六边形网格的船舶轨迹预测。首先，提出本文船舶轨迹预测的总体思路；然后，详述预测方法的主要两个步骤：目标区域预测及路径预测；最后，分别设计定性与定量实验对本文预测方法的可行性与准确性进行验证评估。

第五章：航迹预测原型系统开发与展示。首先，从系统开发环境、系统功能以及数据库表设计来详细说明原型开发系统的主要技术与系统功能。然后，通过主要界面功能以及预测成果进行展示开发系统。

第六章：总结与展望。本章对全文的研究工作和成果进行总结，并对后续值得进一步研究的内容进行展望。

2 全球六边形网格与航迹预测技术

2 Global Hexagonal Grid and Trajectory Prediction Technology

2.1 全球六边形网格 (Global Hexagonal Grid)

网格构建与网格编码是全球六边形离散网格技术主要的两个环节，它们共同构成了这一技术框架的基础，以下从这两个方面对全球六边形离散网格进行详细介绍。

2.1.1 网格构建

多面体全球离散格网数据模型的基本思想是将地球递归剖分为形状、面积近似相等且具有规则层次结构的单元，每个单元在记录位置信息的同时也表达比例尺和精度。构建全球六边形网格系统首先需要在多面体表面通过递归剖分建立规则的格网层次关系，形成多分辨率六角格网，之后再映射至地球表面。全球六边形离散网格的构建包括多面体选择、多面体定位、多面体网格剖分与投影映射共四步，分述如下：

(1) 多面体选择

为了确保多面体各面展现一致的空间属性，并高效管理全球尺度的空间分辨率数据，以适应广阔区域应用的需求，众多学者探索基于柏拉图正多面体构建的球面空间网格模型^[76]。柏拉图固体包括四面体、六面体、八面体、十二面体和二十面体，如图 2-1 所示。事实上，数学上可证明只有 5 种这种类型正多面体投影到球面上能产生形状相同的球面多边形且顶点所在多边形的数目相等^[77]。一般来说，柏拉图固体的表面面积越小，多面体面与相应的球面之间转换时引起的扭曲误差相对来说就越小，同时所得的网格单元的形状及面积就越均匀。正二十面体具有相对最小的面尺寸，是一个基本的柏拉图式固体最常见的选择。考虑到项目与本文实验有单元面积统计的需求，为此，采用正二十面体为正多面体，作为全球六边形网格剖分的基础。



(a) 正四面体 (b) 正六面体 (c) 正四面体 (d) 正四面体 (e) 正二十面体

图 2-1 五种柏拉图式正多面体示意图

Figure 2-1 Schematic diagram of five Platonic regular polyhedrons

(2) 多面体定位

在选择多面体后，需要确定多面体与地球之间的相对位置关系，一般通过正多面体某顶点的地理坐标和相应边的方向角来表示。Goodchild 和 Yang^[78]将八面体顶点定位在南北极点和八分点上保证全球对称以简化坐标与编码转换。Fuller^[79]为“动态最大张力投影”设计的二十面体定位方案，将 12 个顶点放在海洋中以保证陆地毫无破裂地展开到二十面体表面，但该方案并不关于赤道对称。Alborzi 和 Samet^[80]提出了将多面体的面心放置在南北极的定位方案，保证了极地区域的完整性和对称性。Sahr^[9]借鉴 Fuller 的思想，减少落在陆地上的顶点个数并保证多面体关于赤道对称，最理想的情况是只有一个顶点落在我国四川省。考虑到研究区域的地理位置，为减少跨面计算的复杂度，本文多面体定位采用 Sahr 提出来的方案，如图 2-2 所示：



图 2-2 多面体定位
Figure 2-2 Polyhedral positioning

(3) 多面体表面剖分

选定初始正多面体之后，需选择一种递归剖分方法来剖分正多面体。由于正多边形的每个面均一样，为此，仅需对一个面进行递归剖分设计即可。正二十面体的每个面均为三角形。在三角形基础上进行六边形递归剖分的方案有多种，可按照孔径和连接方式来划分。相邻层次单元面积之比称为孔径，常见的孔径有三孔、四孔和七孔三种类型。通过边的中点产生新边的连接称为 I 类，而通过现有单元内缩的方式产生新边的连接方式称为 II 类。低孔径剖分产生的格网分辨率间隔更小，因而分辨率之间过渡比较平滑，而第 II 类具有单元方向不变特性。本项目后续需要对航迹进行网格化，航迹网格化算法将使用到网格

的邻近计算。若网格单元的方向保持不变，将有利于邻近计算的简化。为此，采用 II 类孔径 4 的网格递归划分方法，如图 2-3 所示。

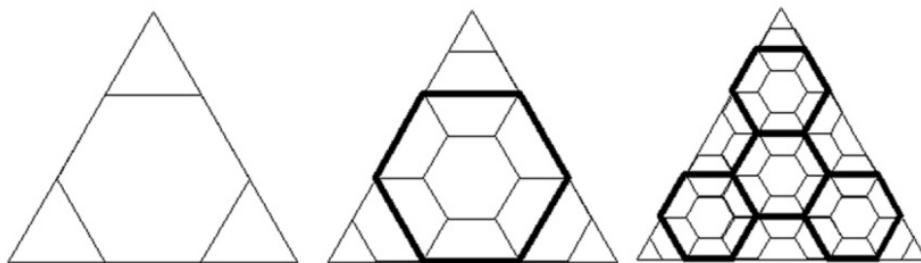


图 2-3 II 类孔径 4 六边形划分示意图

Figure 2-3 Schematic diagram of the hexagonal division of Class II aperture 4

(4) 投影映射

一旦在基多面体的面或面上指定了划分方法，就必须选择一个投影变换来在相应的球面创建类似的拓扑，以球面为原面、多面体表面为投影面建立两者的映射关系。一般有两种的基本方法，一是直接球面细分法，其直接在球面上创建一个分区，该分区映射到平面上相应的分区；二是映射投影方法使用逆映射投影将在平面上定义的分区转换为球面。杨帆等人^[81]通过应用 Fuller 投影法，构建了正八面体与球面之间的映射，与施耐德投影进行对比后发现，Fuller 投影生成的球面网格单元具有更低的形变程度，并且分布更为均匀一致。贲进^[82]通过对施奈德等积多面体投影的研究，建立多面体和球体两者的映射关系，实现了球面能够完整、均匀地映射到正多面体表面。为保证经纬网在各面上的连续性与网格的等积性，Sahr^[9]提出了基于中心投影和施耐德投影相互转换的映射方案。为了确保经纬网在边界处的连续性，考虑到对等面积多面体投影有一定要求，同时考虑到本项目的面积统计需求，因此本文采用 Sahr 提出的映射方法，该投影方法如图 2-4 所示：

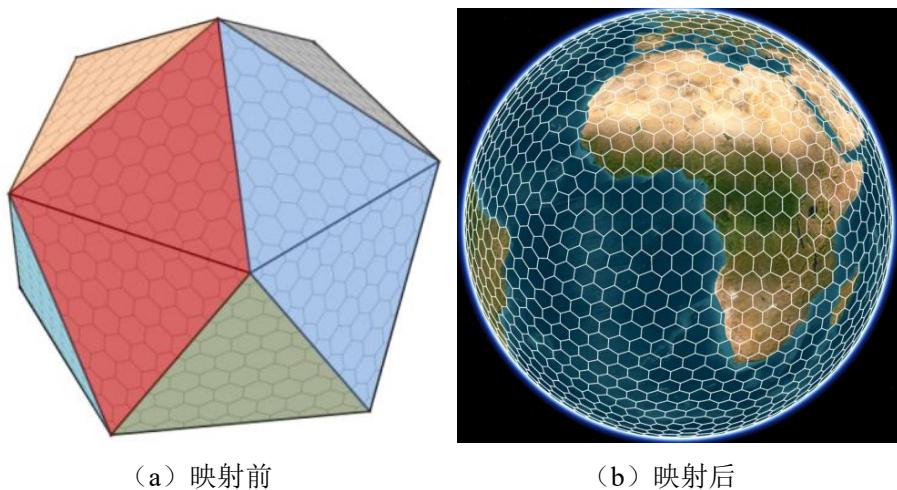


图 2-4 球面与多面体投影映射图

Figure 2-4 Spherical and polyhedral projection mapping diagram

2.1.2 网格编码

在将地表剖分为均匀统一的层次格网单元结构后，需要给每一个单元赋予全局唯一的地址编码，其能够代替传统地理坐标参与数据操作和运算。正二十面体相邻两个三角面可组成一个菱形面，沿着菱形的边展开可得图 2-5 所示结果。展开后共有 10 个菱形区域及两个五边形区域。12 个菱形区域依次编码为 1-10，两个五边形区域分布对应南北极点，北极对应的五边形区域编码为 0，南极对应的五边形区域编码为 11。

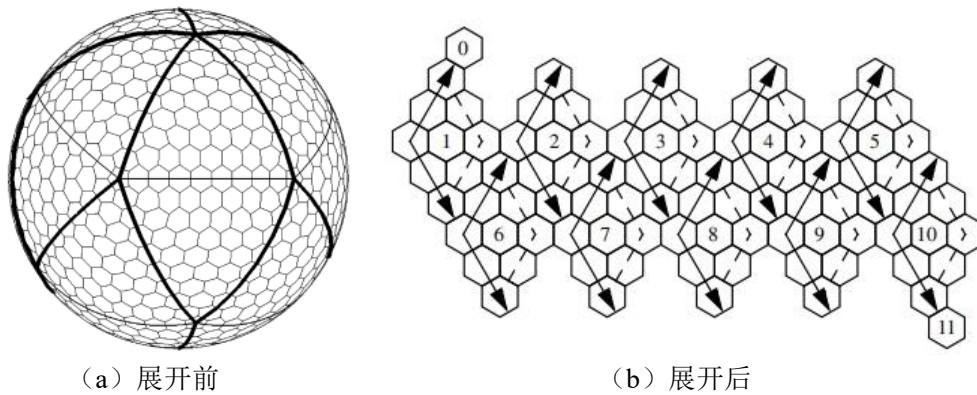


图 2-5 二十面体展开示意图

Figure 2-5 Schematic diagram of icosahedron expansion

传统的平面笛卡尔坐标系采用两个相互垂直的坐标轴。然而，六边形系统自然形成三个轴，彼此形成 60 度角。如图 2-6 所示，相对于传统的笛卡尔直角坐标系，这三个轴有两个自然方向，我们将其称为第 I 类和 II 类，作为 Fuller 为三角形网格系统开发的术语的推广。

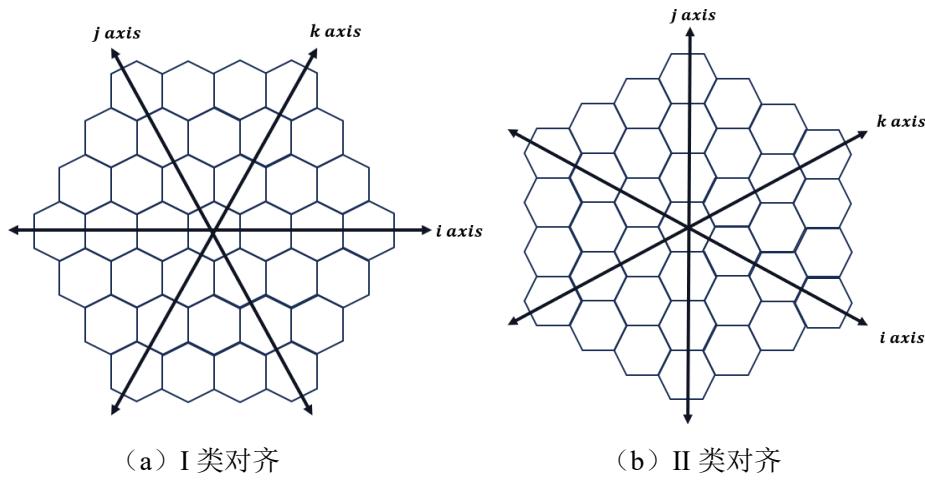


图 2-6 三轴六边形坐标系统

Figure 2-6 Three-axis hexagonal coordinate system

任意一个菱形区域面的网格唯一对应正二十面体的一个三角形区域。为此，菱形区域内网格编码可参照平面空间六边形网格编码方法，在某一个菱形面建立二轴坐标系，按照轴坐标法对菱形区域的每个网格单元给出序号坐标，记为

i 轴与 j 轴，其中 $0 \leq i \leq n$ 与 $0 \leq j \leq n$ (n 为轴上的六边形网格目)，如下图所示：

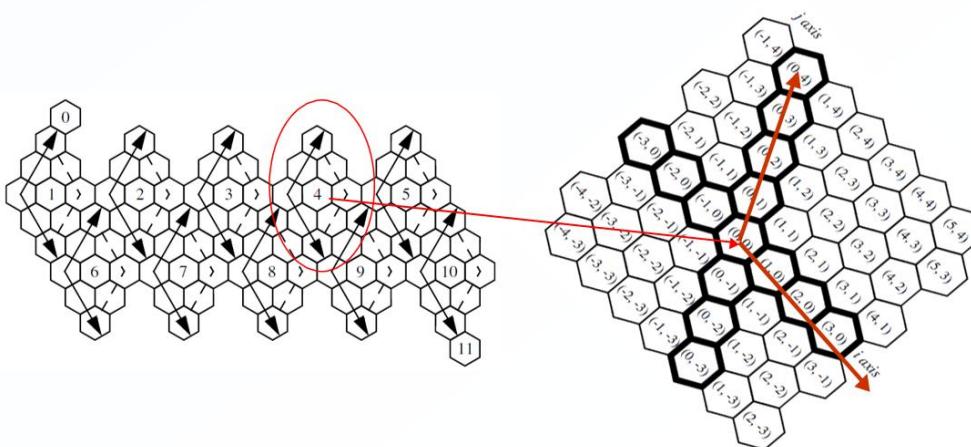


图 2-7 菱形面坐标轴示意图

Figure 2-7 Schematic diagram of the coordinate axis of the rhombus surface

获取轴坐标之后，将层次信息、菱形块编号、序号信息按一定方式组合，形成一个唯一的金字塔地址形式线性编码： $\{r, [q, (i, j)]\}$ ，其中 r 是六边形网格系统的分辨率， q 是单元格的四边形编码， (i, j) 对应于在相应菱形内的轴坐标。

2.2 基于 LSTM 的航迹预测（Trajectory Prediction Based on LSTM）

长短期记忆网络（LSTM）是一种特殊的循环神经网络（RNN），通过引入三个关键门结构实现了对信息流的精细控制，其强大的时序记忆能力使其能够很好地发现时间序列中所隐含的规律。LSTM 网络将 RNN 的隐含层单元替换为 LSTM 细胞，可有效地克服传统 RNN 在处理大型时间序列数据时可能会出现的梯度爆炸和梯度消失等问题，这种结构使 LSTM 特别适合于长期序列数据的预测任务，因为它能够捕捉并维持长期依赖关系，从而在长距离时间序列预测中，相比于其他类型的神经网络模型，提供更高的准确率和稳定性，显著提升了对未来事件的预测能力。LSTM 凭借其具有较强的时序建模能力、记忆能力、处理不规则变长序列的能力与准确的预测等优点，被广泛应用于各种轨迹预测场景领域^[83, 84]。

2.2.1 LSTM 原理

LSTM 网络的基础构成包括细胞状态和门控机制，由三个主要的门控单元——遗忘门 (f_t)、输入门 (i_t) 和输出门 (O_t)，以及记忆细胞组成。这三种门通过独特的交互机制共同作用，实现对网络信息流的精细管理。与传统 RNN 的隐含层单元不同，LSTM 引入了记忆模块，每个模块都含有一个或多个

可以维持内部状态的记忆细胞，从而优化了信息的长期保持与传递。LSTM 网络模型结构如图 2-8 所示。

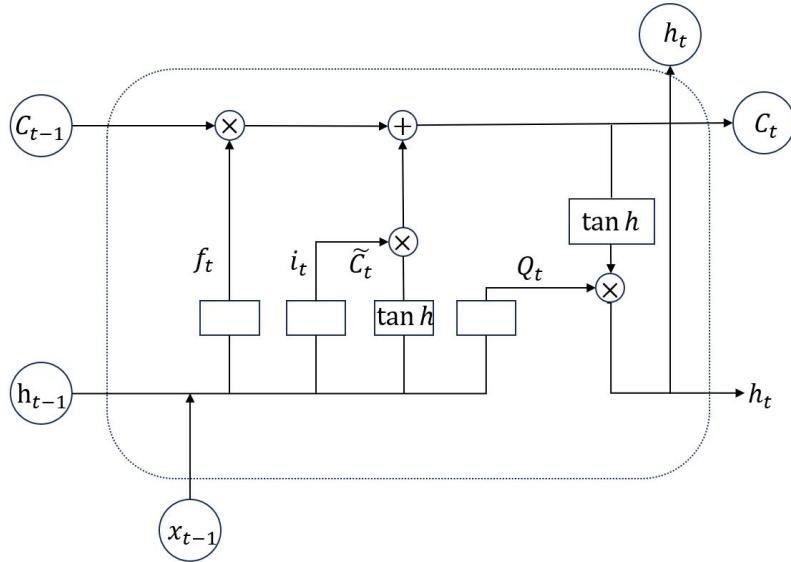


图 2-8 LSTM 网络结构示意图

Figure 2-8 Schematic diagram of LSTM network structure

如图 2-8 所示，其中每一个箭头都表示着一个向量，意味着从一个节点的输出到另一个节点的输入，圆圈则表示向量的相加或相乘计算，箭头分叉表示两个节点复制。LSTM 计算过程如下：

(1) 遗忘门决定哪些信息从细胞状态中被遗忘或丢弃。通过接收前一个隐含状态和当前输入，遗忘门输出一个 0 到 1 之间的值，用于乘以细胞状态，从而过滤掉不重要的信息。计算公式如下：

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (2-1)$$

上式中： x_t 为当前时刻的输入； h_{t-1} 为上一时刻隐含层的输出； f_t 为忘记门的输出； W_f 、 b_f 分别为忘记门的权重矩阵和偏值项； σ 为 sigmoid 激活函数。

(2) 模块中的输入门决定哪些新信息是重要的，并应当被更新或添加到细胞状态中，该过程主要分两步来进行：第一步 sigmoid 层决定哪些值被更新，计算公式如下：

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (2-2)$$

其中， W_i 、 b_i 分别为输入门的权重矩阵和偏值项；第二步 $\tan h$ 层产生了一个候选向量 \tilde{C}_t ，将加入到细胞中，表达式如下：

$$\tilde{C}_t = \tan h(W_c[h_{t-1}, x_t] + b_c) \quad (2-3)$$

上式中， W_c 、 b_c 分别为权重矩阵与偏值项， $\tan h$ 为双曲正切函数。

(3) 将遗忘门与输入门输出的两个值结合起来并更新细胞状态 C_t ，计算如下：

$$C_t = C_{t-1} \times f_t + \tilde{C}_t \times i_t \quad (2-4)$$

(4) 最后, 输出门基于新的细胞状态确定输出内容, 输出门决定控制单元 C_t 输出到 O_t 的信息有多少能输出到隐藏层 h_t 中, 用来确定下一个隐藏状态的值。

$$O_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (2-5)$$

$$h_t = O_t \times \tanh(C_t) \quad (2-6)$$

上式中, W_o 、 b_o 分别为输出门的权重矩阵与偏值项; h_t 为隐含层输出。

2.2.2 LSTM 模型构建

LSTM 网络模型的训练采用了时间反向传播算法 (Back Propagation Through Time, BPTT), 这一算法在本质上与传统的误差反向传播算法 (Back Propagation, BP) 相似, 都涵盖了正向传播和反向传播两个过程。为了深入挖掘航迹历史数据, 以船舶航向、航速和经度、纬度作为 LSTM 网络的输入, 同时以这四个变量作为输出来构造实时航迹预测模型, 模型结构如图 2-9 所示:

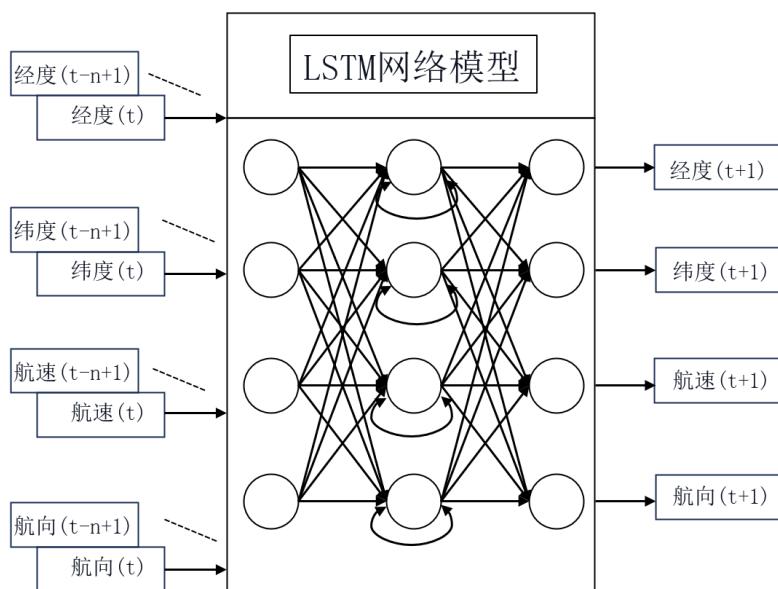


图 2-9 LSTM 航迹预测模型结构示意图

Figure 2-9 Schematic diagram of the structure of the LSTM track prediction model

假设某目标船舶在 t 时刻的航行状态特征表示为:

$$S_t = \{x_t, y_t, v_t, \alpha_t\} \quad (2-7)$$

对于上式: x_t 、 y_t 、 v_t 、 α_t 分别为船舶在 t 时刻的经度、纬度、航速与航向。一般, 采用船舶的 t 时刻及其前 $n - 1$ 个时刻的经度、纬度、航速、航向作为 LSTM 网络的输入, $t + 1$ 时刻船舶的经纬度坐标、航速与航向作为输出, 建立 LSTM 航迹预测模型, 则 $t + 1$ 时刻船舶位置可以表示为:

$$S_{t+1} = f(S_t, S_{t-1}, \dots, S_{t-n+1}) \quad (2-8)$$

在上式中, f 表示所训练的模型。对于 $t + 2$ 的航行状态同理以船舶的 $t + 1$ 时刻及其前 $n - 1$ 个时刻的经度、纬度、航速、航向作为 LSTM 网络的输入进行预测。上述时间序列预测的 n 步预测输出是一个通过不断迭代的过程来进行实

现的，如图 2-10 所示：

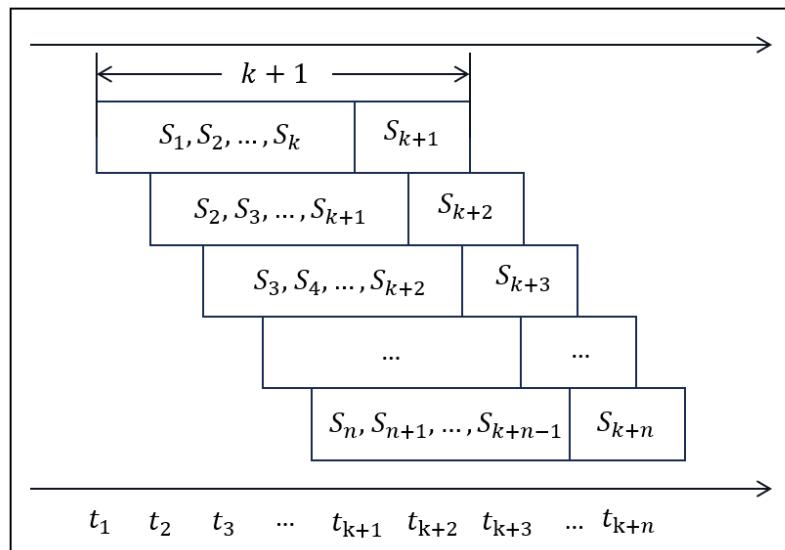


图 2-10 时间序列预测迭代过程示意图

Figure 2-10 Schematic diagram of the iterative process of time series forecasting

2.3 本章小结 (Summary of this Chapter)

本章介绍了全球六边形网格与航迹预测相关技术。首先从网格构建和网格编码两个维度具体介绍了全球六边形离散网格，其中网格构建包括多面体选择、多面体定位、多面体表面剖分与投影映射。之后详细阐述了基于 LSTM 网络模型的航迹预测方法，对其原理与模型构建做了详尽介绍。

3 面向船舶轨迹预测的热点区域挖掘

3 Hot Spot Region Mining Based for Ship Trajectory Prediction

热点区域的识别与分析，在船舶轨迹预测的研究领域中，不仅揭示了船舶在特定时间段内的活动模式和偏好区域，而且还为预测船舶未来运动趋势提供了重要的参照依据。热点区域由于其特定的地理位置、经济活动、港口设施等吸引因素，往往成为船舶航行的目的地或必经之地。在预测模型中，船舶向热点区域移动的概率显著高于其他区域。因此，将热点区域的吸引力作为船舶轨迹预测模型考虑的关键参数，以提高预测的准确性和可靠性。本章同时是本文第四章船舶轨迹预测的研究基础。本章从实验数据、轨迹预处理、轨迹网格化等内容来介绍热点区域的具体挖掘方法，并设计实验来验证相关算法的可行性和准确性。

