

# 基于多源水运大数据的 货运 OD 特征挖掘算法研究

陈冬, 陈奕超, 刘秀彩

(华设计集团股份有限公司, 江苏 南京 210014)

**摘要:** 针对传统水路货运 OD 调查样本少、成本高等问题, 本次融合 AIS 数据、报港数据、港口地理信息数据等多源水运数据, 建立了基于大数据技术的水路货运 OD 特征挖掘算法, 实现了对水上货物运输总量及主要货种运输流量、流向的分析, 对辅助管理部门制定经济发展规划、水运基础设施规划具有重要意义。

**关键词:** AIS 数据; 水路货运 OD; 挖掘算法

中图分类号: U695.2 文献标识码: A 文章编号: 1006—7973 (2021) 09—0094—03

## 1 概述

水运是综合运输体系中的一种重要运输方式, 具有运量大、能耗低、污染少等特点。水路货运的起讫点 (Origin and Destination, OD) 不仅清晰地反映了地区经济的发展情况以及地区经济之间的直接联系, 而且反映了水运基础设施 (包括港口、航道、船闸、锚地等) 的利用情况, 因此及时准确地获取水路货运 OD, 对于辅助交通运输管理部门制定经济发展规划、港口航道基础设施规划具有重要意义。

传统的水路货运 OD 主要通过人工调研的方式获得, 通过人工在主要航道上记录船舶及货物的运输情况, 同

时辅以交通统计年鉴资料、航道观测点的交通量观测资料、港口吞吐量统计资料, 可基本分析出船舶和货物的流量、流向规律。由于传统方法仅为某段时间的调研数据, 同时需要花费较大的人力物力, 局限性较大, 因此本次基于 AIS 数据、报港数据、港口地理信息数据等多源水运大数据, 提出了全新的水路货运 OD 计算方法。

## 2 多源水运数据准备

### 2.1 AIS 数据

AIS 数据来自于船舶自动识别系统, AIS 报文信号经过解析可以解读为静态信息、动态信息、航次信息、

建造工艺设计水平, 降低建造成本、推广船舶建造工艺三维数字化设计具有一定的参考价值 and 意义。

## 参考文献:

[1] 杨润党, 范秀敏, 王文荣. 虚拟仿真技术在船舶建造过程中的研究与应用 [J]. 舰船科学技术 2008 (01): 11-12

[2] 鲁雄飞, 邓中军, 船舶智能制造技术的应用及发展 [J]. 船舶物资与市场 2020 (06): 1-2.

[3] 张志航, 程广伟, 门清毅等. 基于 3DVIA Composer 的变速箱主动轴三维可视化装配工艺研究 [J]. 现代信息技术 2019 (08): 144-147

[4] 沈刚, 官雄明, 冯卫星等. 基于 Catia Composer 的三维模型 Web 浏览平台开发 [J]. 软件工程 2019 (04): 37-39

[5] 杨羲昊, 赵文军, 马涛等. 基于 3DVIA Composer 的大型结构件装焊三维可视化表达 [J]. 新技术新工艺 2017 (06): 72-74

[6] 牛鹏辉, 刘救世, 李大磊. 基于 3DVIA Composer 3D 动画在机床夹具 CAI 中的应用研究 [J]. 科教导刊 (下旬刊) 2017 (12): 32-33.

[7] 马庆恒, 万永丽, 姜贵中等. 基于 3DVIA Composer 的制药外包装线后端智能化生成三维仿真动画制作 [J]. 机电产品开发与创新 2020 (11) 68-70.

**基金项目:** 该论文受科研项目 JCKY2018206A001 支持。

安全信息等四种不同类型的信息。其中，静态信息包括船舶 MMSI 编号、船名、型宽、船上定位天线的位置等；动态信息包括 UTC 时间、对地航向、对地航速、船首向、航行状态、倾角、转向率等；航次相关信息包括船舶吃水深度、船舶运载的危险货物类型、目的港口与到达时间、航线计划等；安全信息主要为与船舶航行相关的航权消息等。

AIS 数据主要用以判断船舶和货物的流向，本次以 2019 年某段时间内的江苏省 AIS 数据为例开展研究，所需要用到的 AIS 数据信息主要有：MMSI（船舶呼号）、NAME1（船舶名称）、L（船舶所在经度）、B（船舶所在纬度）、TYPE1（船舶类型）；STAT（船舶状态）、CTIME（测定该位置时的时间）。

