

# 《软件需求工程》选题报告

软硕232 丁浩宸 2023213911

项目名称

基于AIS数据的分布式船舶轨迹航道交通流特征算法

项目所属的研究领域/工程领域介绍

项目属于分布式大数据计算与交通领域的交叉应用。

- 分布式计算与分布式架构

分布式计算是一种计算方法，和集中式计算是相对的。随着计算技术的发展，有些应用需要非常巨大的计算能力才能完成，如果采用集中式计算，需要耗费相当长的时间来完成。分布式计算将该应用分解成许多小的部分，分配给多台计算机进行处理。这样可以节约整体计算时间，大大提高计算效率。

分布式架构是分布式计算技术的应用和工具。本项目使用的分布式架构为基于Apache Hadoop的Apache Spark，它使用HDFS作为文件系统，容错率高、吞吐量大，非常适合大规模数据集上的应用；使用MapReduce编程模型进行并行计算，允许编程人员在不单独进行分布式并行编程的情况下，直接将自己的程序运行在分布式系统上。

- AIS：

船舶自动识别系统（AIS）系统，是一种实时记录船舶经纬度轨迹的系统。根据国际海事组织的《海上人命安全公约》（SOLAS）规定，所有在国际水域航行的总吨位在300吨以上的船舶、不在国际水域航行但总吨位在500吨或以上的船舶以及任何大小的客船都必须安装AIS转发器。

2010年前后，随着卫星事业的发展，卫星自动识别系统（S-AIS）开始兴起，有取代陆基自动识别系统（T-AIS）的趋势。近两年，利用AIS系统提供的数据进行交通管理数据计算、交通管理统筹规划，逐渐成为了水路运输中的热门研究方向。

- 交通流特征计算

交通流特征，指的是交通体系中人流、车流在不同条件下变化规律及其相互关系的定量或定性描述的总和。在河道中，我们常用航道密度K、船舶运行速度V、航道流量Q这三个最重要的参数加以描述。

在工程上，交通流特征计算的重点在于实现路网匹配算法，即**根据AIS系统提供的经纬度数据实时计算出船舶在航道中的位置**。根据路网匹配算法的计算结果，可以很方便地求取所需的各种交通流特征，并可以根据需要选择进行统计学意义上的分析，例如计算各个参数的均值，方差，中位数，四分位数，众数，图线斜率与峰值等等；另外还可以进行一些相关性检验。

问题描述：研究/工程问题是什么？解决该问题的意义？难度如何？

本项目研究的问题是：

- 根据航道的经纬度数据构建航道模型

- 实现路网匹配算法，根据船舶的AIS经纬度数据，将船舶实时匹配到航道中的具体位置
- 根据匹配结果进行进一步的交通流特征求解。

解决该问题的意义：

- 由于我国幅员辽阔、河流众多，因此需要动态掌握各条河流上的船舶运行情况，方便动态规划各种交通管理。
- 难度：
  - 由于河运有固定的航道（即河流），而海运不存在这方面问题，因此最新的相关进展对固定航道的路网匹配算法并不能提供太多支持。
  - 路网匹配算法有两大难点：
    - AIS系统提供的经纬度坐标数据量极大
    - 计算结果的时效性需求高，需要保证计算效率达到准实时性，即**每秒能处理10万条AIS数据**。因此需要在分布式架构上进行路网匹配算法构建。

该领域的研究文献/竞品数量？主要出处？主要相关工作有哪些？主流方法和最新进展如何？

本领域的相关研究文献或竞品数量不多，主要出处来自于近三年来中国、韩国等国家的海事相关大学的研究著作，例如：

- Kim Y, Jeongseok L, Pititto A, et al. Maritime traffic evaluation using spatial-temporal density analysis based on big ais data[J/OL]. Applied Sciences, 2022, 12: 11246. DOI: 10.3390/app122111246.
- Onyango S, Kim K I, Yoo S. A quasi-intelligent maritime route extraction from ais data[J/OL]. Sensors, 2022, 22: 8639. DOI: 10.3390/s22228639.
- Kim J. Semi-continuous spatial statistical analysis using ais data for vessel traffic flow characteristics in fairway[J/OL]. Journal of Marine Science and Engineering, 2021, 9: 378. DOI: 10.3390/jmse9040378.40

AIS相关讨论是近年出现的新领域，以上最新研究基本解决了**海运**相关的AIS数据提取问题，但对河运层面的讨论目前几乎为零。

该领域的核心技术需求是什么？有哪些相关产品？功能、可用性、安全性、收费情况如何？

本领域的核心技术需求为实时性，即**每秒能处理10万条AIS数据**。因此需要借助Apache Kafka、Apache Spark等开源架构进行分布式计算，同时使用多个worker机处理数据计算与结果上传等任务，以提升效率。这些开源架构能够自动完成局域网内大批量数据的传输和分配计算工作，因此可用性很高、安全性较强（限定在局域网内）。不收费。

你要从哪些角度出发来解决这个问题？你的具体研发内容包括什么？

- 完成航道建模
- 对船舶AIS数据进行预处理，并实现路网匹配算法
- 将路网匹配算法和计算结果导出代码移植到Apache Spark分布式架构上进行部署

- 将计算结果展示在前端界面，并在前端完成各种交通流特征计算

预期成果是什么？如何验证？

- 一个能够展示路网匹配算法计算结果的数据库
- 一个能够进行路网匹配算法后端（可用Apache Spark分布式部署界面进行展示）
- 一个能够简洁计算并展示各种交通流特征的前端界面

使用前端界面的交互进行计算，将计算结果与实际情况做对比进行验证。