3.1 总体思路（General Idea）

本文研究设计的热点区域挖掘方法，核心在于利用全球六边形离散网格对船舶的历史轨迹数据进行高效的空间化处理。该方法首先将连续的船舶轨迹信息映射至预先定义的六边形网格结构中，在此框架下，进一步对每个网格内的船舶姿态参数进行综合计算与分析，包括但不限于船舶的速度、航向以及停留时间等关键指标，最后基于密度与时间两个维度提取热点区域。密度维度反映了船舶在特定网格内的聚集程度，通过计算网格内船舶数量的空间分布密度，可以有效识别出船舶活动频繁的区域。同时，时间维度的考量则允许捕获船舶在这些热点区域内的活动持续时间，从而辨识出长期或周期性的船舶聚集现象。

热点区域提取的具体过程如下：在对海域的轨迹热点区域挖掘之前，首先对历史轨迹进行预处理，包括异常值剔除、轨迹压缩与轨迹插值，为后期的数据挖掘与数据分析提供良好的数据基础；完成轨迹数据预处理后，对每一条历史轨迹先按照大圆弧分段后进行网格化处理，实现轨迹的全球六边形网格表达；随后，计算并统计历史轨迹在每一个六边形网格的停留时间与经过频次，根据设定的阈值来判别网格是否属于热点网格；最后，合并相邻的热点网格为一个热点区域，同时计算热点区域的热度系数，评价与分析相关热点区域。图 3-1 给出了上述过程的基本思路：

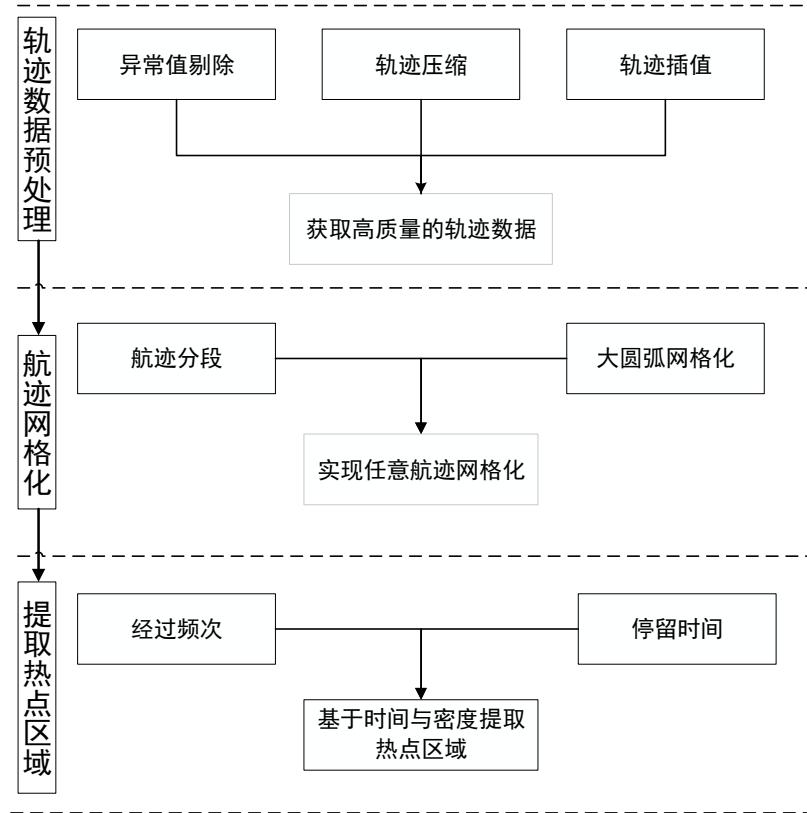


图 3-1 热点区域挖掘总体流程图

Figure 3-1 Overall flow chart of hotspot area mining

3.2 实验数据（Experimental Data）

3.2.1 AIS 数据介绍

最初开发的自动识别系统（AIS）技术主要旨在解决船舶间信息识别的问题，并辅助岸基船舶交通服务（VTS）操作员准确甄别雷达上的船舶目标。到了 20 世纪 90 年代末，国际海事组织（IMO）基于提高船舶航行援助能力和加强海上搜索救援系统的构建理念，确定将船载 AIS 设备作为船舶必须强制安装的设备之一，可适用于内河、港口及外海等各类船舶。整个系统由岸基站、助航站、转发站与船载设备等共同组成，具有船舶身份识别、数据传输、处理信号等功能^[85]。

AIS 数据是各类传感器收集并且通过 AIS 设备按照固定格式发送的船舶航行时的状态信息。根据国际 ITU_R 标准，AIS 数据主要分为静态信息、动态信息、航行相关信息、航行安全信息和安全短信，这些信息均由二进制编码的形式传输，使用时需要进一步解析。

(1) 静态信息。静态信息只有当船舶类型发生重大改变或其名称或呼号更改时才需要更改，一般不会随着船舶航行状态的改变而改变。主要包括军事移动服务标识码（Maritime Mobile Service Identity，MMSI）、船舶类型（客船、

油轮等）、船长、呼叫名称等。

(2) 动态信息。动态信息是随着船舶航行的状态变化而不断变化的信息。主要包括时间戳（UTC）、对地航速（Speed Over the Ground, SOG）、对地航向（Course Over the Ground, COG）、船首向（Heading）、导航状态等。

(3) 航行相关信息。航行相关信息主要包括船舶吃水深度、运载货物类型、目的地与预计到达时间以及路线规划等信息。

(4) 安全短信。一般指带有重要导航安全相关信息的短信，会显示在一个额外的窗口中。

3.2.2 实验区域

本文所有实验均采用欧洲波罗的海的船舶 AIS 数据进行一系列相关实验。波罗的海是位于北欧地区的一个辐射状海域，被认为是欧洲最大的内陆海，其被瑞典、芬兰、爱沙尼亚、拉脱维亚、立陶宛、俄罗斯等国环绕，是一个重要的地缘政治和地理区域。该区域从北纬 54 度起向东北延伸，到近北极圈的地方为止，呈三岔形，西以斯卡格拉克海峡、厄勒海峡、卡特加特海峡、大贝尔特海峡、小贝尔特海峡、里加海峡等海峡和北海以及大西洋相通。实验数据集包含 2017-2019 两年间波罗的海以 AIS 观测形式的船舶跟踪数据，共计 3507 条船舶轨迹，轨迹数据的分布如图 3-2 所示：



图 3-2 实验数据分布
Figure 3-2 Experimental data distribution

3.3 轨迹预处理 (Trajectory Preprocessing)

通过传感器和定位技术收集的 AIS 数据容易受到噪声、缺失和数据不一致的影响，影响数据准确度的原因主要来自两个方面：一是 AIS 本身设备问题，如信号接收能力弱或精度不足；二是数据采集和传输过程遭遇外部不可控因素的干扰，例如恶劣的天气状况和特殊的海域环境。低质量的数据往往将致使低质量的挖掘效果，使得从数据中发现的模式和规律不准确，因此进行有效的数据预处理工作是进行数据分析和数据挖掘工作的重要前提。本文主要从异常值剔除、轨迹压缩与轨迹插值对原始轨迹做预处理。

3.3.1 异常值剔除

异常数据的排除旨在消除非法数据与噪声对后续轨迹挖掘过程的干扰，其中无关数据涵盖船舶抵达港口的时间、气象状况以及船舶的建造日期等；而噪声数据则指的是包含非法值、重复项及缺失值等信息。一般认为，某个点位置偏离正常轨迹较远或者在一定时间范围内速度、方向出现较大偏差则认为该点为异常点，应进行剔除掉。具体清洗方法如下：

- (1) 删除时间相同且 MMSI 相同的数据。这是由于 AIS 传输信道拥挤，导致一条信息可能被重复接收，重复数据会影响到后期的工作量与精度；
- (2) 删除 $SOG > 50$ 或者 $SOG < 0$ 的轨迹数据，说明船舶速度数据不正常，属于噪声数据；
- (3) 删除位置偏离正常轨迹较远或者在一定时间范围内速度出现激变的情况，具体规则如下：

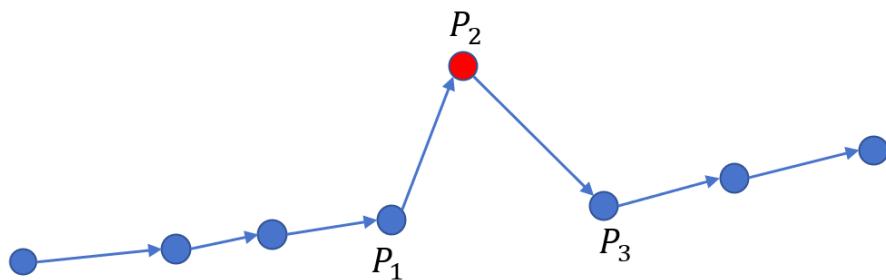


图 3-3 轨迹中的噪声点

Figure 3-3 Noise points in the trajectory

如上图所示，假设三个在时间上连续点分别为 $P_1(t_1, lon_1, lat_1, sog_1, cog_1)$ 、 $P_2(t_2, lon_2, lat_2, sog_2, cog_2)$ 与 $P_3(t_3, lon_3, lat_3, sog_3, cog_3)$ ，其中 t_i 、 lon_i 、 lat_i 、 sog_i 与 cog_i 分别表示该轨迹点的采样时间戳、地理经度、地理纬度、速度与方向。

a、航速异常与航向异常

一般船舶的航速与方向不会在短时间内发生剧烈的变化，在传感器采集和传输的过程中出现异常速度值与异常航向值，因此本文从航速变化率与航向变化率两个指标进行评判：

$$\alpha_s = \frac{|sog_2 - sog_1|}{|t_2 - t_1|} \quad (3-1)$$

$$\beta_c = \frac{|cog_2 - cog_1|}{|t_2 - t_1|} \quad (3-2)$$

上式中， α_s 与 β_c 分别为航速变化率与航向变化率，设定 α_s 的阈值为 α_{max} 与 β_{max} 这两个指标表示船舶运行状态的变化的快慢程度。当船舶轨迹点的 α_s 大于 α_{max} 或 β_c 大于 β_{max} 时说明该点是异常数据，需要进行剔除处理。

b、位置异常

首先是计算两个轨迹点之间的球面距离，通过两个点的经纬度来计算两个点之间的距离，公式如下：

$$d = 2r \sin^{-1} \left(\sqrt{\sin^2 \left(\frac{\varphi_2 - \varphi_1}{2} \right) + \cos(\varphi_1) \cos(\varphi_2) \sin^2 \left(\frac{\lambda_2 - \lambda_1}{2} \right)} \right) \quad (3-3)$$

其中， d 表示两个点之间的距离， r 表示地球的半径，本文地球半径取值为6371km， (λ_1, φ_1) 和 (λ_2, φ_2) 分别表示前后两个点的经纬度坐标。假设 P_1P_2 、 P_2P_3 与 P_1P_3 的距离分别为 d_{1-2} 、 d_{2-3} 与 d_{1-3} ，若 $d_{1-2} + d_{2-3} > 2d_{1-3}$ ，则说明点 P_2 位置异常，其已经严重偏离轨迹的前进方向，我们必须对其进行过滤。

3.3.2 轨迹压缩

轨迹由 AIS 坐标点的离散样本近似、采样率越高，轨迹将越精确，同时也产生了大量数据导致存储、传输和处理的开销问题。因此，设计数据压缩技术以压缩轨迹的大小的同时又不牺牲自动海事航路应用所需的轨迹质量至关重要。Douglas 和 Peucker 在 1973 年提出 Douglas-Peucker (DP) 算法^[86]，该算法以递归方式分割路线数据并通过阈值控制压缩质量，由于其速度快、精度高而被广泛用于简化移动点对象的运动轨迹。Meratnia 等人^[87]在 2004 年提出的基于速度和时间比例的自顶向下轨迹压缩算法 (Top-Down Time-Ratio Algorithm, TD-TR 算法)能够集中对 GPS 轨迹的形状进行压缩。Schmid 等人^[88]研发了一种结合路网信息的 GPS 轨迹压缩技术，该技术通过首先对移动用户的轨迹进行地图匹配，识别出用户所行驶的具体道路，然后以这些路段信息来代表压缩后的轨迹，极大提升原始轨迹数据的压缩效率。为了支持在线轨迹压缩，Meratnia^[89]提出了开放窗口压缩方法 (Opening Window, OW)，该算法实现了压缩率与简化误差之间的有效平衡。鉴于传统的 DP 算法被认为是一种标准的轨迹压缩技术，并且展现出了优秀的轨迹压缩性能，本文采用 DP 算法对原始轨迹点进行压缩。

以下对 DP 算法原理进行具体介绍:

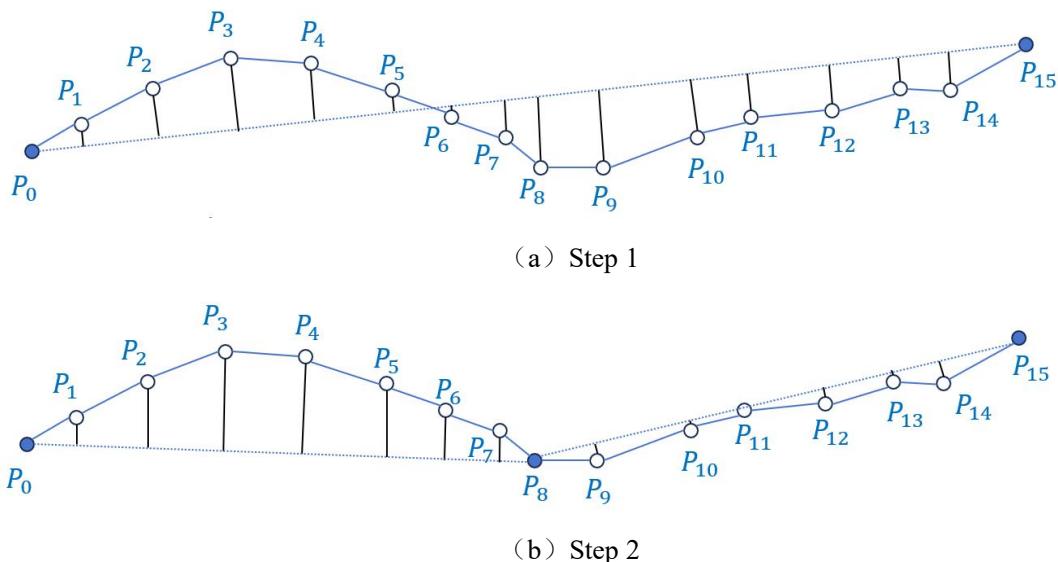


图 3-4 DP 算法轨迹压缩过程示意图

Figure 3-4 Schematic diagram of the trajectory compression process of DP algorithm

- (1) 首先在曲线首尾两点 P_0 与 P_{15} 之间连接一条直线 P_0P_{15} ，该直线为曲线的弦；
- (2) 找到曲线上与 P_0P_{15} 距离最远的点 C （点 P_8 ），并计算 C 与 AB 的距离 d ；
- (3) 将距离 d 与预先设定的阈值 threshold 进行比较，若 d 小于 threshold，则认定直线 P_0P_{15} 为该曲线段的近似表示，完成该段曲线的处理；
- (4) 若 d 大于 threshold，则用 C 将曲线分为两段 P_0C 和 CP_{15} ，并对这两部分分别重复步骤 (1) 至 (3) 的处理；
- (5) 所有曲线段处理结束后，将所有分割点按序连接起来，形成的折线即为原曲线的近似表示。

为了验证 DP 算法应用在本文航迹压缩中的有效性，实验采用欧洲波罗的海的 2017-2019 两年间的船舶 AIS 数据进行实验，进行大量单条船舶轨迹的压缩实验。下图为某一条轨迹数据的压缩展示：

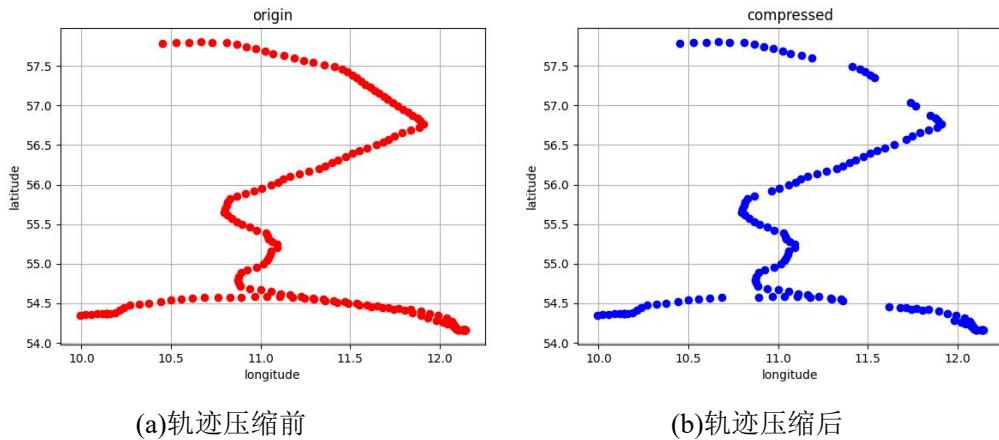


图 3-5 DP 算法轨迹压缩结果

Figure 3-5 DP algorithm trajectory compression result

从图 3-5 可知：压缩后的轨迹基本保持原航线的最基本的形状特征特点，重要的特征点均被保留下，在轨迹直行处压缩率较大，而在轨迹弯曲处部分压缩精度比较低，采用更少的数据量来描述原轨迹特征，提升了处理海量轨迹数据的效率，为后续高效的轨迹处理与数据挖掘奠定基础。

3.3.3 轨迹插值

原始轨迹数据可能会出现记录采集时间间隔较长的情况，这就导致了部分数据处于缺失的状态。在连续的一段时间内轨迹数据处于异常状态，通过异常值剔除之后导致一段时间数据的空缺，这种仍然需要做缺失值处理。以某一条航迹数据为例，其轨迹分布如图 3-6 所示，航线在点 $P1$ 与 $P2$ 的时间间隔超过了最大阈值 T_{max} （本文 T_{max} 设置为 30min），出现了部分数据缺失的情况。针对该问题，需要使用合理的数据插值算法对缺失的数据进行插补处理。



图 3-6 轨迹数据缺失示意图

Figure 3-6 Schematic diagram of missing trajectory data

对于数据缺失的情况，一般采用线性插值、拉格朗日插值与三次样条插值等方法，本文实验采用三次样条插值处理，以下对三次样条插值做详细介绍：

三次样条插值是一种常用的插值方法，它通过在给定的数据点上构造一组三次多项式，以实现平滑的插值效果。与线性插值、拉格朗日插值等方法不同，三次样条插值将插值区间划分为多个小段，并在每个小段上构造一个三次多项

式。这使得整个插值函数在给定数据点上具有较好的平滑性和连续性，避免了拉格朗日插值的曲线波动与极值点过多等缺陷，能够较好地逼近原函数的形状。相对于高次插值方法而言，它所需的计算量较小。三次样条插值的具体步骤如下：

(1) 分割插值区间。首先，将插值区间 $[a, b]$ 分割为 n 个小段，每一个小段对应2个数据点，即 $a = x_0 < x_1 < \dots < x_n = b$ ，假设这些点对应的函数值为 y_0, y_1, \dots, y_n 。

(2) 构造三次多项式。在每一个小段 $[x_i, x_{i+1}]$ 上构造一个三次多项式，其形式如下：

$$S_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3 \quad (3-4)$$

(3) 确定系数。为了确定三次多项式的系数，需要满足以下条件：

- 插值条件： $S_i(x_i) = y_i$ 与 $S_{i+1}(x_{i+1}) = y_{i+1}$
- 平滑条件： $S_i(x_{i+1}) = S_{i+1}(x_{i+1})$
- 曲率条件： $S_i''(x_{i+1}) = S_{i+1}''(x_{i+1})$

这些条件将构成一个线性方程组，通过求解这个方程组可以得到每个小段上的系数。

(4) 插值计算。通过求解线性方程组得到每个小段上的系数后，即可得到整个插值函数。在给定的插值区间内，根据数据点和对应的系数，可以计算出任意点的插值结果。

3.4 航迹网格化 (Trajectory Gridding)

矢量与栅格是空间数据模型中最基本的两种类型，在不同的应用场景中各具优势。矢量数据常用于城市规划、土地管理、地图制作、空间分析和导航系统等需要精确几何信息和拓扑关系的领域。而栅格数据适用于地形建模、遥感图像处理、环境模拟与自然资源管理等需要连续性数据表示和分析的领域。本文需要模拟船舶轨迹的连续动态规律与挖掘相关轨迹模型，因此研究对象由轨迹点矢量数据转换为网格，这对于后续的热点挖掘、轨迹预测是一种更优的做法。

3.4.1 网格化算法

依据一定准则，将矢量数据映射至对应尺度格网上的过程称为离散化（网格化），这是矢量数据与栅格数据同构处理的重要环节。计算机图形学及地理信息系统领域，已研发多种离散线生成算法，具有代表性的有数字微分法^[90]、中点画线法^[91]和 Bresenham 算法^[92]。在此基础上，Freeman^[93]提出了基于顶点的三角形格网离散线生成算法，该算法根据单元顶点到矢量线的距离筛选单元，

导致部分离散线单元偏离矢量线。刘勇奎等^[94]设计了一种基于平面的矢量线六边形格网表达成算法，该算法通过逐点选择距离直线最近格网点的方式计算下一个网格化单元。针对三角形格网离散线生成算法的不足，杜灵瑀等^[95]提出了基于弱对偶的平面三角形格网离散线转化生成算法，借助于弱对偶六边形格网，建立等效三角形格网离散线数学模型，并通过降维的方式实现网格离散线的生成。目前的离散线生成算法主要是基于局部的平面空间生成离散线，但在球面应用网格化算法面临巨大的挑战。球面上处理几何问题需要考虑到曲面特性、距离计算、曲面投影、曲面交叉等困难。针对上述问题，一般有两种方案：一是直接在球面上网格化，按照球面矢量与球面单元的网格单元的几何关系生成网格；二是先把球面上的点投影到平面上，然后在平面空间在应用算法生成离散线，最后反投影到球面上。于文率等^[96]通过在全球六边形离散网格系统中对矢量线的精确表达和误差管理进行精细控制，确保了格网的误差被严格限制在同一层级单个网格单元的范围之内，有效地完成了矢量线的六边形全球网格化处理。文献^[97]提出了一种基于编码运算的矢量线球面六边形网格化方案，将平面网格中的基于向量平移建立矢量线格网表达模型的思想扩展到球面。该方法根据方位角之差选取三个最优向量，这些向量在每一步平移中都会得到三个相应单元，其中网格中心到矢量线的球面最小距离的单元作为后续网格化单元。尽管目前的球面六边形网格化技术已经取得了一些进展，但依然面临着几个主要问题，包括网格化误差大、计算过程繁琐、算法理论模型复杂以及时间性能差等挑战。

鉴于现有方法的局限性和项目需求，本文设计了一种直接基于球面的线段网格化算法，可实现对任意弧段的快速网格化。任意航迹由若干航段构成。为此，航迹网格化的核心在于航段的网格化。视任意航段为两个节点间的大圆弧，航迹网格化的本质是大圆弧的网格化，如下图所示：

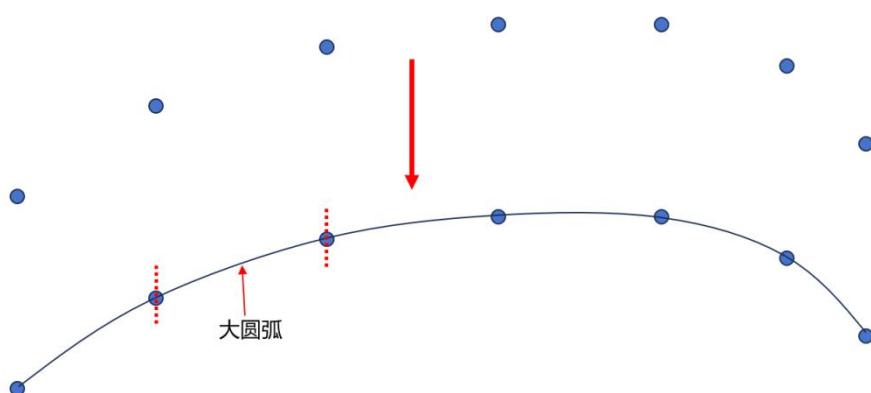


图 3-7 原始航迹由若干大圆弧段组成
Figure 3-7 The original track consists of several large arc segments

对原始轨迹进行分段之后，需对各个分段的弧线进行网格化处理，其核心思想是通过球面几何关系计算出所有与大圆弧相交的网格单元。首先从起点网格开始，下一个网格为当前所处网格的六个相邻网格之一，利用两个六边形共用一条边的拓扑关系计算下一个网格。按照这种方式循环计算下一个网格，直到网格化到终点为止。大圆弧网格化过程如图 3-8 所示：

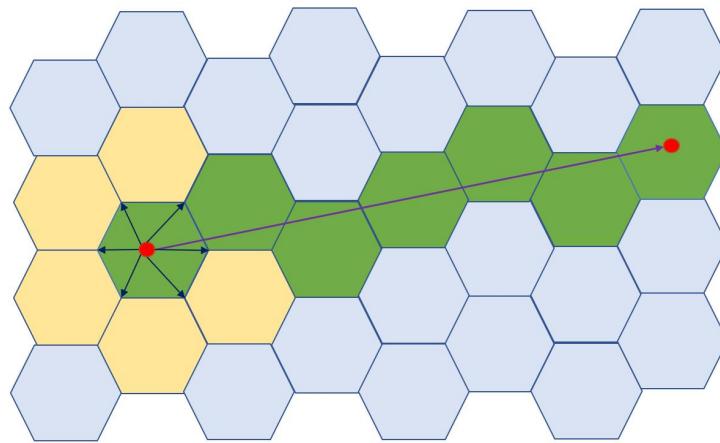


图 3-8 大圆弧网格化示意图

Figure 3-8 Schematic diagram of the large arc meshing algorithm

以下详述大圆弧网格化算法：假设给定大圆弧 AB ，即为大圆弧 AB 的劣弧部分（方向是从 A 到 B ），设 H_1 、 H_2 、 H_3 、 H_4 、 H_5 与 H_6 为网格的 6 个顶点，大圆弧从 A 点开始，到 B 点结束，如图 3-9 所示：