MMSI	NAME1	ANAME	IMO	L	B	CALL1	TYPE1	WIDTH1	HEIGHT
205700000	LA STELLA	(Null)	9748368	9376666667	333333333	ONIS	70-货轮	199	33
205749000	MINERAL GE	(Null)	9435076	121.311129	31.59502	ONJL	70-货轮	292	45
209539000	WARNOV C	(Null)	9437256	0783333333	8666666667	5BFN4	70-货轮	139	23
210698000	NICOLAOS A	(Null)	9232151	2626666667	1666666667	C4KE2	70-货轮	190	32
219024344	CMA CGM G	(Null)	9718117	121.57785	0666666667	OXZO2	70-货轮	299	48
219206000	BELGIA SEA	(Null)	9816842	9383333333	32.23345	OXBT2	70-货轮	238	34
228348900	CMA CGM N	(Null)	9299812	6068333333	31.22365	FMHQ	70-货轮	349	43
229623000	ANDREAS P	(Null)	9664952	0993333333	3333333333	9HA3460	70-货轮	225	32
229657000	SAN PEDRO	(Null)	9625281	122.16583	31.185245	9HA3492	70-货轮	172	27
229788000	KIRAN ADRI	(Null)	9653185	1801666667	8666666667	9HA3610	70-货轮	200	32
232003534	BERGE RISH	(Null)	9713222	2538333333	8666666667	MADD2	70-货轮	180	30
235089912	VECTIS OSPH	(Null)	9594315	121.37485	31.525465	2FBR5	70-货轮	110	18

图 1 船舶 AIS 数据（部分数据）

## 2.2 报港数据

报港数据记录了各港口船舶的装卸信息，主要用于在货物流向的基础上进一步确定货物的流量，报港数据信息具体包括船舶的中文船名、海船内河船标志、船舶种类、进（出）港标志、进（出）港时间、查验机构、海事机构、签证方式、总吨、载重吨、实载货量、载/卸货量、前吃水以及后吃水。

## 2.3 港口基础设施地理信息数据

水路货运 OD 以港口为起讫点，因此，本次应用 GIS 系统采集了江苏省沿江沿海港口各作业区的地理信息数据，主要包括作业区名称、位置坐标、所属港区、所属港口、所属区县、所属城市、所属省等信息。

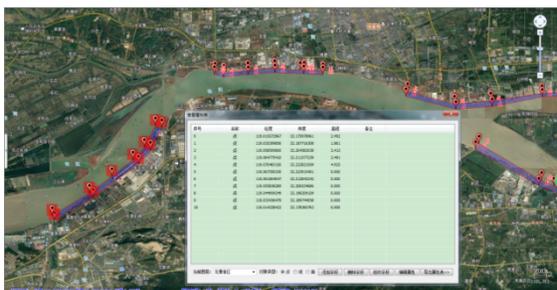


图 2 港口基础设施地理信息数据

## 3 货运 OD 特征挖掘算法

### 3.1 数据预处理

AIS 等运行数据在传输采集的过程中易出现信息缺失、错误、重复、异常等状况，会对大数据的挖掘分析造成干扰，影响分析效果，因此需要对数据进行数据清理、数据集成、数据变换、数据归约等预处理工作，本次主要对 AIS 船舶运行数据进行数据清理操作。

表 1 AIS 数据预处理

噪声种类	数据特征	清洗方法
信号漂移	与前后两个轨迹点的距离都超过距离阈值	剔除
位置异常	前后其中一个轨迹点的距离超过距离阈值	分割
编码错误	与前后两个轨迹点的时间差都大于时间阈值	剔除
多次往返	与前后其中一个轨迹点的时间差超过时间阈值	分割

### 3.2 货运 OD 特征挖掘算法

对水路货运 OD 特征挖掘的关键在于通过船舶 AIS 轨迹数据准确地判断船舶停靠点。本次利用聚类分析方法，将行驶速度小于 0.5km/h，行驶距离大于 10km，停泊时间大于 3h 的停靠点初步判断为水路货运 OD 点；结合港口地理信息数据，应用射线法，进一步检验确定船舶的 OD 位置，同时结合报港数据确定每次货运 OD 的货种及运量。具体货运 OD 特征挖掘技术路线如下图所示。