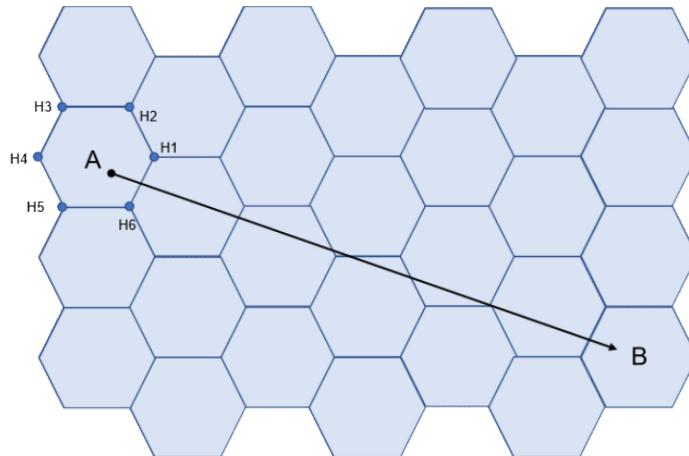


图 3-9 轨迹线 AB 采用大圆弧连接示意图

Figure 3-9 Schematic diagram of trajectory line AB connected by a large arc

(1) 计算 A 点与 B 点所处网格的编码，分别记为 ID_A 与 ID_B 。若 $ID_A = ID_B$ ，说明 A 与 B 点同处一个网格，立即结束大圆弧的网格化。否则，进入第(2)步；

(2) 通过编码加法运算计算当前网格的 6 个相邻网格，同时采用大圆弧连接 AH_1 、 AH_2 、 AH_3 、 AH_4 、 AH_5 与 AH_6 ，计算圆弧 AH_1 、 AH_2 、 AH_3 、 AH_4 、 AH_5 、 AH_6 与 AB 的大地方位角，计算大地方位角的公式如下：

$$\alpha = \arctan \frac{\cos B \sin(L - L_0)}{\cos B_0 \sin B - \sin B_0 \cos B \cos(L - L_0)} \quad (3-5)$$

上式中, (B_0, L_0) 与 (B, L) 分别对应两个大地点的经纬度。

(3) 计算圆弧AH₁、AH₂、AH₃、AH₄、AH₅、AH₆与AB的大地方位角之差, 分别为 ω_1 、 ω_2 、 ω_3 、 ω_4 、 ω_5 、 ω_6 , 大小在 $[-180^\circ, 180^\circ]$ 范围内, 如下图所示:

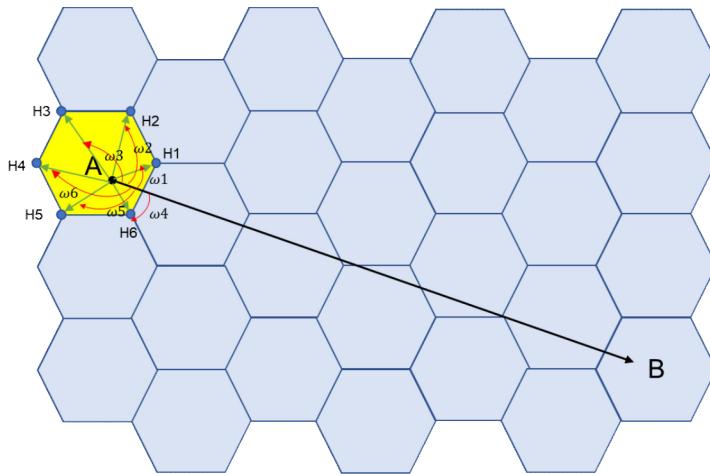


图 3-10 计算 6 个大地方位角夹角示意图

Figure 3-10 Schematic diagram of calculating the included angles of six geodetic azimuth angles

(4) 在(3)的基础上, 判断相邻夹角乘积结果符号的正负性, 若乘积结果为负, 说明该两夹角共用的六边形边为活动轨迹的离开边, 除了当前网格的进入边对应的两个夹角外, 有且仅有一对相邻夹角的乘积结果为负。如图 3-10 所示, $\omega_1 \times \omega_4 < 0$, 则活动轨迹从边 H_1H_6 进入到下一个网格;

(5) 在计算出活动轨迹的离开边后(为下一个网格的进入边), 查找与当前网格共用该边的相邻网格, 该相邻网格即为轨迹网格化的下一个目标网格;

(6) 通过上述步骤, 可能得到三个连续的互相相邻的网格, 如图 3-11 所示, 网格 G_1 、 G_2 、 G_3 两两之间互相相邻, 因此需要从 G_2 、 G_3 中选取一个离矢量线 AB 较近的网格作为前进网格。直接求解格心到矢量线的球面距离过程较为复杂, 本文利用单元格心到矢量线的端点与球心所确定的平面的垂直距离代替格心到矢量线的球面距离;

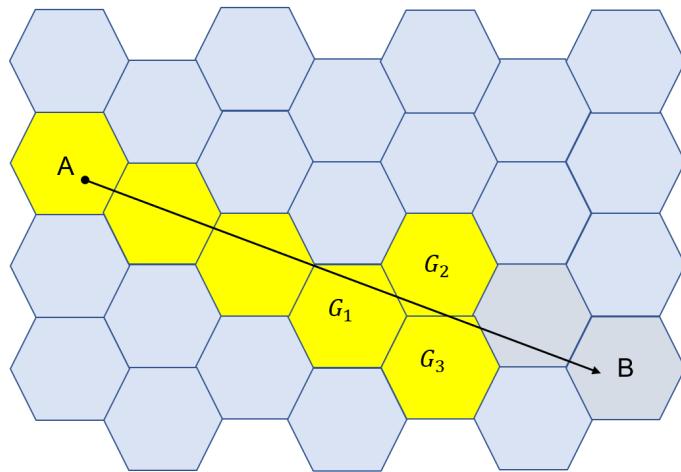


图 3-11 三个连续互相相邻的网格

Figure 3-11 Three consecutive grids adjacent to each other

(7) 进入到下一个目标网格后, 重复(2)到(6)的五个步骤进行计算后续的目标网格, 当目标网格为点B所处在的终点网格时立即停止循环, 过程如图 3-12 所示:

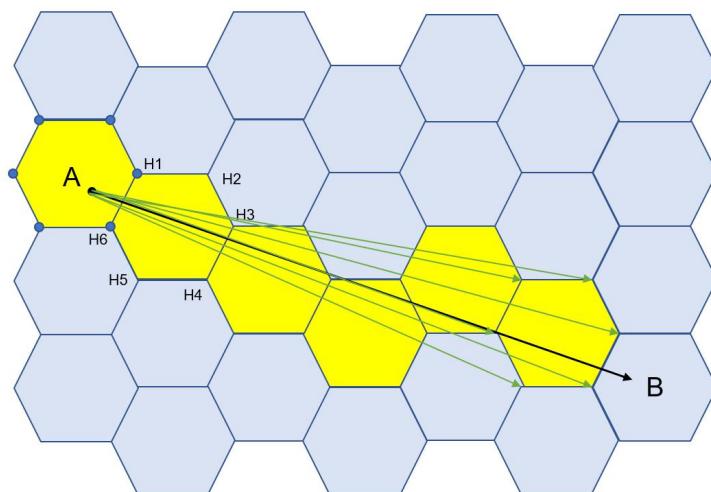


图 3-12 后续网格计算示意图

Figure 3-12 Schematic diagram of subsequent grid calculation

由上所述, 大圆弧网格化算法的伪代码如表 3-1 所示:

表 3-1 大圆弧网格化伪代码

Table 3-1 Pseudo code for large arc gridding

Algorithm 大圆弧网格化

输入变量: 起点A与终点B的经纬度: (B_s, L_s) 与 (B_e, L_e)

输出变量: 网格化编码结果CODEs

1: Begin

2: 将起点与终点经纬度分别转换为网格编码 s_{code} 与 e_{code}

表 3-1 (续)

Algorithm 大圆弧网格化

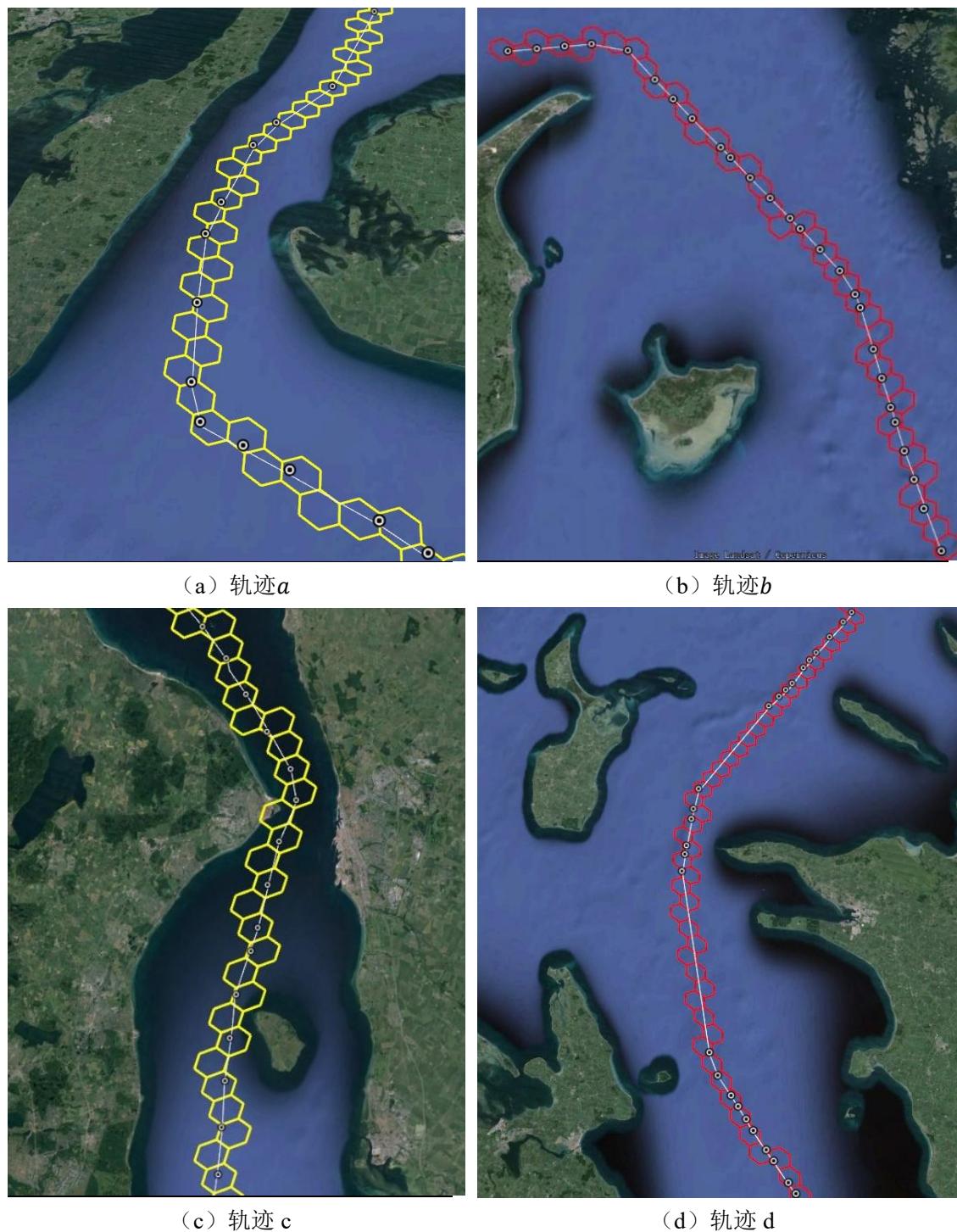
```

3:           If    $s_{code} = e_{code}$ 
4:               Return  CODEs
5:           End If
6:           While 当前网格编码  $cur_{code} \neq e_{code}$ 
7:               计算当前网格的 6 个相邻网格  $NeighborGrid_s$ 
8:               记录当前网格的两个顶点分别为  $H_1$ 、 $H_2$ 、 $H_3$ 、 $H_4$ 、 $H_5$ 、 $H_6$ 
9:               计算圆弧  $AH_1$ 、 $AH_2$ 、 $AH_3$ 、 $AH_4$ 、 $AH_5$ 、 $AH_6$  的大地方位角
10:              计算圆弧 AB (劣弧) 的大地方位角
11:              计算圆弧  $AH_1$ 、 $AH_2$ 、 $AH_3$ 、 $AH_4$ 、 $AH_5$ 、 $AH_6$  与圆弧 AB 的大地
12:                 方位角之差:  $\omega_1$ 、 $\omega_2$ 、 $\omega_3$ 、 $\omega_4$ 、 $\omega_5$ 、 $\omega_6$ , 并存入  $Ws$ 
13:                 For  $Ws$  中的每一个  $\omega_i$ 
14:                     If    $\omega_i \times \omega_{i+1} < 0$ 
15:                         查询活动轨迹的离开边
16:                         查找与当前网格共用离开边的相邻网格  $next_{code}$ 
17:                         Break
18:                     Else Continue
19:                     End If
20:                 End for
21:                  $cur_{code} = next_{code}$ 
22:                 // 判断是否具有三个连续的互相相邻的网格
23:                 If CODEs.size > 2
24:                     If   当前  $cur_{code}$  与网格  $CODEs[CODEs.size-2]$  相邻
25:                         计算当前网格中心到平面 ABO 的距离  $d_1$  与网格
26:                            $CODEs[CODEs.size-1]$  中心到平面 ABO 的距离  $d_2$ 
27:                         If  $d_1 < d_2$ 
28:                             CODEs 弹出  $CODEs[CODEs.size-1]$ , 同时装入  $cur_{code}$ 
29:                         End If
30:                     End If
31:                 End If
32:             End While
33:             Return  CODEs
34:         End

```

3.4.2 实验结果与分析

为验证及评估本文提出的航迹网格化算法的准确性与时效性，本文以波罗的海 AIS 数据为例进行实验，图 3-13 给出了 6 条不同轨迹的网格化结果，其中白色的线为球面矢量线。实验采用 C++ 语言实现，并都在同 1 台联想台式电脑（硬件配置：i5-7500；CPU@3.40GHz；16GBRAM；操作系统：Windows10 专业版；开发工具：Visual Studio 2019）上运行。



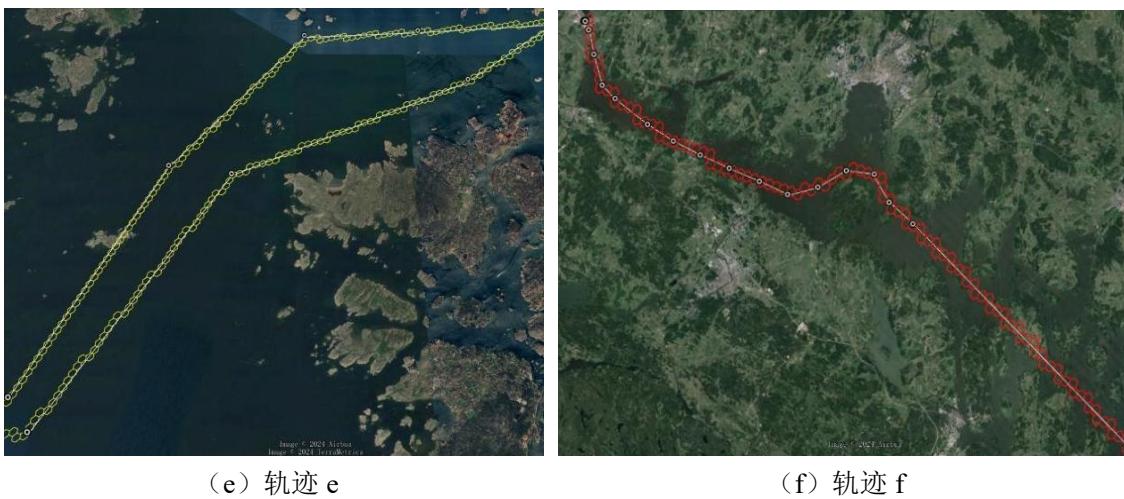


图 3-13 网格化结果

Figure 3-13 Gridding results

由实验结果可知：从整体效果来看，网格化单元均以边相邻的形式连接，具有良好的平滑性与连续性，能够较为准确拟合矢量数据。在不同的层级下，网格化结果误差均小于 1 个网格单元，每一条轨迹均有六边形网格准确覆盖，每一个网格均与原始矢量线相交，实现了活动轨迹的接近无误差网格化。因此，本文网格化算法具有一定的准确性，能够应用于基于球面的航迹网格化，可满足后期轨迹挖掘与轨迹预测的相关需求。

同时，为了评估算法的时效性，针对不同的网格层级数，统计了船舶轨迹在不同网格大小下的生成时间，统计的相关指标包括：最大生成时间、最小生成时间与平均生成时间，具体的统计结果如表 3-2 所示。结果表明：在一定层级范围内，该算法能够快速生成长距离航迹的六边形网格，平均每条轨迹的网格生成时间均小于 50ms。此外，为进一步分析生成时间随轨迹长度的变化情况，以 $Level = 12$ 为例绘制了轨迹长度与网格生成时间的散点图，如图 3-14 所示。结果表明：随着网格半径的增大，网格生成时间有所增加，时间与轨迹长度呈现出正相关（线性）的趋势。基于上述分析可知，在一定的层级范围内，本文航迹网格化算法能够快速实现轨迹的网格化，满足于后期轨迹预测的实时要求。

表 3-2 不同层级下网格生成时间

Table 3-2 Grid generation time at different levels

层级	网格平均大	轨迹条	轨迹平均长	最大生成	最小生成	平均生成
	小(km)	数	度(km)	时间(ms)	时间(ms)	时间(ms)
10	4.2	2159	904	29	1	5.76
11	2.3	1737	911	35	1	9.68
12	1.2	2115	920	98	2	19.04
13	0.6	1987	915	230	3	40.55

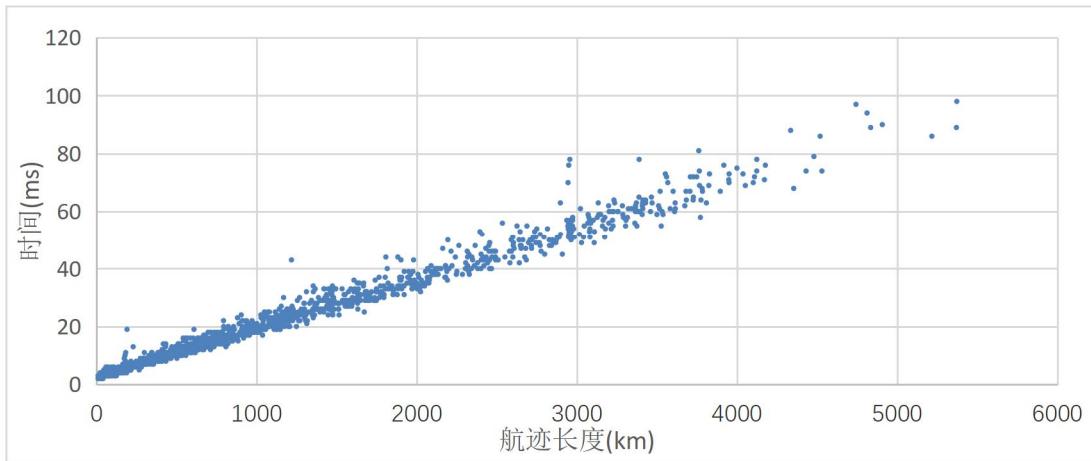


图 3-14 level=12 时网格生成时间
Figure 3-14 Grid generation time when level=12

3.5 热点区域提取 (Hot Spot Area Extraction)

GScan 算法^[41]是一种基于网格密度的热点区域探测方法，面对大规模轨迹数据时，其表现出伸缩性好、计算效率快等优点，被广泛运用于城市中的出租车热点区域挖掘、航迹热点区域挖掘与人行热点区域挖掘等研究领域。其基本思路为：首先，将研究区域被划分为大小一致、不相交的正方形网格单元；然后原始空间数据通过映射函数分配到相应的网格单元中；接着，通过计算各网格单元内的数据密度，根据阈值识别出热点网格单元；最终，根据热点网格单元内数据点的分布确定热点位置并合并相邻的热点网格为热点区域。过程如图 3-15 所示：

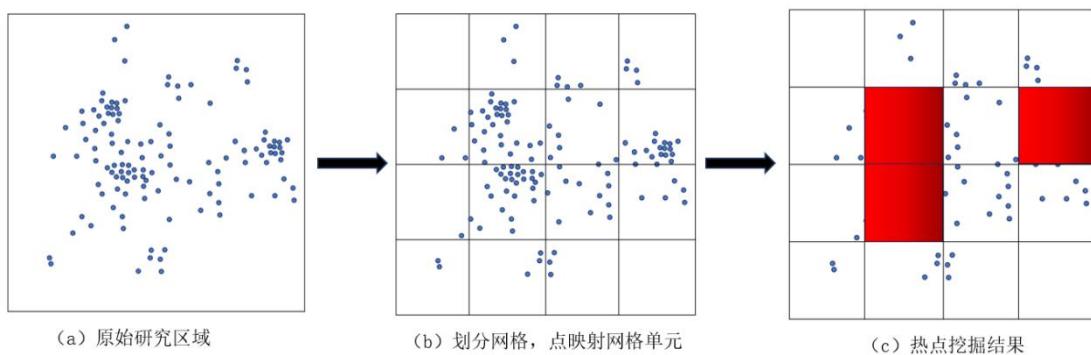


图 3-15 GScan 热点挖掘示意图

Figure 3-15 Schematic diagram of GScan hotspot mining

受到 GScan、CLIQUE 等热点区域挖掘算法的启发，本文设计了一种基于全球六边形网格的热点区域挖掘方法。本文把热点区域定义为船舶频繁经过且停留时间超过一定时长的区域。基于前文提出的航迹网格化算法对船舶的每一条历史轨迹进行网格化，借助于轨迹点的经纬度、速度等信息计算与统计出历史轨迹在每一个网格的经过频次与停留时间，根据阈值筛选出相应的热点网格。

因此，与传统的许多热点区域提取方法有所差异，本文方法充分考虑了时间维度对热点提取结果的影响，为热点区域的精细化挖掘研究提供了参考方案。

3.5.1 算法思想

通过轨迹网格化，已经实现了将离散的点轨迹对象转换为连续的六边形网格。在此基础上，首先借助于轨迹点的经纬度、速度、时间戳等信息，计算历史轨迹在每一个网格的经过频次与停留时间信息；其次，统计计算历史轨迹在每一个网格上的平均经过频次与平均停留时间；接着，随后根据阈值判断热点网格；随后，通过合并相邻的热点网格从而得到热点区域；最后，通过计算热度系数来分析与评价相关热点区域。具体步骤如下：

(1) 根据原始轨迹点的经纬度与各六边形网格的顶点坐标，计算出两个轨迹点之间的大圆弧段与各六边形网格的交点，如下图所示：

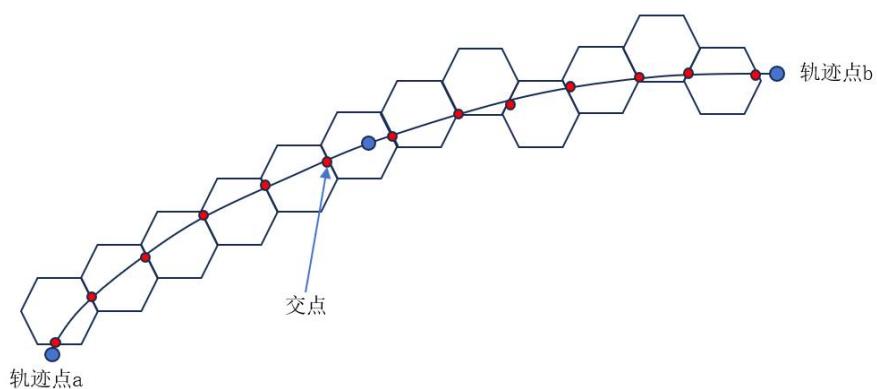


图 3-16 两个轨迹点与各六边形网格相交示意图

Figure 3-16 Schematic diagram of two trajectory points intersecting each hexagonal grid

(2) 求出交点后，根据两个轨迹点的时间戳信息，按照距离加权平均计算出每一个交点的时间戳，由此可以统计每一条轨迹在单一网格上的停留时间，同时统计历史轨迹经过每一个网格的频次次数；

(3) 在(2)的计算基础上，计算出历史轨迹在每一个轨迹网格的经过频次 N 与平均停留时间 T ；

(4) 计算历史轨迹在所有网格的平均频次 ΔN 与平均停留时间 ΔT 。若满足以下条件，则定义该网格为热点网格；

$$\begin{cases} N \geq w_1 \Delta N \\ T \geq w_2 \Delta T \end{cases} \quad (3-6)$$

其中 w_1 、 w_2 中为常数， $w_1 \geq 1$ ， $w_2 \geq 1$ 。

(5) 假设提取的热点网格为 H_1 、 H_2 、 H_3 …… H_n ，从第一个网格开始遍历，判断其余的热点网格与其是否相邻，如果相邻，需要合并为大的热点区域；

(6) 最后对提取的热点区域进行分析与评价，计算每一个热点区域的热度系数（热度系数是针对各热点区域之间相比较而言，如果其热点系数更大，说

明该区域的热度更高) 计算公式如下:

$$HotCoe_i = w_1 * \frac{\Delta N_i}{\sum_{j=0}^n N_j} + w_2 * \frac{\Delta T_i}{\sum_{j=0}^n T_j} \quad (3-7)$$

其中, w_1 与 w_2 为经过频次与停留时间对计算热点区域的热度系数权重, 需满足: $w_1 + w_2 = 1$, $0 < w_1 < 1$, $0 < w_2 < 1$ 。如果值越大, 说明该因素对提取热点区域的影响越大。

上述为热点区域挖掘的过程, 具体的热点区域挖掘方法的伪代码如下表:

表 3-3 基于全球六边形网格的热点区域伪代码

Table 3-3 Pseudo code for hotspot areas based on global hexagonal grid

Algorithm 基于全球六边形网格的热点区域挖掘

输入变量: 区域历史轨迹信息

输出变量: 热点区域与热度系数表

```

1:      Begin
2:          对所有历史轨迹网格化
3:          计算每一条历史轨迹与六边形网格的交点
4:          计算与统计历史轨迹在每一个网格的停留时间与经过频次
5:          计算历史轨迹在所有网格的平均频次 $\Delta N$ 与平均停留时间 $\Delta T$ 
6:          根据阈值 $w_1\Delta N$ 与 $w_2\Delta T$ 筛选出相应的热点网格, 放入动态数组 $V$ 
7:          For 编码集合 $V$ 中的每一个热点网格 $g$ 
8:              If  $g$ 被遍历过 then 继续循环
9:              Else
10:                 查找 $g$ 网格的相邻网格
11:                 将相邻网格放入动态数组 $h$ 中
12:                 If  $h$ 中的网格集合无相邻网格
13:                      $h$ 存入 $H$ 集合中
14:                 Else 继续遍历 $h$ 中每一个网格的相邻网格, 结果存入 $H$ 
15:                 End If
16:             End If
17:         End For
18:         For 热点区域集合 $H$ 
19:             设定 $w_1$ 与 $w_2$ 的值
20:             根据热点区域的经过频次与停留时间计算计算热度系数
21:         End For
22:     End

```

3.5.2 实验结果与分析

在本文的热点区域挖掘算法中涉及到 w_1 与 w_2 两个重要参数，它们用来确定经过频次与停留时间的阈值，进而判断网格是否为热点网格。当 w_1 一定时，如果 w_2 越大，那么挖掘的热点区域数目越少；同理，当 w_2 一定时，如果 w_1 越大，挖掘的热点区域数目也越少。因而， w_1 与 w_2 值的大小直接影响到挖掘结果的精度与数目。文献中^[23]提出了一种基于信息熵的启发式参数设置方法，该方法能够在没有人工干预的情况下自动调整和优化算法的参数。信息熵公式如下：

$$H(X) = \sum_{i=1}^n p(x_i) \log p(x_i) \quad (3-8)$$

其中 $p(x_i)$ 表示随机事件 x_i 发生的概率，本文采用这种方法对算法中的参数 w_1 与 w_2 进行了设置，通过实验计算，本文设置 $w_1 = 2.0$ 与 $w_2 = 1.5$ 较为合适。

为了验证及评估本文热点区域挖掘方法的有效性与准确性，同样以波罗的海的 AIS 数据为研究对象进行实验，实验结果如图 3-17 所示，其中红色网格代表热点区域。

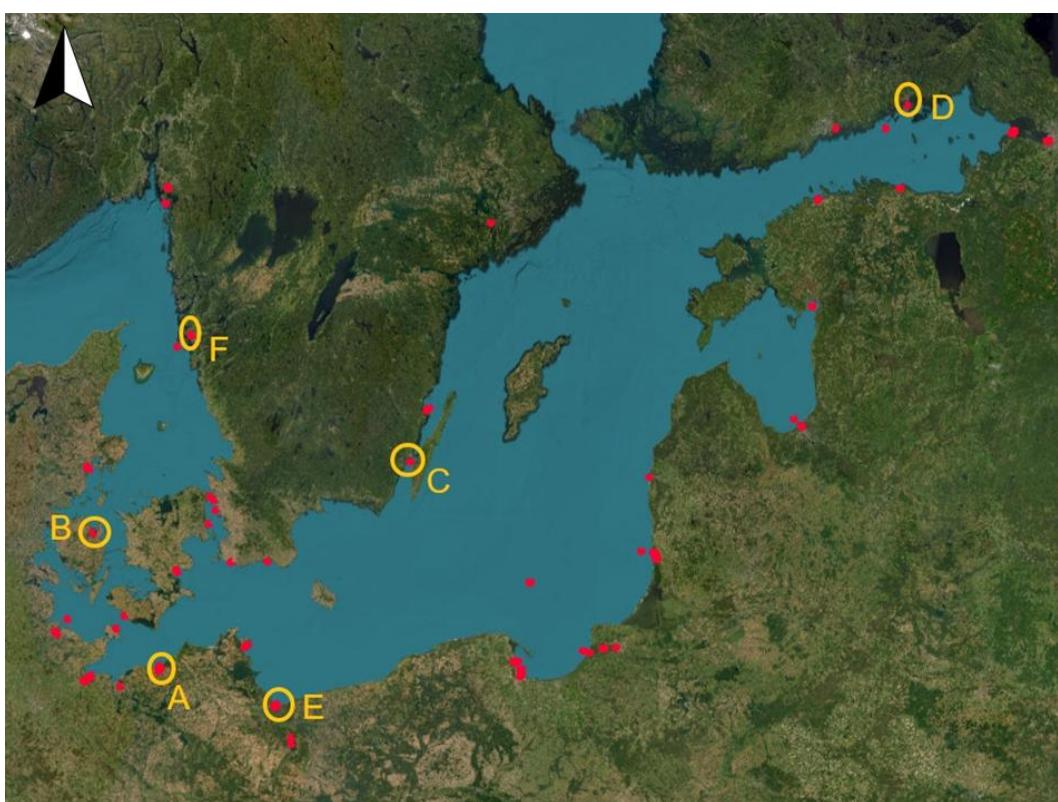


图 3-17 波罗的海热点区域提取结果

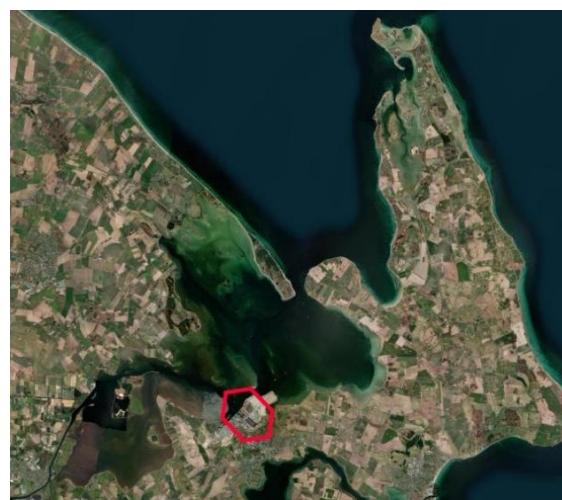
Figure 3-17 Baltic Sea hotspot area extraction results

为进一步评估提取结果的准确性，在地图软件上进行局部放大分析部分热点区域，对图中的 A、B、C、D、E 与 F 区域进行局部放大，如图 3-18 所示。同时，表 3-4 统计了不同热点区域类型的数量。实验结果表明：提取结果绝大

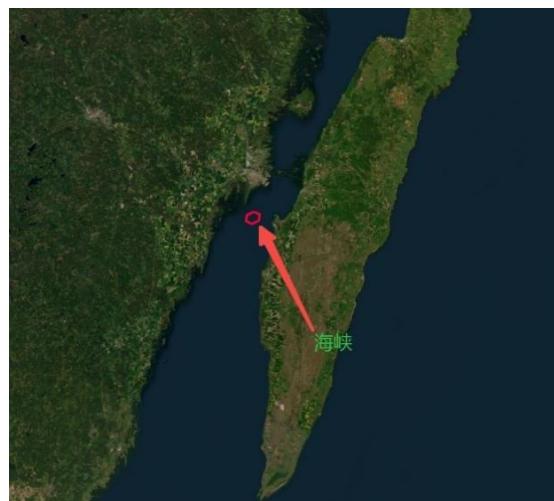
多数为港口，少部分为海峡、三角洲、航道交汇处与交通枢纽，极少部分为其他区域。一般来说，重要的港口、船舶行驶必经的海峡、连接内陆与外海的交通枢纽区域等区域都是船舶密集度高、活动频繁的区域，具有重要的航运功能。这说明提取结果符合从空间与时间两个维度限制下的热点区域。因此，从热点区域挖掘的结果结合实际情况分析来看，证明了提取的结果符合“热点区域”这一事实，同时也说明了本文方法具有一定的可行性与准确性，能够实现对船舶活动高频区域的精确识别。



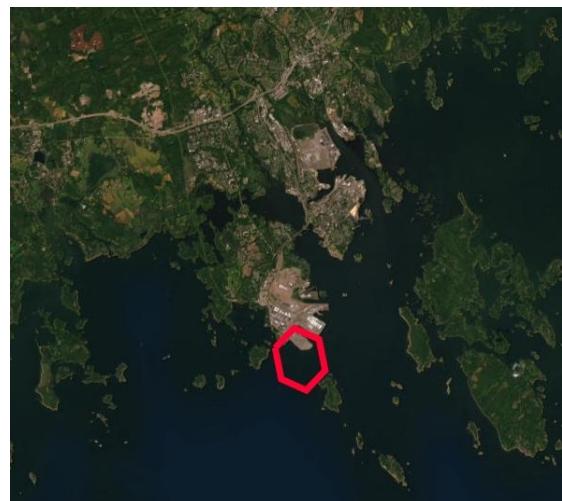
热点区域 A



热点区域 B



热点区域 C



热点区域 D



热点区域 E

热点区域 F

图 3-18 热点区域局部放大图

Figure 3-18 Partial enlargement of hotspot area

表 3-4 各热点区域类型数量统计表

Table 3-4 Statistics of the number of hotspot area types

]热点区域类型	数量	占比
港口	35	67.31%
海峡	4	7.69%
三角洲	5	9.62%
航道交汇处	3	5.77%
交通枢纽	2	3.85%
其他	3	5.77%

为了对比分析各热点区域之间的相对热度大小，计算每一个热点区域的热度系数，如表 3-5 所示，由于挖掘的热点区域结果较多，表中只是展示了部分热度系数较大的区域。其中，热度系数越高说明该区域相对热度越大。同时，热度系数的计算结果也是作为第四章船舶轨迹预测的研究基础。

表 3-5 区域挖掘结果热度系数表

Table 3-5 Heat coefficient table of regional mining results

热点区域编码	频次	停留总时间(s)	热度系数
40906	124	30097	0.009074
40917	91	2561	0.002895
33718258	60	1523	0.001877
33734653	149	86516	0.020611
33755026	208	76240	0.020187
33763219	66	2569	0.002237

表 3-5 (续)

热点区域编码	频次	停留总时间(s)	热度系数
33886084	42	22546	0.005455
33996633	42	2933	0.001674
34090968	114	1039	0.003208
52998140	813	112310	0.043104
67124624	715	15211	0.021798
34107353	210	8083	0.007099

3.6 本章小结 (Summary of this Chapter)

本章设计了一种基于全球六边形网格的热点区域挖掘方法，能够有效挖掘出船舶历史轨迹的热点区域。首先是介绍了本文热点区域挖掘方法的基本思路；然后，介绍了实验数据与数据预处理，数据预处理包括异常值剔除、轨迹压缩与轨迹插值；之后，设计了一种直接基于球面的线段网格化算法，实验表明：网格化算法具有一定的准确性与时效性，实现对任意弧段的快速网格化，且网格化误差小于一个网格单元；最后，阐述了热点区域提取方案的具体步骤并设计实验进行验证。结果表明：热点提取结果绝大多数为船舶密集度高、活动频繁的区域，符合从空间与时间两个维度限制下的热点区域，证实了挖掘方法具有一定的可行性与较高的准确性。

4 基于全球六边形网格的船舶航迹预测

4 Technology Ship Trajectory Prediction Based on Global Hexagonal Grid

在轨迹网格化与历史轨迹热点区域挖掘的研究基础上，本章提出了一种以目的地为导向的船舶航迹预测方法。首先，介绍了航迹预测的基本思想；之后，从目标区域预测、路径预测等内容介绍航迹预测技术；最后，设计了可行性分析实验与预测精度分析实验，分别从定性与定量两个方面对本文预测方法进行验证评估。

4.1 总体思路（General Idea）

当前对于船舶航迹预测的研究主要聚焦于两类方法：一是依托于船舶现有航迹信息的预测算法，通过分析船只当前的运动状态来预测其未来位置；二是基于历史 AIS 数据的预测方法，通过对船只当前航迹与过去记录进行比较匹配，以此预测船舶的未来运动轨迹。然而，仅依赖当前与历史轨迹信息来预测船舶的未来路径是不够充分的。决定船舶活动未来方向的因素极为复杂且多变，包括但不限于气候条件、航行环境、预定航向目的地等，这些因素在不同程度上对船舶的航向产生影响。通常，短期航迹预测关注于预测船舶在短时间内的具体行程，而从长期视角来看，预测的焦点转向船舶最终将到达的目的地。这些目的地通常对应于船舶历史轨迹中的热点区域，即船舶频繁访问的区域。换言之，航行预测的目的最终是为预测其奔赴哪个目的地服务的。目前的方法主要是基于历史轨迹和当前轨迹建立数学模型逐步预测，其预测结果最终可能奔向目的地，也可能奔向其他地方，这就忽略了目的地对船舶活动轨迹的影响。因而，当历史轨迹不足或者长远距离预测时，预测面临着更大的预测误差和不确定性。随着已知轨迹信息减少，我们对于未来航行路径的预测就变得更加困难。这是因为我们无法准确捕捉到船舶航行行为的变化模式，也无法充分了解可能的外部影响因素。对比，需要更加精细的建模方法和更多的考虑因素来提高预测的准确性和可靠性。

综上，本文提出一种基于全球六边形网格的和以目的地为导向的船舶航迹预测方法，预测时顾及到目的地及航道权等因素的多重约束。预测方法的具体过程如下：首先，确定船舶当前可能奔赴的目标区域，之后再以此为导向进行路径的预测。其基本思路为：首先，从海量历史航迹中挖掘出热点区域，从各个热点区域中筛选出船舶可能奔向的目标区域；之后，根据当前轨迹的航向、位置、相似历史轨迹以及各热点区域的热度系数构建计算模型来预知奔向各个

目的地的概率；最后，根据当前位置到目的地的相似历史轨迹情况进行路径预测，若当前位置到目标区域的历史轨迹多则以历史轨迹为参考进行路径预测，否则以 A*寻址算法进行路径的预测。其基本流程如图 4-1 所示：

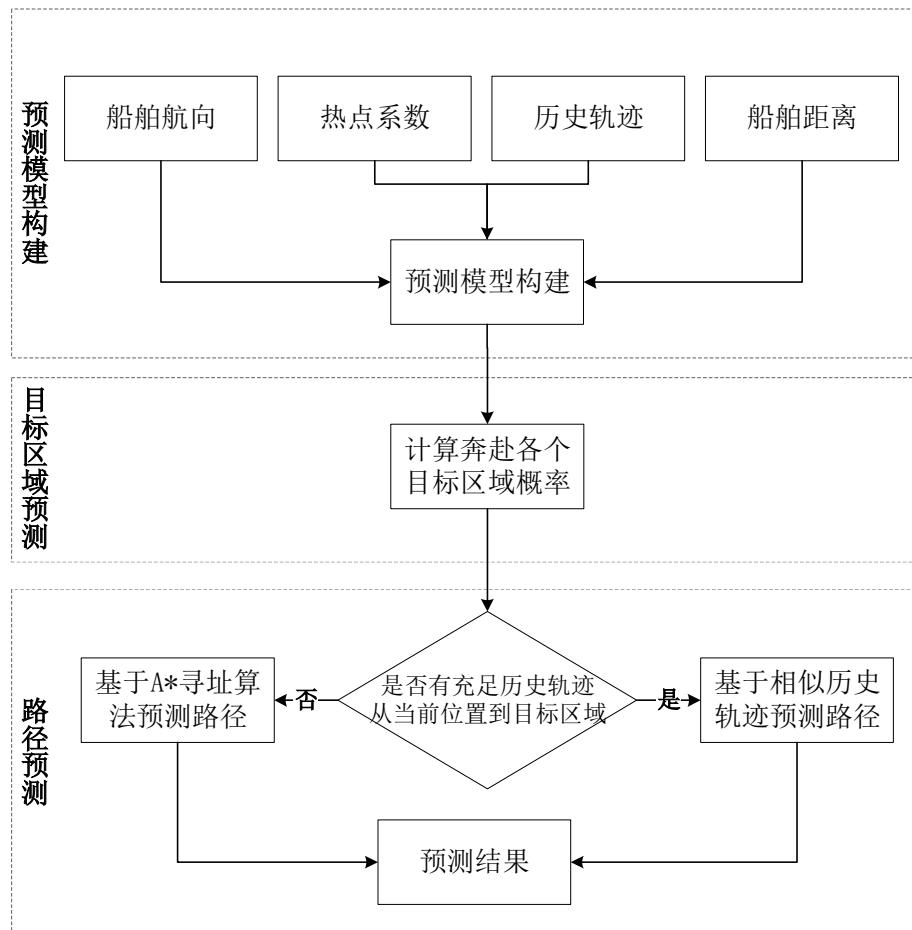


图 4-1 航迹预测总体流程图

Figure 4-1 Overall flow chart of track prediction

4.2 目标区域预测（Target Area Prediction）

目标区域预测指的是预测从当前位置奔赴各个目标区域的概率（以下简称目标奔赴概率）大小。目标奔赴概率的大小会受到天气、路线、距离等众多因素的影响。本文主要考虑历史轨迹、目标区域的热度系数、航向与船舶当前位置到目标区域的距离（下文简称距离）这四个主要因素的影响。目标区域预测首先是目标区域的筛选，从历史轨迹中的热点区域筛选出船舶可能奔向的区域；之后，进行预测模型的构建，通过融合前述的这四个因素为一个统计概率模型，计算出船只奔赴各个目标区域的概率。

4.2.1 目标区域的筛选

在上一章的研究基础上挖掘出历史轨迹的热点区域，但并不是每一个热点区域都可以作为船舶可能奔赴的目标区域，一般与当前航向相背的热点区域基

本可以排除，具体判断方法如下：

首先对当前轨迹进行网格化，计算出与当前航向方向最接近的三个当前网格的相邻网格，如图 4-2 所示，黄色网格代表与当前航向方向最为接近。

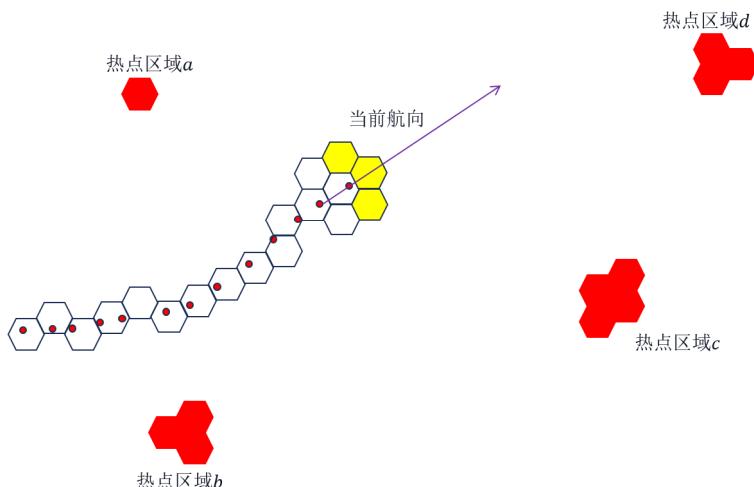


图 4-2 查找与当前航向最为接近的三个相邻网格示意图

Figure 4-2 Schematic diagram of finding the three adjacent grids closest to the current heading

随后预先通过 A*寻址算法求得当前位置到目标区域的初步路线，若该路线未穿过上一步所得的三个网格则认为热点区域为船舶基本上不可能去的区域，需要从可能奔向的目标区域列表中进行删除，否则视为可能奔向的目标区域。判断过程如图 4-3 所示，其中热点区域a、热点区域b为船舶不可能奔向的区域，热点区域c与热点区域d则为目标区域。

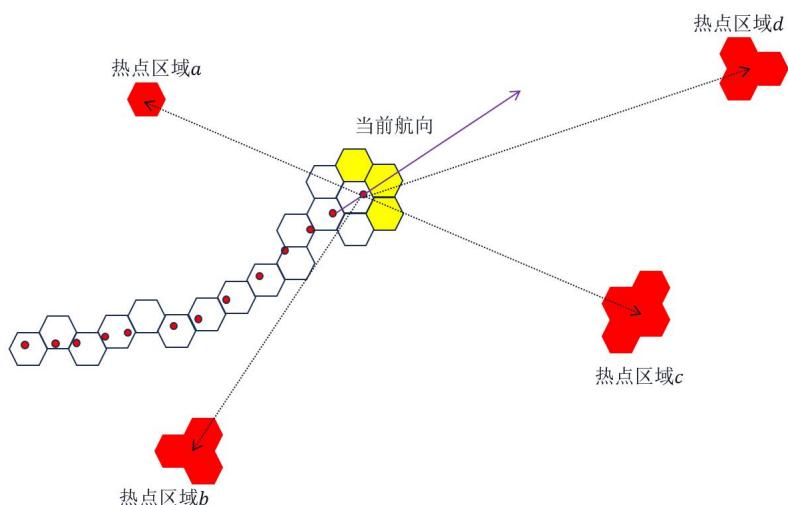


图 4-3 选取目标区域示意图

Figure 4-3 Schematic diagram of selecting target area

4.2.2 预测模型的构建

目标区域筛选之后，需要综合考虑历史轨迹、目标区域的热度系数、航向与距离来构建预测模型。通常而言，当前位置至特定目标区域的历史航迹数量与船舶前往该区域的可能性之间存在正向相关关系；正如前文所讨论，一个区域的热度系数较高意味着其相对吸引度较大，从而增加船舶航向该区域的可能性；此外，船舶的当前航向也会对其未来的行进轨迹产生影响，与当前航线更为一致的目标区域，其成为船舶未来目的地的概率相对较高；最后，距离因素同样不容忽视，目标区域与当前位置的距离越远，其成为下一目的地的概率越低，表现为与距离成负相关的关系。奔赴目标区域的预测模型如图 4-4 所示：

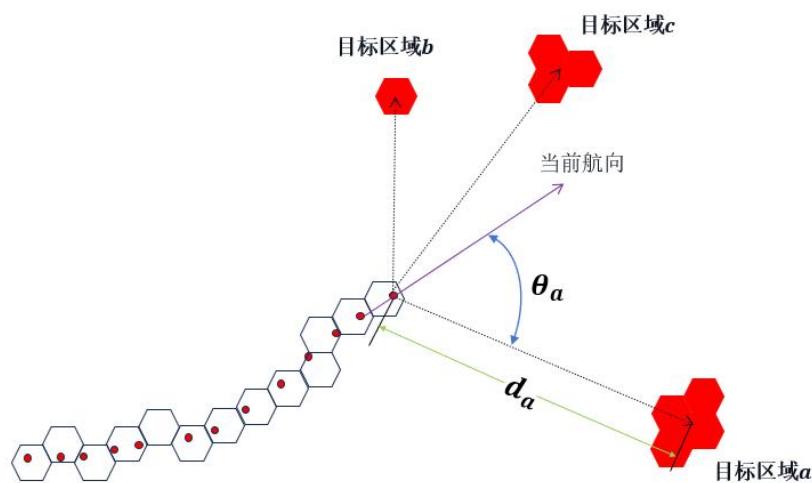


图 4-4 奔向目标区域的预测模型示意图

Figure 4-4 Schematic diagram of the prediction model heading towards the target area

首先是当前轨迹的相似历史轨迹的查找。需要从海量历史轨迹中查找从轨迹出发点处到目标区域的相似历史轨迹，而当前位置到目标区域的轨迹是未知的，因此本文需要重新定义相似历史轨迹。若轨迹 $tracA$ 满足三个条件：（1）轨迹 $tracA$ 经过当前网格；（2）轨迹 $tracA$ 从起点到当前网格的部分与当前轨迹相似；（3）轨迹 $tracA$ 通过目标区域 $targetB$ ，则认为轨迹 $tracA$ 奔向目标区域 $targetB$ 的历史轨迹。判断两条轨迹是否相似下图所示：