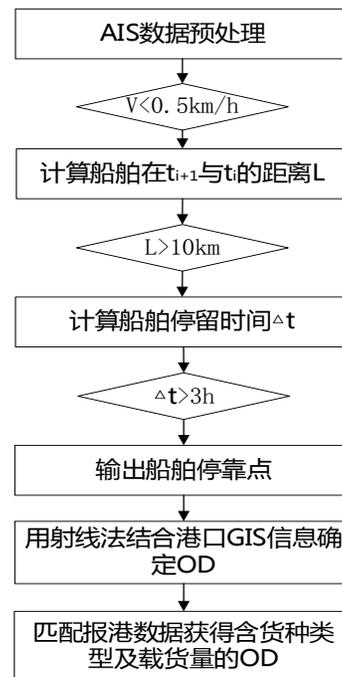


图 3 货运 OD 特征挖掘算法

### 3.3 货运 OD 特征挖掘计算结果

图 4 - 图 7 为本次江苏省内水路货运 OD 挖掘分析

结果，可以看出该时段内江苏省内沿江沿海船舶运输主要在长江沿线城市间进行，泰州市、苏州市、南通市等城市之间货运联系强度较大；同时研究时段内，干、散货运输区县间货运量较大的运输线路主要为太仓市—靖江市、太仓市—镇江市区、太仓市—南通市区、南京市—太仓市区；集装箱运输区县间货运量较大的运输线路主要为海门市—南京市区、南京市区—常熟市、镇江市—海门市。

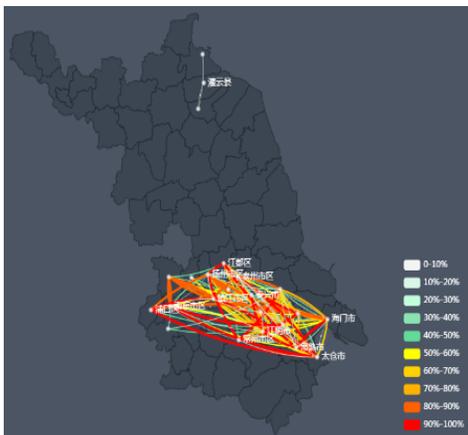


图4 水路货运 OD 总体特征

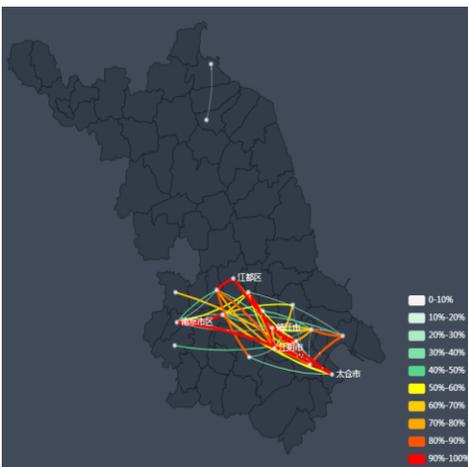


图5 散货运输 OD 特征

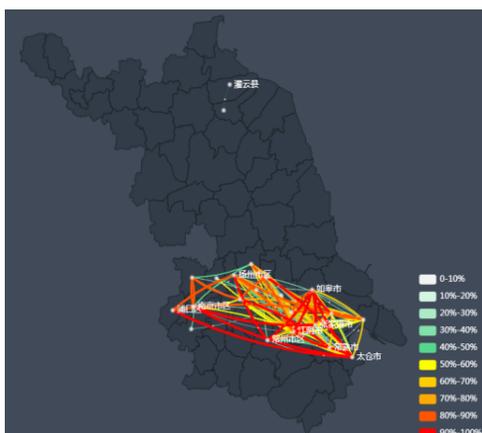


图6 干货运输 OD 特征

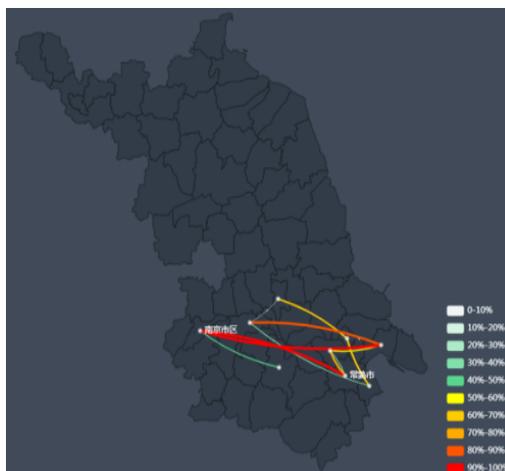


图7 集装箱运输 OD 特征

#### 4 结语

本次利用 AIS 数据、报港数据、港口地理信息数据等多源数据，提出了基于大数据方法的水路货运 OD 计算框架，实现了对水上货物运输总量及主要货种运输的流量、流向分析。由于原始数据信息内容的限制，目前船舶货运种类只能区分出散货、集装箱等货运大类，因此，本计算方法在船舶货运种类的细分上仍有进一步提升的需要，未来可以考虑通过进一步融合船舶身份信息、港口类别信息的方法，提升各水运货种 OD 计算的准确性。

#### 参考文献：

- [1] 姜鹏. AIS 船舶时空大数据分析和可视化技术研究 [D]. 大连：大连海事大学，2020.
- [2] 陈龙彪，张大庆，李石坚等. 基于海事大数据的港口感知计算 [J]. 地球信息科学, 2016, 18(11): 1485-1493.