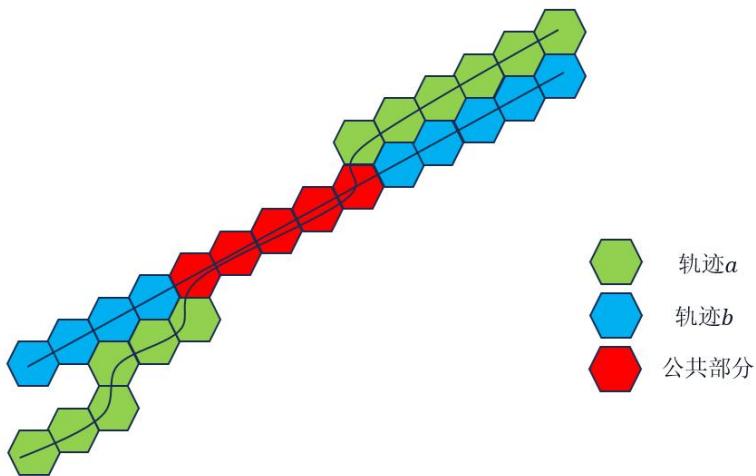


图 4-5 判断两条轨迹相似示意图

Figure 4-5 Schematic diagram of judging whether two trajectories are similar

如图 4-5 所示，对两条轨迹进行网格化之后，从网格数目与相同网格数目进行判断两条轨迹是否相似，具体判断规则如下：

- (1) 若 a 和 b 经过总网格数量之比低于 0.8 或者超过 1.2，则两条航线不相似，否则转规则 (2)；
- (2) 若 a 和 b 共同经过的网格数目超过航线 a 网格总数的 80%，则航线 a 和 b 相似。

完成目标区域的选取与历史轨迹相似计算之后，可计算奔向各个目标区域的概率，计算公式如下：

$$\rho_i = \omega_1 \tau_i + \omega_2 \gamma_i + \omega_3 \alpha_i + \omega_4 \beta_i \quad (4-1)$$

其中， ρ_i 为奔向第 i 个目标区域的概率， ω_1 、 ω_2 、 ω_3 与 ω_4 为历史轨迹、热点系数、航向以及目标区域距离这四个因素的影响因子，通过大量测试与实验，本文取 $\omega_1 = 0.33$ ； $\omega_2 = 0.32$ ； $\omega_3 = 0.1$ ； $\omega_4 = 0.25$ 较为合适。 γ_i 为目标区域的热度系数， τ_i 、 α_i 、 β_i 为第 i 个目标区域在对应影响因素里面所占的比重，计算公式如下：

$$\tau_i = \frac{N_i}{\sum N_j} \quad (4-2)$$

$$\alpha_i = \frac{1/\theta_i}{\sum 1/\theta_j} \quad (4-3)$$

$$\beta_i = \frac{1/d_i}{\sum 1/d_j} \quad (4-4)$$

在以上式子中， N_i 为奔向第 i 个目标区域的历史轨迹数目， θ_i 为航向与当前位置到第 i 个目标区域中心的方向的夹角， d_i 为当前位置到第 i 个目标区域中心的距离。

4.3 路径预测 (Path Prediction)

完成目标区域预测之后，需要预测从当前位置到目标区域的路径。根据当前位置到目标区域的历史轨迹的数量，分为两种情况进行预测：（1）在历史轨迹充足的情况下，依据船舶的历史轨迹进行预测；（2）在历史轨迹欠缺的情况下，采用 A*寻址算法进行路径的预测。

4.3.1 历史轨迹充足的路径预测

历史轨迹充足时，依据船舶的历史轨迹进行路径的预测。目前，在拥有充足的历史轨迹数据的前提下，轨迹预测的研究通常依赖于移动目标的历史移动模式和行为规律。当前的预测算法大多建立在对历史轨迹数据的分析之上，这是基于历史轨迹能够较为准确地揭示出移动目标的行为模式和运动趋势。本文充分借鉴现有船舶轨迹预测方法以及考虑到本文研究特色，研究并提出了一种历史轨迹充足时的路径预测方法，其基本思路为：首先对从当前位置到目标区域所有历史轨迹进行路径的网格化；之后统计历史轨迹穿过每一个网格的频次；最后，从当前网格出发，依次选择穿过频次高的相邻网格作为下一个网格，一直遍历到目标区域网格。具体过程如图 4-6 所示，网格内的数字代表历史穿过频次，灰色网格为预测路径网格。

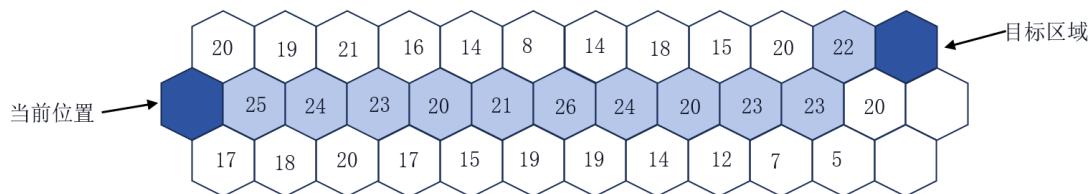


图 4-6 历史轨迹充足的路径预测方法示意图

Figure 4-6 Schematic diagram of path prediction method with sufficient historical trajectories

4.3.2 历史轨迹欠缺的路径预测

当历史轨迹不足或缺失的情况下，现有研究方法的预测精度可能受到严重影响。缺乏充足的历史数据限制了模型对轨迹变化模式的准确捕捉，进而导致预测结果的不稳定性和误差增加。因而，面对历史轨迹数据不足的挑战，本文采用 A* (A-Star)寻址算法进行路径的预测。A*算法是一种启发式搜索算法，它是在状态空间中的搜索，首先对每一个搜索的位置进行评估，得到最优的位置，再从这个位置进行搜索直到目标。A*算法充分考虑了方向、距离等因素对结果的影响，是静态路网中求解最短路径最有效的一种直接搜索方法，也是解决许多轨迹预测问题的有效算法。本文的基于 A*寻址算法的路径预测方法的基本过程为：首先对海域环境的障碍物进行网格化，以排除船只无法到达的区域；

之后对海域环境进行政治代价建模，计算海域内每一个六边形网格的政治代价；最后，从当前位置出发进行位置的搜索，完成路径的预测。其预测步骤具体如下：

为了排除在海域环境中船只无法穿过的障碍物，首选对海域中的障碍物进行网格化。本文的障碍物具体指的是大陆、岛屿与半岛的海岸线，网格化的效果如图 4-7 所示：

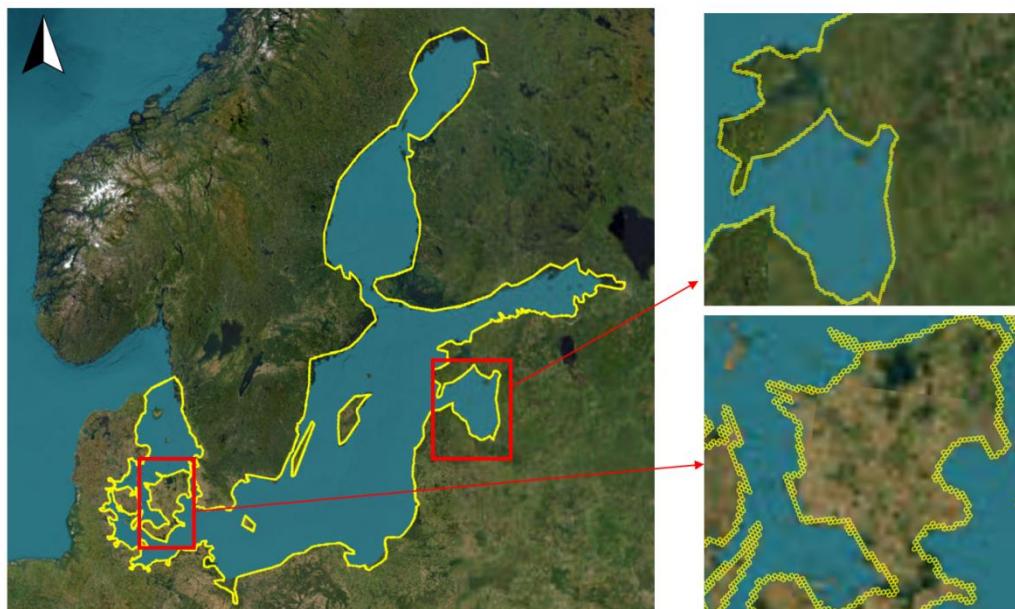


图 4-7 海域障碍物网格化结果
Figure 4-7 Sea area obstacle gridding results

根据距离海岸线距离的不同，依次将海域各区域分为内海、领海、毗连区、专属经济区与公海。船只在穿过各不同区域时政治代价（即航道权）不一样，其中内海的政治代价最高，公海的政治代价最低。换言之，如果在距离成本相同的情况下，船只优先考虑政治成本较低的路线进行航行。因此，本文 A*寻址算法的代价计算不仅涉及到距离代价，同时也需要考虑政治成本代价。对不同区域的政治成本代价具体计算方法如下：

由于内海作为船只进入或者离开港口的必经区域，视穿过内海所有的六边形网格的政治成本为一致；对于领海、毗邻区，穿过六边形网格的政治成本随着距离海岸线的增大而减小；公海内的每一个六边形网格的政治成本认定为 0；另外，研究海域未涉及到专属经济区。假设海域内网格为 g ，内海区域 A_1 ，领海区域 A_2 ，毗邻区 A_3 ，领海区的网格中心到内海与领海区分界线的最小距离为 d_1 ，毗邻区的网格中心到毗邻区与领海区分界线的最小距离为 d_2 ，示意图如图 4-8 所示。

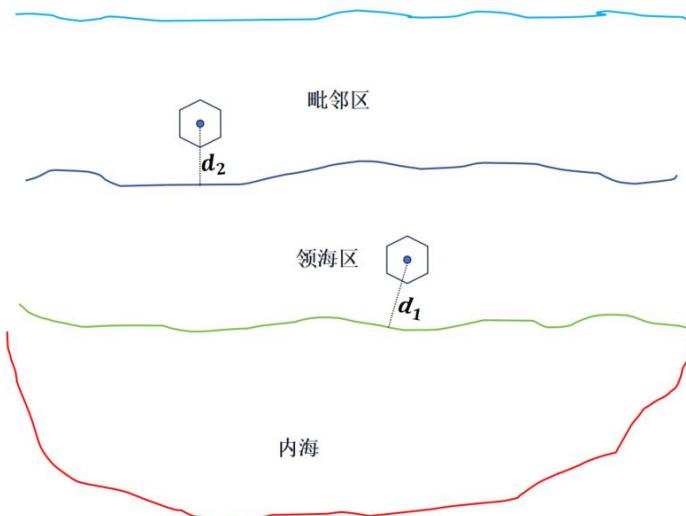


图 4-8 网格至分界线最短距离示意图

Figure 4-8 Schematic diagram of the shortest distance from the grid to the dividing line

得到网格中心到分界线最小距离之后，可计算海域内的网格对应的政治成本大小，具体计算公式如下：

$$cost_p = \begin{cases} 0.7R & g \in A_1 \\ R(0.7 - \frac{d_1}{30}) & g \in A_2 \\ R(0.3 - \frac{d_2}{40}) & g \in A_3 \end{cases} \quad (4-5)$$

上式中， $cost_p$ 为政治权重大小，R为六边形网格的边长，R、 d_1 与 d_2 的单位均为km。

为了展示海域环境政治代价建模的效果，根据每一个六边形网格的政治代价大小采用不同的颜色进行填充网格，颜色越深表示政治代价越大。建模效果如图 4-9 所示。

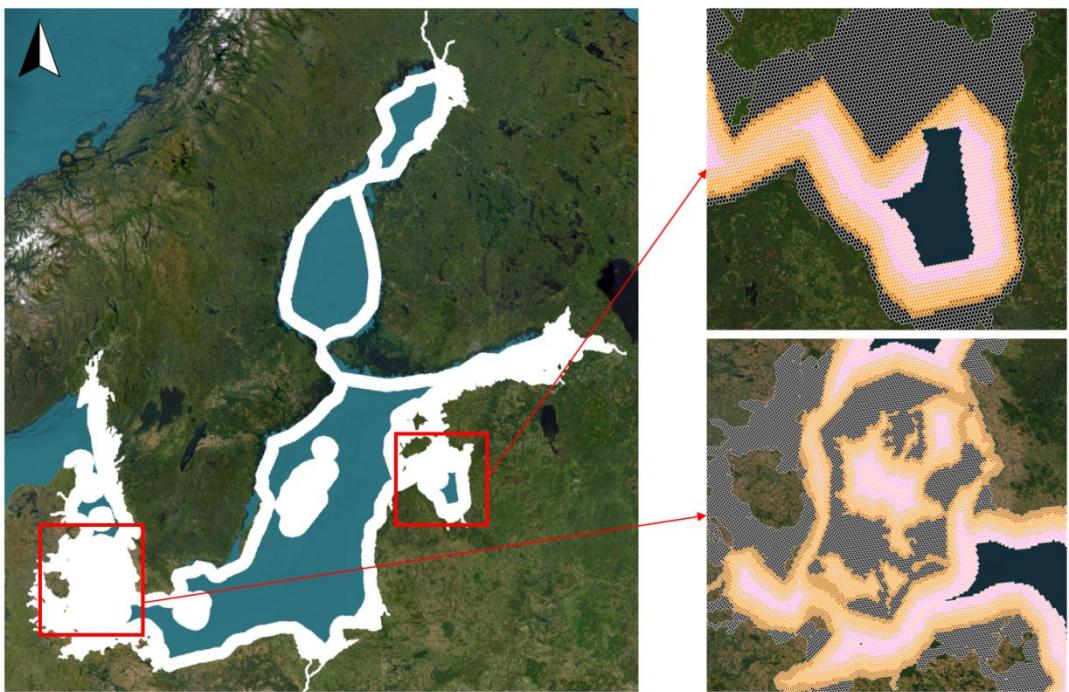


图 4-9 海域环境政治代价建模结果

Figure 4-9 Maritime environmental political cost modeling result

在完成海域障碍物的网格化与网格政治代价计算之后，基于六边形网格采用 A*算法进行最优路径的搜索。A*算法是一种建立在 Dijkstra 算法和最佳优先搜索(best-first-search, BFS)算法基础上的路径搜索算法，其函数表现形式为：

$$f(n) = g(n) + h(n) \quad (4-6)$$

在上式中， $f(n)$ 为估价函数值， $g(n)$ 为从起始节点到节点 n 的代价， $h(n)$ 为从节点 n 到目标节点的估计代价。代价包括距离成本代价与政治成本代价，其中政治成本代价的大小按照前文所述方法进行计算，距离成本代价的大小采用曼哈顿距离公式进行计算，具体计算如下：假设网格 A 与网格 B 的编码分别为 (i_A, j_A) 与 (i_B, j_B) ，网格半径为 r ，距离成本代价 $cost_d$ 计算如下所示：

$$cost_d = r(|i_A - i_B| + |j_A - j_B|) \quad (4-7)$$

A*算法主要通过两个表（open 表与 close 表）来优化路径搜索过程。open 表存储待评估的节点，即那些已被发现但尚未评估的节点；而 close 表包含已经评估的节点。在每一步，算法从 open 表中选取具有最低成本估计的节点进行扩展，将其移至 close 表，并将其邻近未评估的节点添加到 open 表。这一过程持续进行，直到找到目标节点或 open 表为空。通过这种机制，A*算法能够有效地遍历搜索空间，同时避免重复评估节点，从而高效地找到最佳路径。下面给出了 A*算法的寻址算法的具体步骤：

- (1) 初始化起点和终点节点，并将起始节点放入 open 表中；
- (2) 从 open 表中选择具有最小总代价的节点作为当前节点，并将其移动到 close 表；

- (3) 对当前节点的六个相邻节点进行遍历，计算各相邻节点的代价，并更新其父节点和总代价；
 (4) 将所有未处理的相邻节点加入开启列表，并按总代价排序；
 (5) 重复步骤(2)到(4)，直到找到终点节点或 open 表为空；
 (6) 如果找到终点节点，通过回溯父节点路径可以得到最优路径；否则说明无法到达终点。

上述过程为 A* 算法的寻址过程，具体的伪代码如表 4-1 所示：

表 4-1 基于六边形网格的 A* 寻址算法伪代码

Table 4-1 Pseudo code of A* addressing algorithm based on hexagonal grid

Algorithm 基于六边形网格的 A* 寻址算法伪代码

输入变量：起点节点与目标区域网格节点

输出变量：预测路径 *path*

```

1:      Begin
2:          // 初始化起点和终点节点，通过编码转换得到起点与终点的编码
3:          start_node = Node(start, nil)
4:          goal_node = Node(goal, nil)
5:          // 初始化开启列表和关闭列表
6:          open_list = [start_node]
7:          closed_list = []
8:          While open_list is not empty
9:              // 选取开启列表中代价最小的节点作为当前节点
10:             current_node = node in open_list with the lowest f_cost
11:             // 将当前节点从开启列表中移除，加入关闭列表
12:             Remove current_node from open_list
13:             Add current_node to closed_list
14:             // 到达终点节点，返回路径
15:             If current_node == goal_node
16:                 Return path
17:             End If
18:             // 扩展当前节点的相邻节点
19:             For neighbor in current_node.neighbors
20:                 // 如果相邻节点已经在关闭列表中，跳过
21:                 If neighbor in closed_list
22:                     Continue
23:                 End If

```

表 4-1 (续)

Algorithm	基于六边形网格的 A*寻址算法伪代码
24:	// 计算到起点的代价
25:	$new_g_cost = current_node.g_cost + cost(current_node, neighbor)$
26:	// 如果相邻节点不在开启列表中，将其加入
27:	If $neighbor$ not in $open_list$
28:	Add $neighbor$ to $open_list$
29:	$neighbor.g_cost = new_g_cost$
30:	$neighbor.h_cost = heuristic(neighbor, goal_node)$
31:	$neighbor.parent = current_node$
32:	// 否则，如果新的代价更小，则更新
33:	Else
34:	If $new_g_cost < neighbor.g_cost$
35:	$neighbor.g_cost = new_g_cost$
36:	$neighbor.parent = current_node$
37:	End If
38:	End If
39:	End for
40:	End while
41:	// 开启列表为空，无法到达终点
42:	Return $null$
43:	End

4.4 可行性分析实验（Analysis Feasibility Analysis Experiment）

为验证本章提出的目标区域预测以及路径预测方法的可行性，本节采用波罗的海 AIS 数据进行实验，选取若干条具有典型特征的真实航迹的多个关键时间点的预测路径与真实轨迹进行比较分析。本文选取船舶 MMSI 分别为 205366000、211190000 与 209155000 的三条轨迹进行实例分析，简称轨迹 L₁、L₂、L₃，各轨迹分布如图 4-10 所示，其中：轨迹 L₁ 与 L₂ 的历史轨迹充足，L₁ 是长距离航程，而 L₂ 则短距离航迹；相反，L₃ 缺乏历史轨迹，并呈现出广阔的空间覆盖度。以下对各轨迹做实例分析：

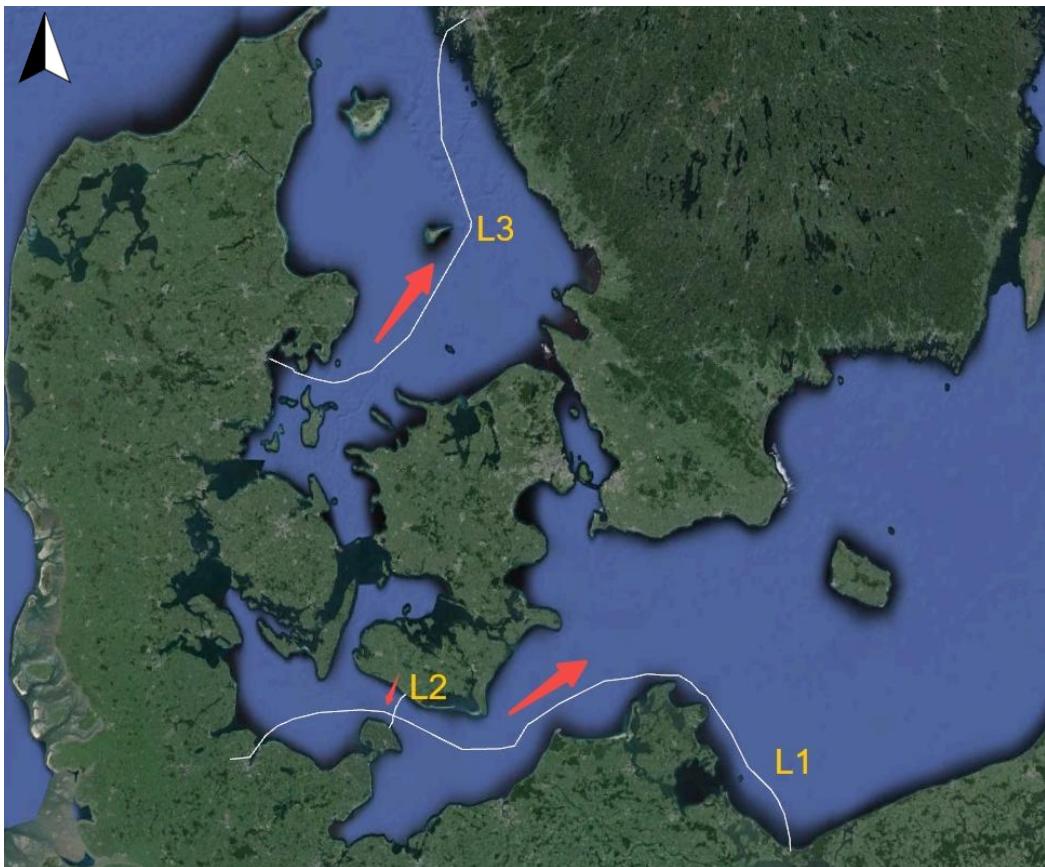


图 4-10 三条测试轨迹分布图

Figure 4-10 Distribution diagram of three test tracks

(1) 轨迹 L_1

轨迹 L_1 覆盖的范围广泛，导致其奔赴的目标区域存在显著较大的不确定性。选取该航迹的四个时间点进行实验分析，四个时刻的预测结果如图 4-11、4-12、图 4-13 与图 4-14 所示，其中白色圆点整条原始轨迹，黄色网格为已航行的网格，绿色区域为筛选后的目标区域，红色网格为预测路径，同时图中也给出了奔赴目标区域的概率大小（为了方便展示说明，图中只展示奔赴概率最高的 6 个目标区域的概率）。从图看出：在初始预测时刻 A，船只奔赴各个目标区域的概率分布相对均衡，分别为 13%，15%，12%，15%，12% 和 9%，这表明了船只奔赴特定区域的可能性存在微小的差异。随着船只的航行进展，预测时刻所指向的目标区域呈现动态变化，尤其表现为目标区域在数量上的减少。当船只逐渐接近预定目的地时，各时刻赴目的地的概率呈现递增趋势，分别为 12%，24%，36%，直至在预测时刻 D 达到 70%，显示出船只奔赴目的地的倾向性显著增强。这一趋势与实际航行场景相吻合，意味着可能奔赴的目标区域随着与终点距离的缩短而减少，对于船只最终奔赴的目的地能够以更高的可能性进行确认。此外，观察预测所得轨迹与实际航迹的对比，可以明显观察到预测轨迹展现出一定程度的平滑性，在各个预测时刻与真实航迹的偏差较小，尤其是预

测时刻 D 显示出较高的一致性。这表明所采用的预测模型能够有效模拟船只的实际航行行为，捕捉其航行动态，并且在航迹预测的准确性上取得较为满意的结果。

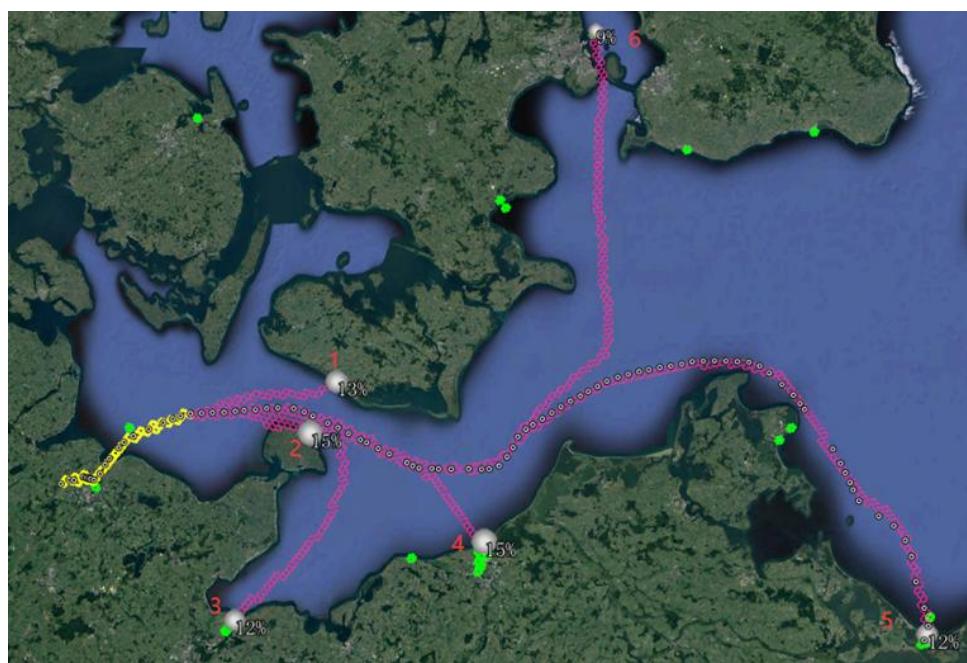


图 4-11 轨迹 L_1 时刻 A 预测结果

Figure 4-11 Prediction results at trajectory moment A

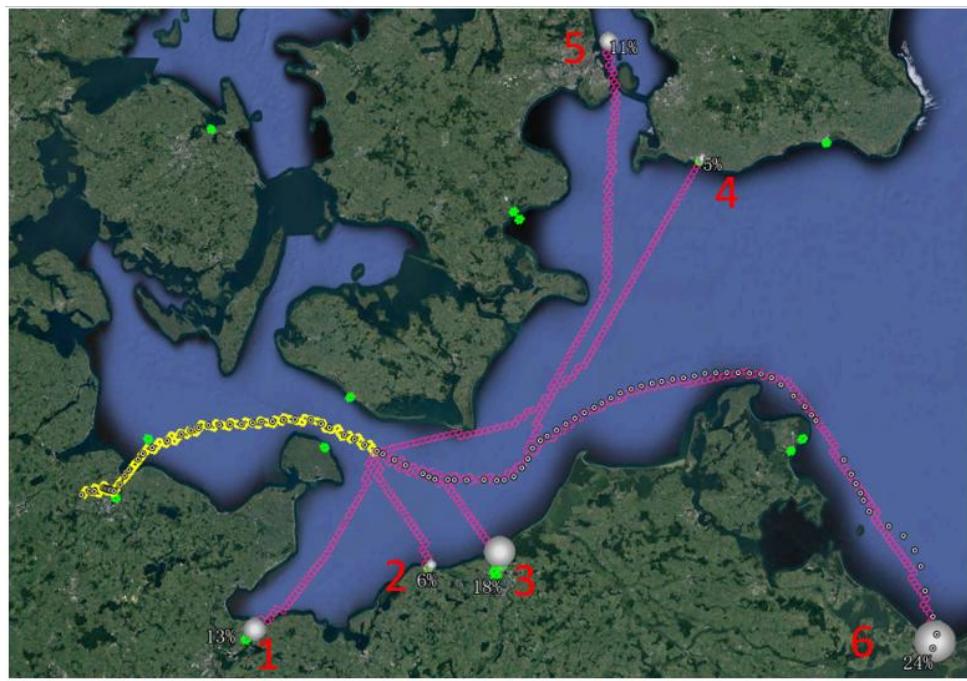


图 4-12 轨迹 L_1 时刻 B 预测结果

Figure 4-12 Prediction results at trajectory moment B

图 4-13 轨迹 L_1 时刻 C 预测结果

Figure 4-13 Prediction results at trajectory moment C

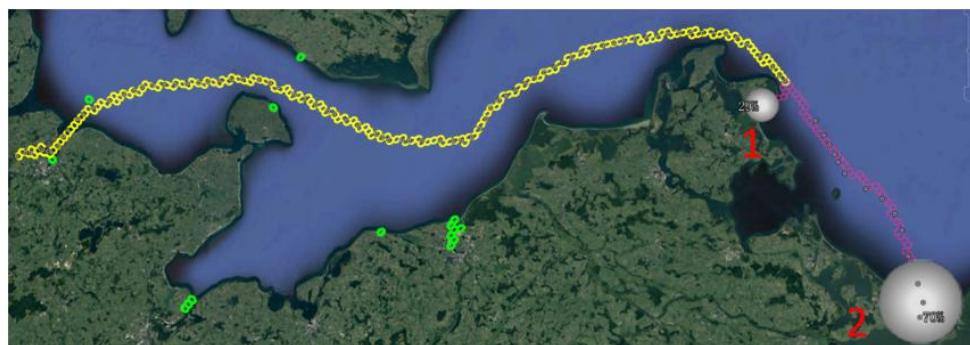
图 4-14 轨迹 L_1 时刻 D 预测结果

Figure 4-14 Prediction results at trajectory moment D

(2) 轨迹 L_2

鉴于轨迹 L_2 具有较小的空间跨度和特定的航向特征，从而导致其目标区域的不确定性相对较低，本实验选择针对该航迹的单一时间点进行预测分析。实验结果如图 4-15 所示，由此可知：船只向目标区域 6 进发的概率显著高于对其他目标区域的，这一现象可以归因于目标区域 6 在航向定位、距离近远、历史轨迹分析以及热度系数等方面相比其他区域展现出显著的综合优势。因此，在该预测时刻，可以较为准确地确定船只未来的目的地，这一预测与实际情况的一致性验证了预测模型的有效性。

图 4-15 轨迹 L_2 某时刻预测结果

Figure 4-15 Prediction results at a certain moment in the trajectory

(3) 轨迹 L_3

轨迹 L_3 与轨迹 L_1 、 L_2 相比，缺乏历史轨迹数据，导致预测具有较高的不确定性和随机性。本次实验选取预测时刻 A、B 与 C 对该轨迹进行实验分析，图 4-16 给出了实验结果，由图可知：在预测时刻 A，由于缺乏历史轨迹数据、与目的地的相对距离较远（数百公里）以及存在多条可行航线，预测路径与实际轨迹之间存在显著偏差。此外，可以明显观察到，目标区域 1 和目标区域 6 的奔向概率高于其他区域，这一现象归因于这两个区域较高的热度系数。进入预测时刻 B，潜在的目标区域发生了显著变化，终点（目标区域 6）的奔向概率增至 26%，且相较于预测时刻 A，预测路径与实际轨迹之间的偏差程度有所降低。进入预测时刻 C，奔向终点（目标区域 3）的概率高达 46%，且预测路径与真实航迹偏差显著减小。随着船只向终点的靠近，奔向的目标区域越来越清晰，这也现象与现实情况是相符合的。这一观察说明，尽管存在不确定性，所采用的预测方法在一定程度上能够反映实际航行情况，证明了其可行性和实用价值。

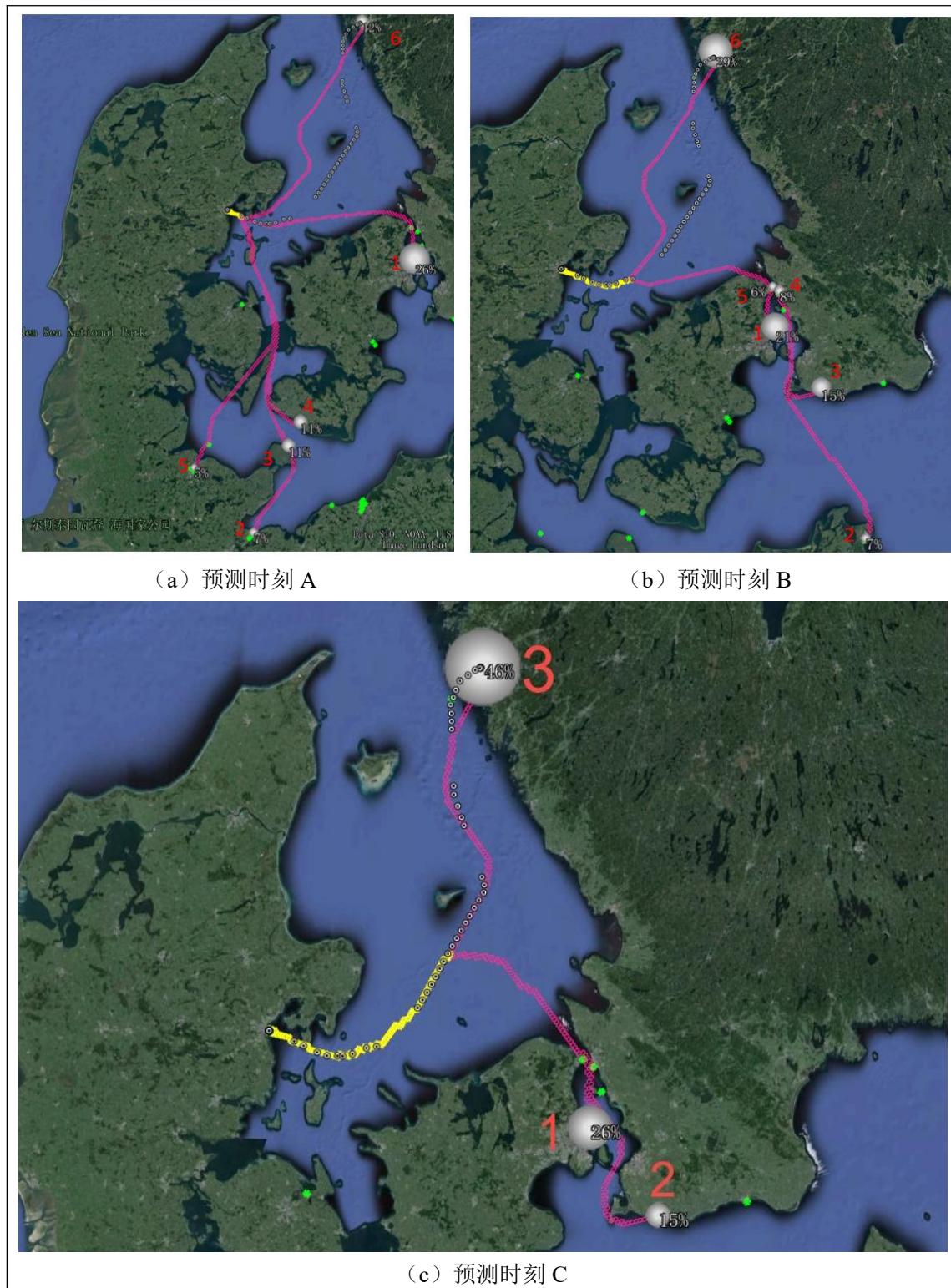
图 4-16 轨迹 L_3 预测结果

Figure 4-16 Trajectory prediction results

基于上述实验结果及其分析可知：本文提出的船舶轨迹预测方法展现出了一定的可行性与可靠性。在广阔海域中进行长距离预测时，该方法能够生成平滑的轨迹预测路径，一般情况下与实际轨迹之间没有显著偏差，特别是在临近预测终点时刻，真实航迹与预测轨迹的显著吻合程度进一步验证了方法的有效性。

性。这表明本方法在处理海洋广域范围内的航迹预测问题时，能够提供较为精确的结果，这对于海上航行安全管理和航迹优化等应用场景具有重要的实际意义和应用价值。

4.5 预测精度分析实验（Prediction Accuracy Analysis Experiment）

4.5.1 实验方案

为进一步验证评估船舶轨迹预测方法的准确性，实验采用第二章所述的LSTM预测方法作为比较对象，展开了一系列的数据预处理、参数优化与设置工作，确保实验的严谨性和结果的可靠性。

本文将从两个评价指标来对比两种预测方法，包括均方根误差（RMSE）、距离最大偏差（ DIS_{max} ）。RMSE被广泛用于对测试样本的预测情况进行评估，它提供了一个整体的度量方法，用于评估预测结果与真实值之间的偏离程度。RMSE通过计算预测值与真实值之间的差异的均方根误差，并以平方根的形式呈现，以全面评估预测结果与真实值之间的整体准确性。 DIS_{max} 则主要用于检测预测结果的极端情况，它能够探测出那些与真实值相比偏离过大的预测结果。距离最大偏差的目的在于评估方法的稳定性，通过发现那些具有较大偏离的预测结果，从而更好地了解模型的预测能力和异常情况。以下具体介绍这两个评价指标的具体含义：

(1) 均方根误差。使用均方根误差 $RMSE$ 作为误差评价指标， $RMSE$ 值越小说明误差越小，其数学表达式如下：

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (4-9)$$

上式中，其中 m 为样本数量， y_i 为实际值， \hat{y}_i 为预测值。

(2) 距离最大偏差。距离最大偏差 DIS_{max} 指的是所有测试数据中的预测轨迹与真实轨迹距离偏离最大值，其具体计算公式如下：

$$DIS_{max} = \max_{i=1} \sqrt{(lon_t - lon_p)^2 + (lat_t - lat_p)^2} \quad (4-10)$$

上式中， lon_t 为真实经度， lat_t 为真实纬度， lon_p 为预测经度， lat_p 为预测纬度。

在采用LSTM方法训练数据之前，需要对实验数据预处理。继第三章轨迹预处理之后，进一步对数据做时间对齐与归一化处理，以确保数据质量和一致性。为获得时间间隔相等的航迹数据，使用三次样条函数对数据进行插值与平滑处理，获得时间间隔为10min的等间隔数据。由于不同的特征向量可能在数量级上差别很大，可能导致绝对值小的数据被大数据“吃掉”的情况。因此，为了将不同尺度的特征值统一到一个固定范围内，故将采用min-max标准化方

法对数据进行归一化处理。具体转换公式如下：

$$X_n = \frac{X_f - X_{min}}{X_{max} - X_{min}} \quad (4-8)$$

上式中， X_f 为原始值， X_{max} 、 X_{min} 分别为样本数据的最大值与最小值， X_n 为归一化的数据。处理后，原始数据值范围转换为[0,1]，提高了后续神经网络训练的效率和效果。

实验数据继续沿用 2017-2019 两年间波罗的海的 AIS 数据，鉴于实验区域涵盖广泛的经纬度范围，为了平衡时间性能和误差精度之间的关系，实验统一设置六边形网格的半径平均为 0.6km（每一个球面六边形网格半径大小有微小差异）。根据前文所述，本文预测方法涉及到多个目标区域的奔赴概率计算，为了便于与 LSTM 预测方法进行比较，将采用奔赴概率最高的目标区域的预测路径与 LSTM 方法结果进行比较分析。在数据集训练之前，需要初始化 LSTM 网络模型的参数，通过 Xavier 方法对网络权值初始化，使用网格搜索的方法对 LSTM 网络时间步、隐含层节点、学习率、批大小等参数进行寻优设置。本文 LSTM 网络具体参数设置如下：隐藏层数为 3，节点数为 100，学习率为 0.001，输入层节点为 4，输出层节点为 4，激活函数为 Relu 函数。实验选取 80% 的数据集进行训练（本文方法用于构建预测模型，LSTM 用于模型训练）和 20% 的数据集进行测试。为了评估本文方法在历史轨迹充足与不足的条件下的预测性能，测试数据集被分为充足历史轨迹组和欠缺历史轨迹组，各占测试数据集的 50%。同时，采用前 30min 的航行轨迹预测未来 120min 的轨迹，并通过评估均方根误差和最大距离偏差两个指标来进行量化分析。

4.5.2 实验结果与分析

基于上述实验方案，以下对两种预测方法分别在测试数据历史轨迹充足与欠缺时的实验结果进行讨论分析：

(1) 历史轨迹充足

对于本文方法预测结果，借助速度（后续速度一直保持当前时刻的速度）与航行时长推导出相应时间节点的点坐标。表 4-2 与图 4-17 给出了两种预测方法的 RMSE 的对比情况，由此可以看出：在时间 10min~40min 的每个预测节点，本文方法相较于 LSTM 分别高了近 32%、12%、33% 与 15%，具有更大的预测误差。但是，在 70min 时之后，LSTM 的 RMSE 均高于本文方法，体现了本文方法在长远距离预测的优越性；随着距离的增大，尤其是到了 90min 之后，两种方法的 RMSE 的值大小有了更为明显的上升，误差达到了接近甚至超过了 2km，这说明长距离给轨迹预测带来了极大的随机性；从整体看，两者的 RMSE 都相对较大，这与海域环境的复杂性有关。海域的轨迹预测面临的不确定性远远大

于河流预测，尤其是在长距离预测时，这种不确定性变得更为显著。本文方法在一定程度上能够减少这种随机性，从而提高了长距离预测的准确性和可靠性。

表 4-2 历史轨迹充足时 RMSE 对比

Table 4-2 RMSE Comparison when historical trajectories are sufficient

时间(min)	本文方法(km)	LSTM(km)
10	0.42	0.317
20	0.783	0.695
30	1.118	0.839
40	1.29	1.113
50	1.386	1.271
60	1.527	1.465
70	1.559	1.529
80	1.817	1.905
90	1.933	2.165
100	2.456	2.604
110	2.588	2.798
120	2.949	3.208

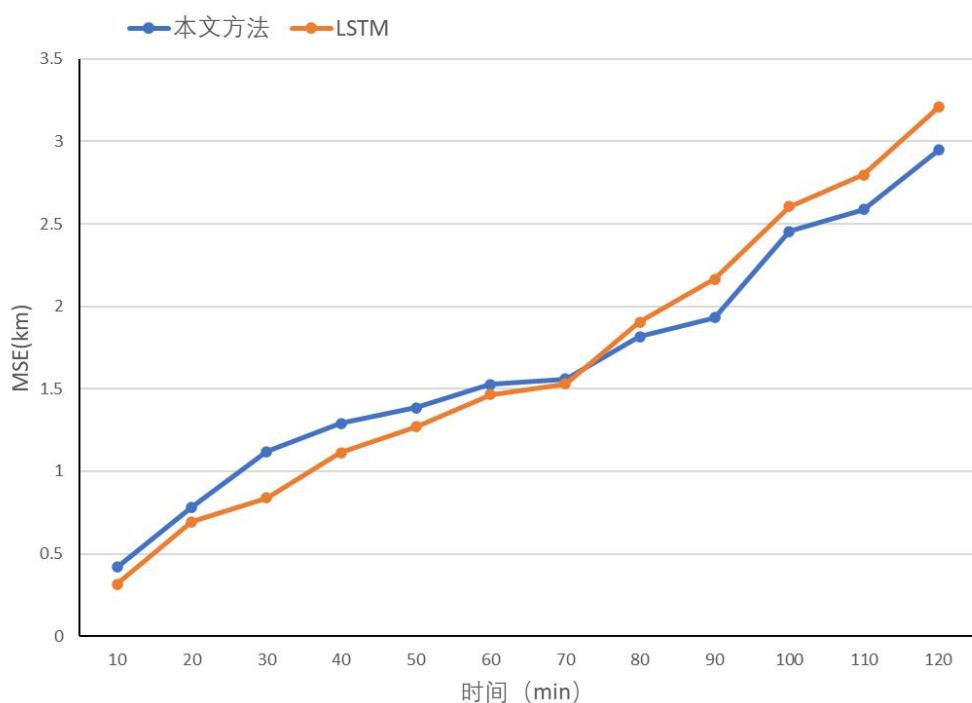


图 4-17 历史轨迹充足时 RMSE 对比结果

Figure 4-17 RMSE comparison results when historical trajectories are sufficient

同时，为探究分析两种预测方法在长短距离的稳定性，本文按照时间序列来计算分析每一个时间节点的 DIS_{max} 指标。表 4-3 与图 4-18 给出了两种方法在

历史轨迹充足时的 DIS_{max} 对比结果，由此可知：在 20min 与 40min 节点处，本文方法的距离最大偏差值高于 LSTM，但是在 80min 以后，LSTM 的最大距离偏差显著高于本文方法。这一分析结果表明，在历史轨迹充足时，LSTM 方法在短距离预测上具有较好的性能；然而，对于长距离预测，本方法展现出更加优越的稳定性。这种稳定性主要归因于对目的地以及航道等因素的约束，这些约束有效地降低了长距离预测中的随机性。

表 4-3 历史轨迹充足时 DIS_{max} 对比

Table 4-3 Comparison when historical trajectories are sufficient

时长(min)	本文方法(km)	LSTM (km)
20	0.979	0.8
40	1.882	1.658
60	2.767	2.881
80	3.313	4.275
100	3.898	5.411
120	4.527	6.552

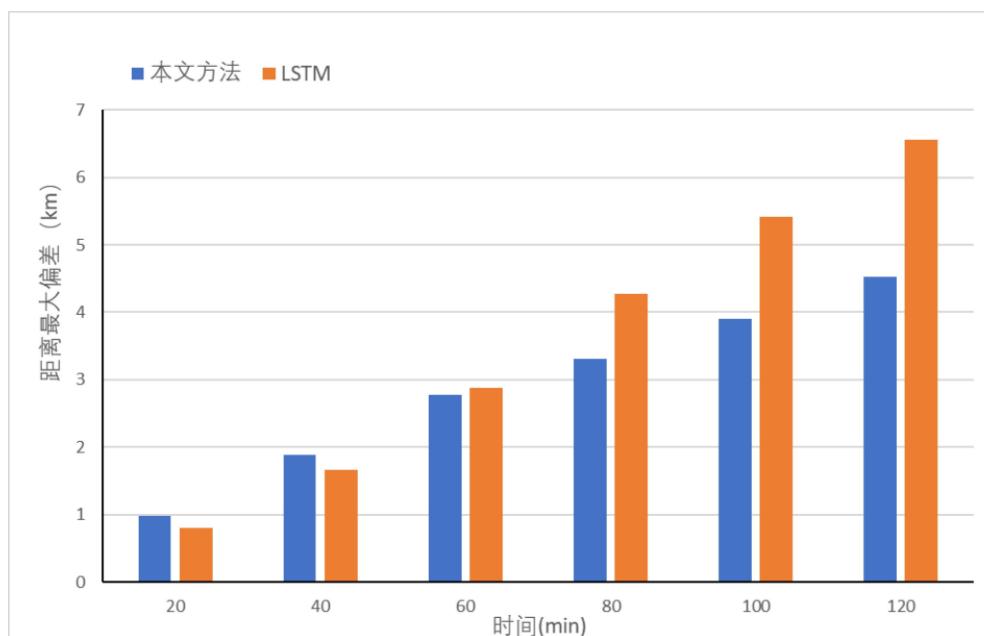
图 4-18 历史轨迹充足时 DIS_{max} 对比结果

Figure 4-18 Comparison results when historical trajectories are sufficient

(2) 历史轨迹欠缺

表 4-4 与图 4-19 给出了历史轨迹不足时两种预测方法 $RMSE$ 的对比情况，由图表看出：从整体上来看，二者的 $RMSE$ 值都相对较大，说明缺乏历史轨迹对两种预测方法均造成一定的精度损失；在 10min~120min 的每一个预测节点，本文方法均优于 LSTM，到了后期优势愈加显著，两种方法的 $RMSE$ 值差距逐渐拉大，在 120min 处本文方法相对比 LSTM 约小 2.264km。由此可知，历

史轨迹不足时，不管长距离预测还是短距离预测，本文方法相对于 LSTM 都具有更高的预测精度。

表 4-4 历史轨迹不足时 RMSE 对比

Table 4-4 RMSE comparison when historical trajectories are insufficient

时间(min)	本文方法(km)	LSTM(km)
10	0.451	0.601
20	0.877	1.025
30	1.222	1.377
40	1.511	1.548
50	1.799	1.971
60	2.044	2.465
70	2.423	2.852
80	2.681	3.444
90	3.024	4.149
100	3.579	4.895
110	3.906	5.72
120	4.731	6.995

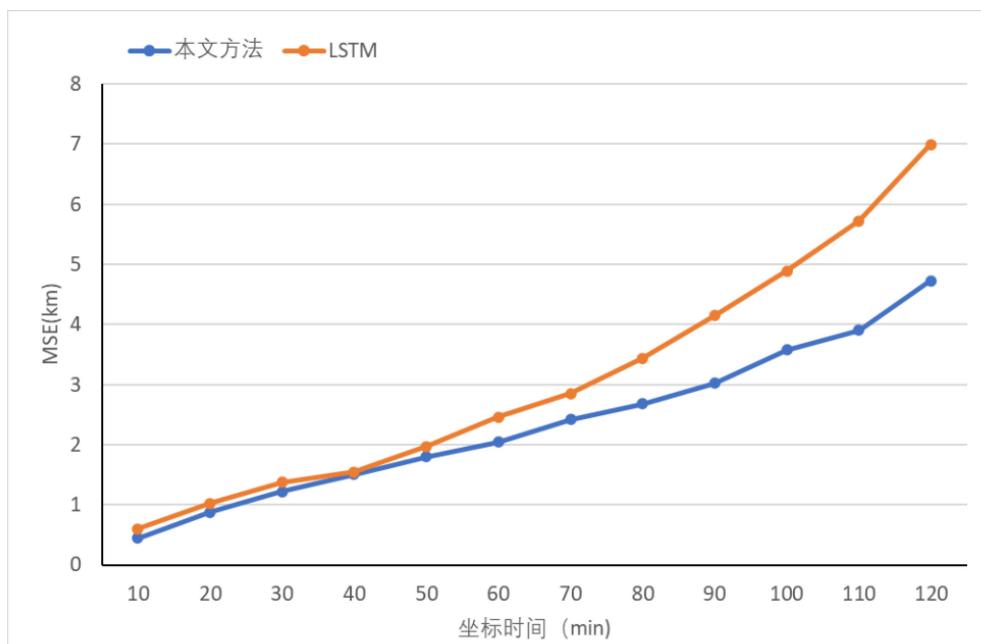
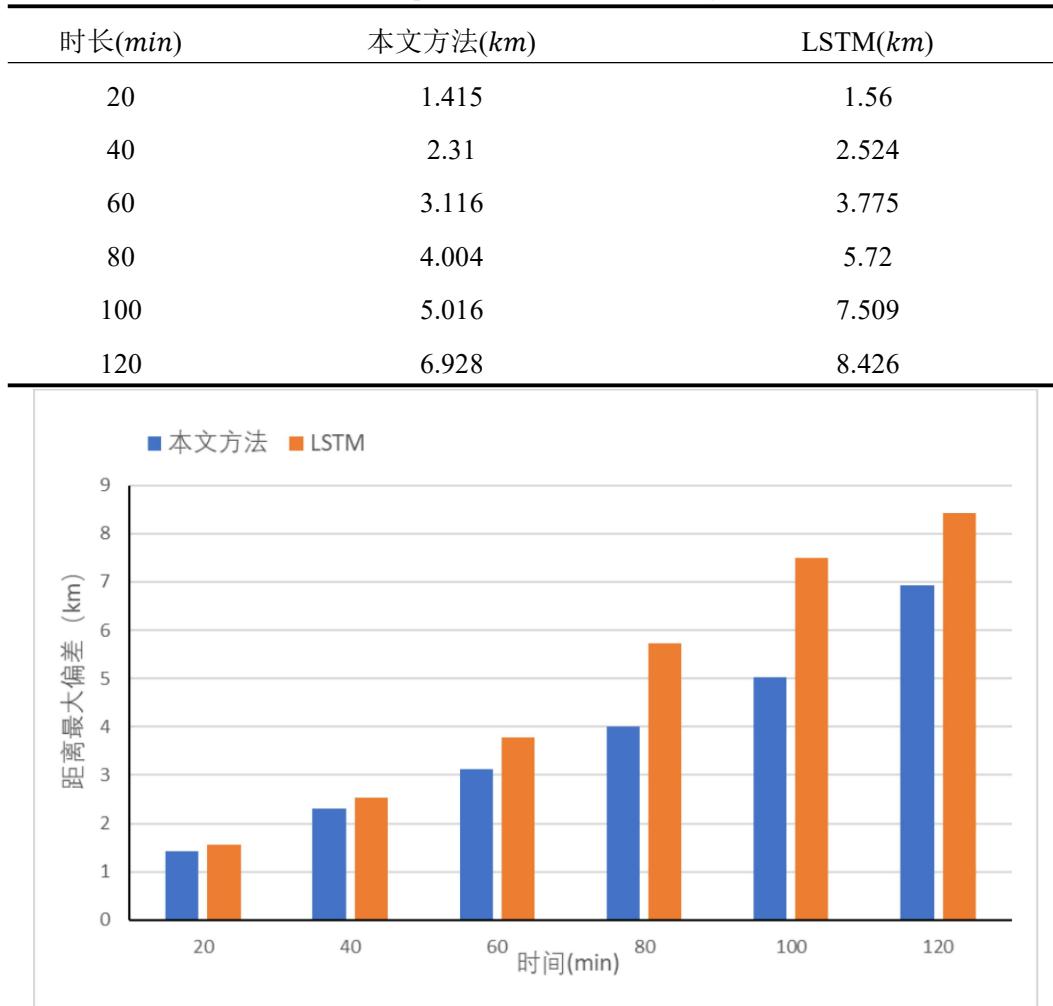


图 4-19 历史轨迹不足时 RMSE 对比结果

Figure 4-19 RMSE comparison results when historical trajectories are insufficient

表 4-5 与图 4-20 给出了两种方法在历史轨迹不足时的 DIS_{max} 对比结果，由此看出：从整体结果来看，随着时间的增长，两种预测方法在各时刻的 DIS_{max} 均呈现出增大的趋势，但是对应时刻两者的距离最大偏差都保持在相对较小的范围内，这也从侧面反映了两种预测方法都具有一定的稳定性，没有出现与真

实值偏离过大的极端值；在 20 min~120 min 的每个时间节点，本文方法的 DIS_{max} 值均小于 LSTM。基于上述分析可知：在历史轨迹不足时，本文方法更具有优越性。

表 4-5 历史轨迹不足时 DIS_{max} 对比Table 4-5 DIS_{max} comparison when historical trajectories are insufficient图 4-20 历史轨迹不足时 DIS_{max} 对比结果Figure 4-20 DIS_{max} comparison results when historical trajectories are insufficient

综上分析可知：两种预测方法在海域环境下的误差都较大，这与海域预测时船舶前进方向具有极大的随机性和长远距离预测的不确定性有关；在历史轨迹充足的情况下，LSTM 在短距离预测相对于本文方法更具有优越性，但面对长远距离时，本文方法预测精度相对更高，这一优势源于方法本身采用了以目的地为导向的预测策略，目的地的约束作用在一定程度上降低了长距离预测过程中的不确定性；在历史轨迹不足时，本文方法相对于 LSTM 具有一定的优势，在长短距离的预测上都表现出较高的预测精度。

4.6 本章小结 (Summary of this Chapter)

本章基于全球六边形网格提出了以目的地为导向的船舶轨迹预测方法，并从该方法的基本思路、具体内容、对比实验等方面来详细分析。首先，针对现有研究方法长距离预测精度低、忽略目的地影响、历史轨迹少误差大等问题，提出了一种以目的地为导向的船舶轨迹预测方法并给出了基本思路；然后，具体阐述预测方法的主要两个步骤，先从目标区域的选取、预测模型的构建介绍目标区域预测，再从历史轨迹充足与不足来介绍路径预测具体过程；之后，为验证本文预测方法的可行性，设计了可行性分析实验。实验结果表明：本文预测方法具有一定的可行性，预测轨迹与实际航行路径展现出较高的一致性。面向海洋广域范围内的长远距离预测时，能够提供较为精确的结果。最后，在验证可行性的基础上，为进一步评估本文方法的准确性，开展了预测精度分析实验。实验结果表明：在历史轨迹充足的情况下，本文预测方法相对于 LSTM 在短距离的预测误差较大，但是在面向长距离预测时，本文方法预测精度更高；历史轨迹不足时，本文方法相对于 LSTM 预测方法在长短距离都具有更优的预测精度。

5 航迹预测原型系统开发与展示

5 Development and Demonstration of a Trajectory Prediction Prototype System

前几章从理论层面具体介绍了航迹预测涉及到的数据预处理、航迹网格化、历史轨迹热点区域挖掘与航迹预测，并设计相关实验来验证方法的可行性与准确性。这一章整合前几章的相关技术与算法整合到航迹预测原型系统中，对系统开发过程中所使用的技术及其应用进行详细介绍，并展示系统实现的界面功能。

5.1 系统开发环境（System Development Environment）

航迹预测原型系统的硬件配置环境为一台台式电脑，其具体配置为：i5-7500 CPU@3.40GHz,16GBRAM;操作系统;Windows10专业版。本系统采用C/S架构，开发所使用的IDE为Visual Studio 2019。后端采用C++语言进行开发，基于GDAL、boost、DGGRID与jsoncpp等相关第三方框架完成算法与原型系统后台的实现。前端界面基于Qt框架进行开发，同时借助OSG与osgEarth完成相关地图功能的开发。本系统基于http服务实现后端与前端的交互，数据库选用MySQL进行系统数据的管理。涉及到的软件开发环境如下表所示：

表 5-1 软件环境表

Table 5-1 Software environment table

类型	技术与依赖库	版本
后端	C++	C++17
前端	Qt	Qt5.12.0
算法	C++	C++17
专业软件	ArcGIS QGIS	ArcGIS10.7 QGIS3.22.4
开发工具	Visual Studio	Visual Studio 2019
数据库	MySQL	MySQL8.3

5.2 系统功能（System Function）

该系统的功能模块主要包括数据库管理模块、网格划分与编码模块、轨迹网格化模块、轨迹数据挖掘模块与轨迹预测模块，轨迹数据、处理中间结果与最终结果存放于轨迹网格数据库中，整体的设计如下图所示：

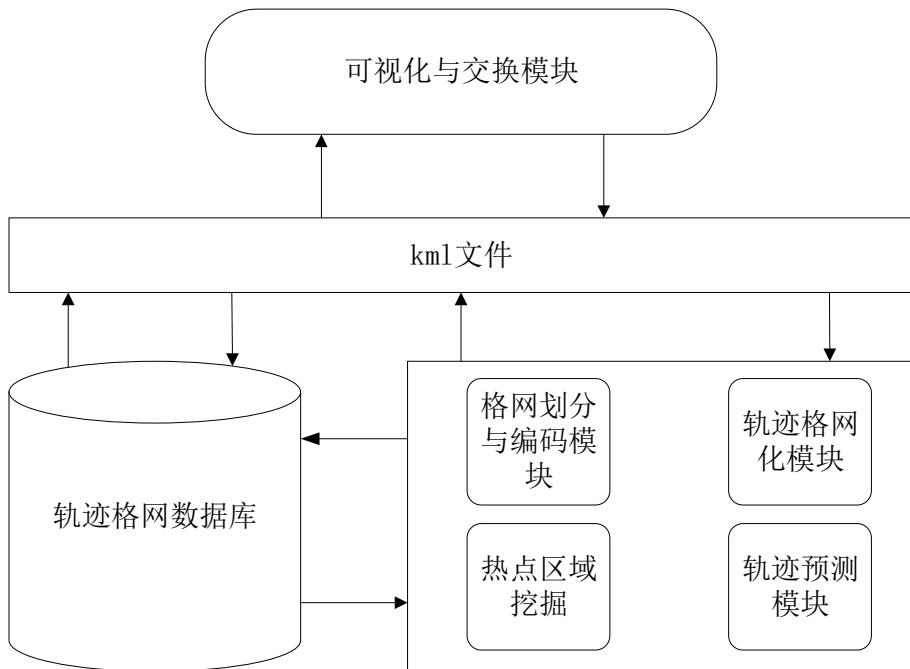


图 5-1 原型系统模块设计图

Figure 5-1 Prototype system module design diagram

该系统功能主要包括轨迹处理模块、轨迹分析模块与轨迹预测模块。在轨迹处理模块中，具备数据清洗、异常值剔除与数据压缩等数据预处理功能，为后期数据的网格化、热点区域挖掘等做好数据准备。轨迹网格化模块能够实现单轨迹与批量轨迹的网格化，批量轨迹网格化的结果同时存储到数据库。轨迹分析模块中包含了规律统计与热点区域挖掘两个模块，规律统计是计算与统计轨迹在每一个网格的进入时间、离开时间、停留时间、进入速度、离开速度等姿态参数。热点挖掘模块不仅能够挖掘出满足一定条件的热点区域，同时可计算各个热点区域的热度系数。轨迹预测模块实现船舶轨迹的实时预测，在挖掘出热点区域的基础上，导入当前轨迹，同时进行预测环境的建模，建模包括导入障碍物与构建政治代价环境，随后计算目标从当前轨迹奔向各目的地的概率，同时还能获取当前位置到目的地的路径。此外，本系统生成的相关可视化文件以 *kml* 文件格式进行存储，可借助 ArcGIS、QGIS 等地图软件进行打开展示。

系统主要功能如图 5-2 所示：

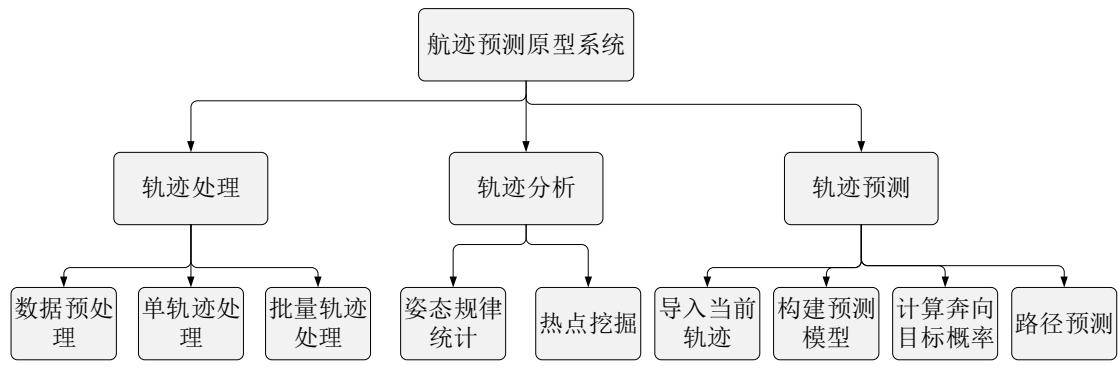


图 5-2 系统主要功能结构图

Figure 5-2 Main functional structure diagram of the system

5.3 数据库表设计 (Database Table Design)

数据库是一种可以存储和管理海量数据信息的技术。MySQL 数据库是一种关系型数据库，具有体积小、速度快、跨平台、开源免费、提供多种编程语言接口等优点，广泛运用于各种系统开发。因此本系统的数据库以 MySQL 数据库技术为基础进行开发。本系统的数据包括船舶原始轨迹数据、轨迹网格化结果数据、姿态统计数据、预测结果等本系统中所有涉及的数据均有对应数据表格进行管理，主要的数据库表结构如下：

船舶历史原始历史轨迹表记为 `trac_table`，用于存储与管理原始轨迹，船舶历史原始历史轨迹表结构如表 5-2 所示。

表 5-2 原始轨迹数据表
Table 5-2 Original trajectory data table

字段名	描述	类型	长度
<code>id</code>	数据项 id	Int4	4
<code>ship_id</code>	船舶 mmsi	Varchar	255
<code>trac_id</code>	轨迹 id	Int4	32
<code>lon</code>	地理经度	Float4	4
<code>lat</code>	地理纬度	Float4	4
<code>stamptime</code>	采样时间戳	Varchar	255
<code>speed</code>	对地速度	Float4	4
<code>direction</code>	航向	Float4	4

网格姿态统计表记为 `statistics_table`，用于统计船舶在每一个网格上的停留时间、平均穿过速度、进入与离开方向等姿态参数，以便于分析船舶的活动轨迹规律。其表结果如下表所示：

表 5-3 网格姿态统计表
Table 5-3 Grid posture statistics table

字段名	描述	类型	长度
id	数据项 id	Int4	4
ship_id	船舶 mmsi	Varchar	255
trac_id	轨迹 id	Int4	4
level	网格层级	Int4	4
seqnum	六边形网格编码	Varchar	255
entertime	进入网格时间戳	Int4	4
leavetime	离开网格时间戳	Int4	4
enterspeed	进入网格速度	Float4	4
leavespeed	离开网格速度	Float4	4
enterdirection	进入网格方向	Float4	4
leavedirection	离开网格方向	Float4	4
maxspeed	最大速度	Float4	4
minspeed	最小速度	Float4	4
avgspeed	平均速度	Float4	4
staytime	停留时间	Int4	4

热点区域挖掘结果表记为 hotarea_table，用于存储历史轨迹热点区域挖掘的结果，其表结构如下：

表 5-4 热点区域提取结果
Table 5-4 Hotspot area extraction results

字段名	描述	类型	长度
id	热点区域 id	Int4	4
include_grids	包含网格	Varchar	255
number	频次	Int4	4
time	总共停留时间	Int4	4
heat_coefficient	热度系数	Float4	4

预测结果表记为 predict_result_table，用于存储奔向各个目标点的概率，其表结构如下：

表 5-5 目标区域预测结果
Table 5-5 Target area prediction results

字段名	描述	类型	长度
id	热点区域 id	Int4	4
pro	概率大小	Float4	4
notes	备注	Varchar	255

5.4 系统展示 (System Presentation)

软件界面由五个部分组成，从左到右依次为数据库管理、网格与编码、轨迹预处理、轨迹分析与轨迹预测，软件主界面如下图所示：

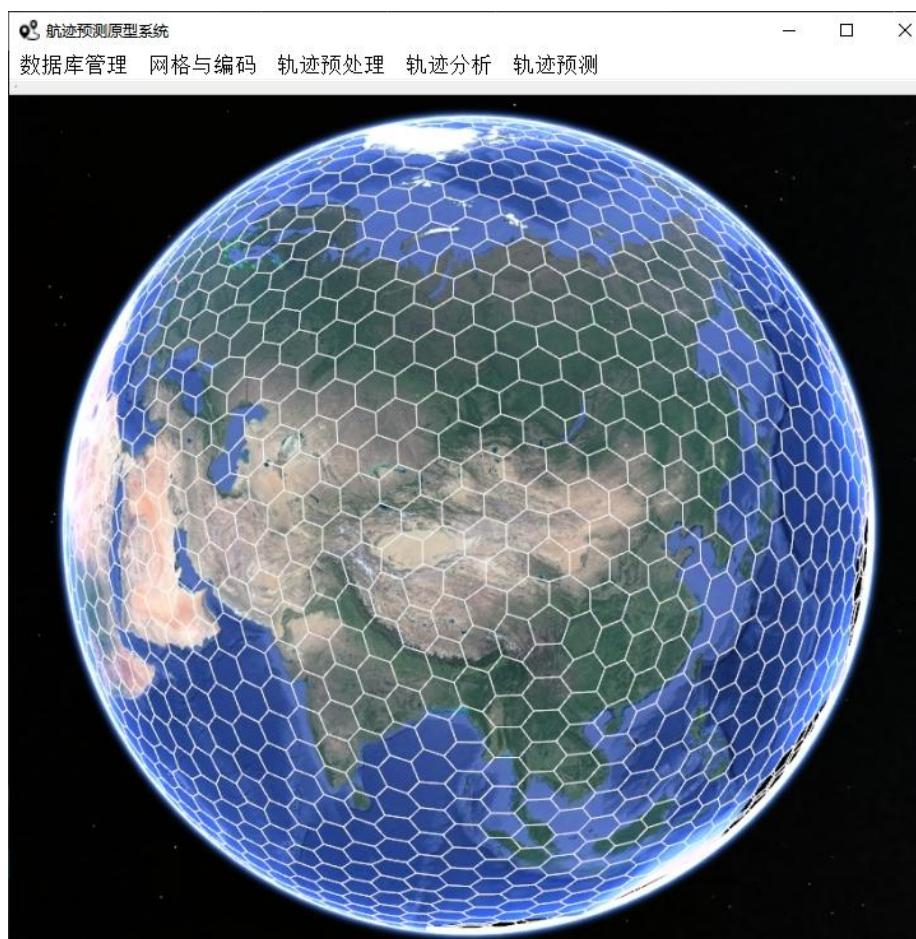


图 5-3 软件系统主界面图
Figure 5-3 Software system main interface diagram

轨迹预处理模块包括单轨迹网格化与批量轨迹网格化，批量轨迹网格化的结果同时存入数据库中相关表格，具体功能如图 5-4 所示：

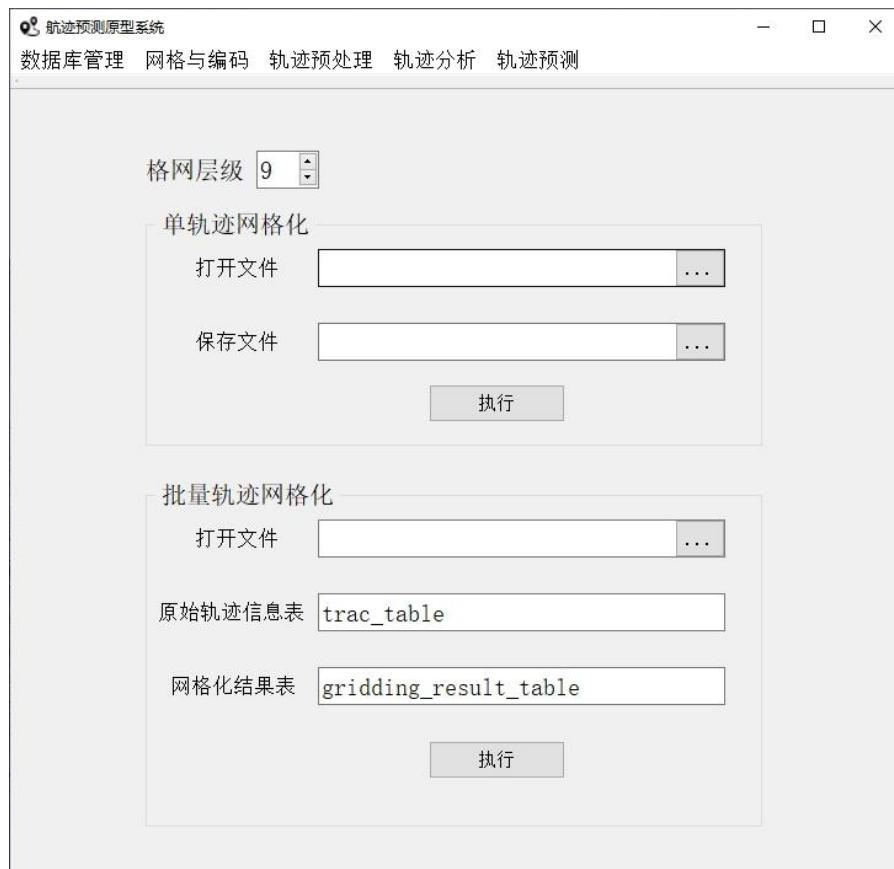


图 5-4 轨迹预处理功能

Figure 5-4 Trajectory preprocessing function

轨迹分析模块中包括轨迹姿态轨迹统计与热点挖掘功能，通过统计进入时间、离开时间、停留时间、速度等姿态参数进一步分析与挖掘船舶轨迹规律，姿态统计结果如图 5-5 所示，并以姿态规律统计结果为基础进行热点区域的挖掘，热点挖掘功能如图 5-6 所示：

id	ship_id	trac_ID	sequnum	enterTime	leaveTime	stayTime	enterSpeed	leaveSpeed	maxSpeed	minSpeed	avgSpeed	enterDirection	leaveDirection
1	205366000	0 33984327	-842105451	1509593169	429	-6277438562204...	4.11411314632442	4.11411314632442	4.01	4.06205657316221...	56.65783870260044...		
2	205366000	0 33984328	1509593169	1509593532	363	4.11411314632442	4.0603552477792	4.17	4.2675638295722	2.429910382126093	236.6578387131025	92.87392540719965	
3	205366000	0 33984325	1509593332	1509594037	505	4.2675638295722	4.33064381326356	4.272	4.2675638295722	2.429910382126093	272.873924357010...	52.40228163220625	
4	205366000	0 33988426	1509594037	1509594374	363	4.11411314632442	4.1328020810101	4.35482778327869	3.81	4.24381493214485	242.402281641562...	79.209163861978303	
5	205366000	0 33982523	1509594374	1509594915	541	3.60822985961341	2.79685660419738	3.60822985961341	2.598	2.091638627379	52.909123149212135		
6	205366000	0 33982524	1509594915	1509594981	66	2.79685660419738	3.77536859486597	2.73	2.38611259595266	259.2330215119888	102.94965562136362		
7	205366000	0 33986620	1509594981	1509595853	872	2.6902380163276	2.6902380163276	1.49	2.36570996422555	282.9946557167593	102.1059983475073		
8	205366000	0 33986621	1509595853	1509595899	114	1.09004111026524	1.370068149315815	1.370068149315815	1.02	1.300546292117	282.1099981309...	119.7582316597434	
9	205366000	0 34000717	1509598747	1509598754	1755	0.07662317255653	0.123770930946	0.237	0.07662317255653	0.1495011033254	299.7759246181207	83.5307764455076	
10	205366000	0 34000718	1509598754	1509599911	437	4.3111167244147	3.16904450637306	5.09	4.16904450637306	3.4000061530361	263.530770164567881	31.79886686742601	
11	205366000	0 34000719	1509599911	1509599917	326	5.06062336373628	5.01538229975061	5.06062336373628	5.01538229975061	5.0380028174343	211.768866868464...	31.80978475835913	
12	205366000	0 34000720	1509599917	1509599913	296	5.01538229975061	5.07950764120708	5.07950764120708	4.99	5.0447740747884	211.809784759348...	25.44777320665572	
13	205366000	0 34000721	1509599913	1509600071	256	5.227112730871997	5.76735575985966	5.227112730871997	5.490724153299891	205.947771208477...	25.56961544988777		
14	205366000	0 34000722	1509600071	1509600329	256	5.76735575985966	6.30756305390949	5.76735575985966	5.76735575985966	6.30756305390949	205.956616469773...	25.66619737205979	
15	205366000	0 34000723	1509600329	1509600567	238	6.30756305390949	6.80509129395214	6.80509129395214	6.55632717467582	205.9661973720881	25.7469177608536		
16	205366000	0 33986627	1509600521	1509600716	149	6.80509129395214	7.1778367741731	6.80509129395214	6.9614355684743	205.974691775986...	25.9800304542966		
17	205366000	0 33986628	1509600716	1509600952	236	7.17180587714731	6.0741071300430	7.25	6.0741071300430	6.6505715083561	205.9800304538527	27.06668701974761	
18	205366000	0 33986629	1509600952	1509601181	229	7.22601152878199	7.24680450657525	7.2	7.23342424467862	207.066870199259...	58.31769393358516		
19	205366000	0 34000726	1509601181	1509601223	42	7.19683453713764	7.19683456057526	7.19367153713764	7.195225304885645	28.3176939299803...	58.3204753479815		
20	205366000	0 33986630	1509601223	1509601364	141	7.19367153713764	7.1829985905428	7.19367153713764	7.1829985905428	7.18833506386596	238.32047353262...	58.32966212111978	
21	205366000	0 34000727	1509601364	1509601615	251	7.1629985905428	7.1639496459459	7.1629985905428	7.1639496459459	7.1639496459459	238.3298621211145...	58.3466216193275	
22	205366000	0 34000728	1509601615	1509601899	284	7.1639496459459	7.18616084631606	7.18616084631606	7.18616084631606	7.18616084631606	238.3466216182058	69.94486530735951	
23	205366000	0 34004625	1509601899	1509602059	160	7.17367830736787	7.2118057059256	7.2118057059256	7.2118057059256	7.17367830736788	249.944853073919...	69.9569500519426	
24	205366000	0 34004626	1509602059	1509602344	285	7.2118057059256	7.1673128901388	7.25	7.1673128901388	7.1895593203882	249.9569500548867	66.20640067391912	
25	205366000	0 34004627	1509602344	1509602519	175	7.2690441792028	7.29571003381079	7.29571003381079	7.29571003381079	7.29571003381079	246.206400673765...	66.2207842972035	
26	205366000	0 34008924	1509602519	1509602738	219	7.29571003381079	7.3291031622902	7.29571003381079	7.3124517501994	246.2207842285865	66.323883695818841		
27	205366000	0 34008925	1509602738	1509602963	225	7.3291031622902	7.2570931622902	7.32570931622902	7.3672570931622902	246.238836956828...	63.5408402913286		
28	205366000	0 34013022	1509602963	1509603017	154	7.33577607517804	7.33577607517804	7.33577607517804	7.34067760976420	243.5408403407184	63.53593715719829		
29	205366000	0 34013023	1509603017	1509603365	244	7.33537906195048	7.35581998884499	7.35581998884499	7.35581998884499	7.35581998884499	61.95277362787776		
30	205366000	0 34013024	1509603365	1509603566	201	7.30498398661399	7.28849656487649	7.30498398661399	7.28849656487649	7.28849656487649	241.952773631503...	61.96804247301906	
31	205366000	0 34017121	1509603566	1509603962	196	7.28949656487649	7.2744395680569	7.28949656487649	7.2744395680569	7.28949656487649	61.98289217865207		

图 5-5 姿态轨迹统计结果

Figure 5-5 Attitude trajectory statistical result



图 5-6 热点挖掘功能

Figure 5-6 Hotspot mining function

轨迹预测模块中包括政治代价建模、障碍物建模、目的地预测与路径预测等功能，对奔向目标区域的概率以及路径预测结果进行了展示，如图 5-7 所示：



图 5-7 轨迹预测结果

Figure 5-7 Trajectory prediction results

5.5 本章小结 (Summary of this Chapter)

本章首先从硬件配置与软件开发环境具体阐述了系统的相关开发环境；接着，介绍了系统功能，并从对各个子模块功能做了详尽说明；之后，具体介绍

了系统主要数据表的设计；最后，在软件界面系统中，展示了主界面、轨迹预处理、轨迹预测等主要功能，能够充分挖掘船舶历史轨迹的航行规律，可实现对船舶轨迹的实时预测，并同时展示轨迹预测结果。

6 总结与展望

6 Conclusion and Prospects

6.1 工作总结 (Conclusion)

在经济全球化的背景下，航运业得到飞速的发展，但随之而来也带来了船道阻塞、船舶碰撞、海盗劫持增多等一系列公共问题。因而有必要通过船舶轨迹预测技术来提前预知船舶的航行轨迹，实现对船舶的有效监督。目前，已有的船舶轨迹预测方法在精度、性能、通用性等方面取得显著进展，但仍然存在限制局部范围分析、历史轨迹不足预测误差大、长距离预测精度差等问题。在此背景下，本文基于全球六边形网格技术，研究基于球面空间下的航迹网格化技术与热点区域挖掘方案，并以此提出一种以目的地为导向的船舶轨迹预测方法。基于全球六边形网格的船舶轨迹预测方法在长远距离与历史轨迹不足相对于LSTM具有一定的优势，可为长远距离的轨迹预测研究提供新的思路。本文的主要工作如下：

(1) 设计了一种直接基于球面的航迹网格化算法，能够准确与快速地拟合球面矢量线。针对原始轨迹AIS数据的噪声大、易缺失和数据不一致等问题，通过异常值剔除、轨迹压缩与轨迹插值等步骤进行轨迹预处理，为后期的数据挖掘与数据分析提供良好的数据基础。在数据预处理的基础上，研究了一种基于全球六边形网格的航迹网格化算法。实验结果表明：该网格化算法具有一定的准确性与时效性，实现对任意弧段的快速网格化，网格化误差小于一个网格单元。

(2) 针对轨迹预测的目的地预测需求，研究了一种基于全球六边形网格的热点区域挖掘方法。在轨迹状态规律统计的基础上，提取方法基于时间与密度这两个维度进行挖掘精细化的热点区域，满足后续轨迹预测的要求。并设计相关实验来进行验证方法的可行性。实验结果表明：提取方法具有一定的可行性和准确性，结果基本为港口、海峡、交通枢纽等船舶频繁活动区域，实现了对船舶活动高频区域的精确识别。

(3) 针对已有船舶轨迹预测方法历史轨迹不足精度低、长远距离预测误差大等问题，提出了一种基于全球六边形网格的以目的地为导向的船舶轨迹预测模型，该模型充分考虑多种因素对船舶轨迹预测的影响，能够在一定程度上减少现有预测模型的问题，并且设计了定性与定量实验验证预测方法的可行性和准确性。实验结果表明：本文方法在处理海洋广域范围内的航迹预测问题时，能够提供较为精确的结果，预测轨迹与实际航行路径具有较高的吻合度。相对

于 LSTM 航迹预测方法，本文方法在长远距离的预测上更具有优越性，为长远距离的预测开辟新思路与提供新途径；针对历史轨迹不足的情况，本文方法相对于 LSTM 具有更高的预测精度。

(4) 在验证相关理论与技术可行的基础上，同时结合项目需求，基于 C++ 语言设计与开发了航迹预测原型系统，系统可实现编码转换、网格化、数据库管理、轨迹分析与热点区域挖掘等功能，能够充分挖掘船舶历史轨迹的姿态规律，满足船舶轨迹实时预测的要求。

本文创新点如下：

(1) 设计了一种球面六边形网格的轨迹快速网格化算法，为热点区域挖掘与轨迹预测提供技术支撑；

(2) 提出了一种顾及目的地及航道权等多重约束的船舶轨迹预测方法，为历史数据匮乏及长远距离预测的船舶轨迹预测提供新的思路与技术途径。

6.2 未来展望（Prospects）

虽然在航迹网格化、热点区域挖掘以及轨迹预测取得了一定的成果，但仍存在一些不足和有待继续研究的内容。因此，本文提出以下几点展望：

(1) 目前使用的网格系统为六边形网格，没有考虑到网格划分的多层次性与相似性，导致建立多尺度层次结构非常困难。下一阶段拟尝试将本文研究方法推广到三角形、四边形等网格系统，实现网格系统的多层次建模；

(2) 本文对热点区域挖掘与航迹预测的研究，仅使用了 AIS 数据中少部分信息，且并没有对不同类型船舶做各自的相关研究。在未来的研究中，期望可以通过数据挖掘和特征工程技术发现数据中更多的隐含信息来用于加强方法的预测能力，同时细化不同类型船舶的轨迹预测研究；

(3) 本文对于目标区域预测中历史轨迹、热点系数、目标区域距离与航向这四个影响因素的定权是基于大量实验的不断尝试，缺乏理论性的支撑与技术指导，建议在后续的工作参考相关理论模型进行更为科学的定权评估。

参考文献

- [1] 童晓冲. 空间信息剖分组织的全球离散格网理论与方法[J]. 测绘学报, 2011,40(4): 1.
- [2] CHENG C, TONG X, CHEN B, et al. A subdivision method to unify the existing latitude and longitude grids[J]. ISPRS international journal of geo-information, 2016,5(9): 161.
- [3] 程承旗. 空间信息剖分组织导论[M]. 科学出版社, 2012.
- [4] BAI M R, CHI L. Adaptive beamforming for acoustic echo cancellation using measured array models and subband filtering[J]. The Journal of the Acoustical Society of America, 2015,137: 2238.
- [5] MA T, ZHOU C, XIE Y, et al. A discrete square global grid system based on the parallels plane projection[J]. International Journal of Geographical Information Science, 2009,23(10): 1297-1313.
- [6] MERLAND R, CAUMON G, LÉVY B, et al. Voronoi grids conforming to 3D structural features[J]. Computational Geosciences, 2014,18(3-4): 373-383.
- [7] LUKATELA H. Hipparchus geopositioning model: An overview[C]//. Proceedings of the Eighth International Symposium on Computer-Assisted Cartography: ASPRS & ACSM, 1987: 87-96.
- [8] HUANG X, DING J, BEN J, et al. Unified Multi-structural Element Modeling Methods Using Icosahedral Discrete Global Grid Systems[J]. Mathematical Geosciences, 2023,56(2): 1-26.
- [9] SAHR K, WHITE D, KIMERLING A J. Geodesic Discrete Global Grid Systems[J]. Taylor & Francis Group, 2003(2).
- [10] de SOUSA L M, LEITÃO J P. HexASCII: A file format for cartographical hexagonal rasters[J]. Transactions in GIS, 2018,22(1): 217-232.
- [11] 童晓冲, 贲进, 汪滢. 利用数值投影变换构建全球六边形离散格网[J]. 测绘学报, 2013,42(02): 268-276.
- [12] 梁晓宇, 贲进, 周建彬, 等. 菱形三十面体六边形全球离散格网系统构建算法[J]. 地理与地理信息科学, 2023,39(02): 8-16.
- [13] SAHR K. Location coding on icosahedral aperture 3 hexagon discrete global grids[J]. Computers, Environment and Urban Systems, 2008,32(3): 174-187.
- [14] 王蕊, 贲进, 杜灵瑀, 等. 平面四孔六边形格网系统编码运算[J]. 测绘学报, 2018,47(07): 1018-1025.
- [15] 贲进, 李亚路, 周成虎, 等. 三孔六边形全球离散格网系统代数编码方法[J]. 中国科学: 地球科学, 2018,48(03): 340-352.
- [16] 赵学胜, 贲进, 孙文彬, 等. 地球剖分格网研究进展综述[J]. 测绘学报, 2016,45(S1): 1-14.

- [17] 邬群勇, 曾庆权, 张爱国. 一种全球离散格网系统框架下的室内空间网格数据模型[J]. 导航定位学报, 2020,8(02): 55-62.
- [18] ROBERTSON C, CHAUDHURI C, HOJATI M, et al. An integrated environmental analytics system (IDEAS) based on a DGGS[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020,162: 214-228.
- [19] LI M, MCGRATH H, STEFANAKIS E. Integration of heterogeneous terrain data into Discrete Global Grid Systems[J]. Cartography and Geographic Information Science, 2021,48(6): 546-564.
- [20] 周建彬, 贲进, 黄心海, 等. 广域六角格兵棋地图构建方法与机动推演应用[J]. 系统工程与电子技术, 2023,45(03): 769-776.
- [21] 梁启爽, 陈艺航, 贲进, 等. 遥感图像菱形三十面体六边形全球离散格网建模及存储[J]. 地球信息科学学报, 2023,25(12): 2361-2373.
- [22] LIU K, YUAN Z, XIN X, et al. Conflict detection method based on dynamic ship domain model for visualization of collision risk Hot-Spots[J]. Ocean Engineering, 2021,242: 110143.
- [23] BI S, XU R, LIU A, et al. Mining Taxi Pick-Up Hotspots Based on Grid Information Entropy Clustering Algorithm[J]. Journal of Advanced Transportation, 2021,2021: 5814879.
- [24] 杨乐, 包磊, 罗兵. 一种高效的船舶活动热点海域探测算法[J]. 计算机应用与软件, 2017,34(8): 5.
- [25] LORBEER B, KOSAREVA A, DEVA B, et al. A-BIRCH: Automatic Threshold Estimation for the BIRCH Clustering Algorithm[C]//. International Neural Network Society Conference on Big Data, 2017.
- [26] LIANG S. Research on the Method and Application of MapReduce in Mobile Track Big Data Mining[J]. Recent Advances in Electrical & Electronic Engineering (Formerly Recent Patents on Electrical & Electronic Engineering), 2021,14(1): 20-28.
- [27] WANG W, YANG J, MUNTZ R R. STING: A Statistical Information Grid Approach to Spatial Data Mining[C]//. Proceedings of the 23rd International Conference on Very Large Data Bases: Morgan Kaufmann Publishers Inc., 1997: 186-195.
- [28] 郁舒昊, 周辉, 叶春杨, 等. SDFA: 基于多特征融合的船舶轨迹聚类方法研究[J]. 计算机科学, 2022, 49(S1):256-260.
- [29] REYES G, LANZARINI L, HASPERUÉ W, et al. Proposal for a Pivot-Based Vehicle Trajectory Clustering Method[J]. Transportation Research Record, 2022,2676(4): 281-295.
- [30] LEE J G, HAN J, WHANG K Y. Trajectory Clustering: A Partition-and-Group Framework[C]//. Acm Sigmod International Conference on Management of Data, 2019.
- [31] MACQUEEN J. Some methods for classification and analysis of multivariate observations[J].

Proc. Symp. Math. Statist. and Probability, 5th, 1967,1.

- [32] AGRAWAL R, GEHRKE J E, GUNOPULOS D, et al. Automatic Subspace Clustering of High Dimensional Data for data mining applications[J]. Data Mining and Knowledge Discovery, 1998,27(2): 94-105.
- [33] ESTER M, KRIEGEL H P, SANDER J, et al. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise[C]//. National Conferences on Artificial Intelligence, 1999.
- [34] AGGARWAL C C, YU P S. Finding generalized projected clusters in high dimensional spaces[C]//. Proceedings of the 2000 ACM SIGMOD international conference on Management of data, 2000: 70-81.
- [35] LI Y, HAN J, YANG J. Clustering moving objects[J]. Department of Computer Science University of Illinois Urbana-Champaign, IL 61801 USA, 2004.
- [36] TU Y, LIU Y, LI Z. Online Segmentation Algorithm for Time Series Based on BIRCH Clustering Features[C]//. International Conference on Computational Intelligence & Security, 2010.
- [37] DENG C, SONG J, SUN R, et al. GRIDEN: An effective grid-based and density-based spatial clustering algorithm to support parallel computing[J]. Pattern Recognition Letters, 2018,109: 81-88.
- [38] IZAKIAN Z, SAADI MESGARI M, ABRAHAM A. Automated clustering of trajectory data using a particle swarm optimization[J]. Computers, Environment and Urban Systems, 2016,55: 55-65.
- [39] AGRAWAL K P, GARG S, SHARMA S, et al. Development and Validation of OPTICS Based Spatio-Temporal Clustering Technique[J]. Information Sciences, 2016,369: 388-401.
- [40] CAI M, LIANG Y. An Improved CURE Algorithm[C]//. IFIP TC12 ICIS, 2018.
- [41] LINJIANG Z, XIN Z, ZHAOHUI J, et al. MINING URBAN ATTRACTIVE AREAS USING TAXI TRAJECTORY DATA[J]. Computer Applications and Software, 2018.
- [42] 王培晓, 张恒才, 王海波, 等. ST-CFSFDP:快速搜索密度峰值的时空聚类算法[J]. 测绘学报, 2019,48(11): 11.
- [43] Y. W, J. R. Taxi Passenger Hot Spot Mining Based on a Refined K-Means++ Algorithm[J]. IEEE Access, 2021,9: 66587-66598.
- [44] WANG L, CHEN P, CHEN L, et al. Ship AIS Trajectory Clustering: An HDBSCAN-Based Approach[J]. 2021,9.
- [45] 胡昕源, 谢磊, 常吉亮, 等. 改进QuickBundles算法在船舶轨迹聚类中的应用[J]. 中国航海, 2023,46(03): 145-152.

- [46] XUE Q, XING Y, LU J. An integrated lane change prediction model incorporating traffic context based on trajectory data[J]. *Transportation Research Part C: Emerging Technologies*, 2022,141: 103738.
- [47] MILLEFIORI L M, BRACA P, BRYAN K, et al. Modeling vessel kinematics using a stochastic mean-reverting process for long-term prediction[J]. *IEEE Transactions on Aerospace & Electronic Systems*, 2017,52(5): 2313-2330.
- [48] R. W L, M. L, J. N, et al. Deep Learning-Powered Vessel Trajectory Prediction for Improving Smart Traffic Services in Maritime Internet of Things[J]. *IEEE Transactions on Network Science and Engineering*, 2022,9(5): 3080-3094.
- [49] HUANG Y, DU J, YANG Z, et al. A Survey on Trajectory-Prediction Methods for Autonomous Driving[J]. *IEEE Transactions on Intelligent Vehicles*, 2022,7: 652-674.
- [50] 戴礼灿, 刘欣, 张海瀛, 等. 基于卡尔曼滤波算法展开的飞行目标轨迹预测[J]. *系统工程与电子技术*, 2023,45(6): 1814-1820.
- [51] QIAN L, ZHENG Y, LI L, et al. A New Method of Inland Water Ship Trajectory Prediction Based on Long Short-Term Memory Network Optimized by Genetic Algorithm: Applied Sciences[Z]. 2022: 12.
- [52] PENG Qu,DING Zhi-ming, GUO Li min. Prediction of Trajectory Based on Markov Chains[J].*Computer Science*, 2010, 37(8): 189-193.
- [53] YE N, ZHANG Y, WANG R. Vehicle trajectory prediction based on Hidden Markov Model[J]. *KSII Transactions on Internet and Information Systems*, 2016,10: 3150-3170.
- [54] 乔少杰, 金琨, 韩楠, 等. 一种基于高斯混合模型的轨迹预测算法[J]. *软件学报*, 2015,26(5): 16.
- [55] 周巧娟. 基于贝叶斯网络的船舶航行安全性预测[J]. *舰船科学技术*, 2020(12): 34-36.
- [56] 乔少杰, 彭京, 李天瑞, 等. 基于CTBN的移动对象不确定轨迹预测算法[J]. *电子科技大学学报*, 2012,41(5): 5.
- [57] YANG G, JIE Q M, TAO N. Prediction of ship motion attitude based on BP network[J]. *2017 29th Chinese Control And Decision Conference (CCDC)*, 2017: 1596-1600.
- [58] K. M, D. K, J. P, et al. RNN-Based Path Prediction of Obstacle Vehicles With Deep Ensemble[J]. *IEEE Transactions on Vehicular Technology*, 2019,68(10): 10252-10256.
- [59] TANG H, YIN Y, SHEN H. A model for vessel trajectory prediction based on long short-term memory neural network[J]. *Journal of Marine Engineering and Technology: Proceedings of the Institute of Marine Engineering, Science and Technology*, 2022(3): 21.
- [60] ZHANG Z, NI G, XU Y. Ship Trajectory Prediction based on LSTM Neural Network[C]//, 2020.

- [61] Z. S, M. X, Q. P, et al. LSTM-based Flight Trajectory Prediction[C]//. 2018 International Joint Conference on Neural Networks (IJCNN), 2018: 1-8.
- [62] CHEN L, ZHOU Q, CAI Y, et al. CAE - GAN: A hybrid model for vehicle trajectory prediction[J]. IET Intelligent Transport Systems, 2022,16(12): 1682-1696.
- [63] XIN L, WANG P, CHAN C Y, et al. Intention-aware Long Horizon Trajectory Prediction of Surrounding Vehicles using Dual LSTM Networks[J]. 2019.
- [64] 任宇翔, 赵建森, 刘卫, 等. 基于AIS数据和LSTM网络的船舶航行动态预测[J]. 上海海事大学学报, 2019,40(03): 32-37.
- [65] LE Q I, ZHENG Z. Trajectory prediction of vessels based on data mining and machine learning[J]. Journal of Digital Information Management, 2016,14(1): 33-40.
- [66] 郭逸婕, 张君毅, 王鹏. 基于注意力机制的TCN-BiLSTM船舶轨迹预测[J]. 计算机测量与控制, 2024,32(01): 30-36.
- [67] BAKLACIOGLU T, CAVCAR M. Aero-propulsive modelling for climb and descent trajectory prediction of transport aircraft using genetic algorithms[J]. Aeronautical Journal - New Series-, 2014,118(1199): 65-79.
- [68] 谭伟, 陆百川, 黄美灵. 神经网络结合遗传算法用于航迹预测[J]. 重庆交通大学学报: 自然科学版, 2010,29(1): 4.
- [69] CHENG Z, WANG E, TANG Y, et al. Real-time Path Planning Strategy for UAV Based on Improved Particle Swarm Optimization[J]. Journal of Computers, 2014(1).
- [70] EBERHART R C, KENNEDY J. A New Optimizer Using Particle Swarm Theory[C]//, 1995.
- [71] C. G P, A. D, E. G. Extended Kalman Filter for State Estimation and Trajectory Prediction of a Moving Object Detected by an Unmanned Aerial Vehicle[C]//. 2007 American Control Conference, 2007: 1805-1810.
- [72] WU H, WANG L, ZHENG S, et al. Crossing-Road Pedestrian Trajectory Prediction Based on Intention and Behavior Identification[C]//, 2020.
- [73] QIAO S, HAN N, ZHU X, et al. A Dynamic Trajectory Prediction Algorithm Based on Kalman Filter[J]. ACTA ELECTRONICA SINICA, 2018,46(2): 418-423.
- [74] 徐义晗. 基于无迹卡尔曼滤波的移动目标跟踪算法[J]. 火力与指挥控制, 2020,45(12): 4.
- [75] 乔少杰, 韩楠, 朱新文, 等. 基于卡尔曼滤波的动态轨迹预测算法[J]. 电子学报, 2018,46(02): 418-423.
- [76] STOUGH T, CRESSIE N, KANG E L, et al. Spatial analysis and visualization of global data on multi-resolution hexagonal grids[J]. Japanese Journal of Statistics and Data Science, 2020,3(22).
- [77] RICHESON D S. Euler's Gem: The Polyhedron Formula and the Birth of Topology[J].

Mathematical Intelligencer, 2010,32(11): 56-57.

- [78] GOODCHILD M F, SHIREN Y. A hierarchical spatial data structure for global geographic information systems[J]. CVGIP: Graphical Models and Image Processing, 1992.
- [79] FERKISS V, FULLER R B, APPLEWHITE E J. Synergetics: Explorations in the Geometry of Thinking[J]. Technology and Culture, 1976,17(1): 104.
- [80] ALBORZI H, SAMET H. Augmenting SAND with a spherical data model[C]//. Proceedings of the First International Conference on Discrete Global Grids, Santa Barbara, CA, USA, 2000.
- [81] 杨帆, 贲进, 潘思宇, 等. 基于Fuller投影的六边形球面离散格网生成算法[J]. 测绘与空间地理信息, 2014,37(8): 3.
- [82] 贲进, 童晓冲, 张永生, 等. 对施奈德等积多面体投影的研究[J]. 武汉大学学报: 信息科学版, 2006,31(10): 4.
- [83] ALAHI A, GOEL K, RAMANATHAN V, et al. Social lstm: Human trajectory prediction in crowded spaces[C]//. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 961-971.
- [84] 王涵. 基于AIS数据和改进长短期记忆网络的船舶航迹预测[D]. 武汉大学, 2020.
- [85] 郭保坤. AIS在当代航海技术中的应用研究[J]. 中国水运, 2024(02): 83-86.
- [86] DOUGLAS D H. Algorithms for the reduction of the number of points required to represent a line or its caricature[J]. Canadian Cartographer, 1973,10.
- [87] HANSUDDHISUNTORN K, HORANONT T. Improvement of TD-TR Algorithm for Simplifying GPS Trajectory Data[C]//, 2019.
- [88] SCHMID F, RICHTER K F, LAUBE P. Semantic Trajectory Compression[C]//. Advances in Spatial and Temporal Databases --- 11th International Symposium, SSTD 2009, 2009.
- [89] LI S, ZHANG K, YIN H, et al. ROPW: An Online Trajectory Compression Algorithm[C]//, Cham: Springer International Publishing, 2021: 16-28.
- [90] MCCREA P G, BAKER P W. On Digital Differential Analyzer (DDA) Circle Generation for Computer Graphics[J]. IEEE Transactions on Computers, 2006,C-24(11): 1109-1110.
- [91] K. S R. Three early algorithms[J]. IEEE Annals of the History of Computing, 2002,24(4): 10-13.
- [92] J. E B. Algorithm for computer control of a digital plotter[J]. IBM Systems Journal, 1965,4(1): 25-30.
- [93] FREEMAN H. Algorithm for Generating a Digital Straight Line on a Triangular Grid[J]. IEEE Transactions on Computers, 1979,C-28(2): 150-152.
- [94] 刘勇奎, 沈红, 周晓敏. 六角网格上的直线生成算法[J]. 计算机工程与设计, 2001(01): 60-64.

- [95] 杜灵瑀, 贲进, 马秋禾, 等. 基于弱对偶的平面三角形格网离散线转化生成算法[J]. 武汉大学学报(信息科学版), 2020,45(01): 105-110.
- [96] 于文率, 童晓冲, 贲进, 等. 全球六边形离散格网的矢量线数据绘制精度控制[J]. 地球信息科学学报, 2015,17(07): 804-809.
- [97] 杜灵瑀. 矢量线六边形全球离散格网表达的理论与方法[D]. 战略支援部队信息工程大学, 2019.

学位论文数据集

关键词*	密级*	中图分类号*	UDC	论文资助
船舶轨迹预测；热点区域挖掘；AIS 数据；全球六边形网格；轨迹网格化	公开	P237	528	XXXX 网格 XX 生成技术项目
学位授予单位名称*	学位授予单位代码* *		学位类别*	学位级别*
中国矿业大学	10290		资源与环境	硕士
论文题名*		并列题名		论文语种*
基于全球六边形网格的船舶轨迹预测		Ship Trajectory Prediction Based on Global Hexagonal Grid		中文
作者姓名*	吴名荣		学号*	TS21160183P31
培养单位名称*	培养单位代码*		培养单位地址	邮编
中国矿业大学	10290		江苏省徐州市	221116
学科专业*	研究方向*		学制*	学位授予年*
测绘工程	地理信息工程及应 用		三年	2024
论文提交日期*		2024 年 5 月		
导师姓名*	余接情		职称*	教授
评阅人		答辩委员会主席*		答辩委员会成员
盲审		张书毕		范洪冬、王中元、李增科、尹鹏程
电子版论文提交格式 文本 (<input checked="" type="checkbox"/>) 图像 (<input type="checkbox"/>) 视频 (<input type="checkbox"/>) 音频 (<input type="checkbox"/>) 多媒体 (<input type="checkbox"/>) 其他 (<input type="checkbox"/>)				
推荐格式: application/msword; application/pdf				
电子版论文出版(发布)者	电子版论文出版(发布)地			权限声明
论文总页数*		79		
注: 共 33 项, 其中带*为必填数据, 共 22 项。				