**技 术 交 底 书**

所属部门：

发 明 人：

联系电话：

----------------------------------------------------------------

1. **发明名称**

一种基于机器学习的船舶航行异常和流程异常检测方法

2、**技术领域**

本发明涉及一种船舶航行异常和流程异常检测方法，其主要通过借助机器学习领域的相关算法，分析船舶的历史和实时轨迹数据，对船舶进行航行异常和流程异常检测。

3、**背景技术**

随着在传感、通信、存储和计算领域的进步，人们在网络和实体空间活动时留下的记录正以前所未有的速度累积。这些记录的规模和多样性为人们在不同环境下了解社会行为和社会动态提供了新的机会，尤其在公共安全、城市规划、交通管理等方面。

船舶自动识别系统（Automatic Identification System,简称AIS系统）由岸基（基站）设施和船载设备共同组成，是一种新型的集网络技术、现代通讯技术、计算机技术、电子信息显示技术为一体的数字助航系统和设备。收集和分析这些AIS信息，有利于人们揭示城市的动态变化和人类行为等信息。比如说发现交通模式、城市动态、行车路线等等。

船舶在航行过程中往往需要遵循特定的航线或者不同工程中的船舶需要在特定起始和终止区域之间航行。但是出于利益问题，某些船只会选择违规行驶或者不按照制定流程完成作业任务，这些违规操作会给工程质量和委托方带来损失。因此检测这些行为是确保工程质量和提高工程效率的关键。当前这些违规行为是委托监理单位安排经验丰富的监理人根据视频监控以及相关数据报表记录进行的，但是不仅人力成本大，而且时间上也十分低效。并且监理人的工作质量本身也不可控，因此准确地判断船舶异常航行行为，既有利于维护船舶公司的整体声誉，有效的约束船只负责人的行为，同时也能保护工程方合法权益，提高工程质量和效率，具有十分重要的意义。目前关于船舶航行异常问题急需一种可靠可行的方法。

近年来学术界对异常轨迹检测研究的热度越来越高，越来越多的检测方法被提出，凸显出其实用价值。如构建集群来检测连续轨迹流，对一个全局的异常事件进行检测，使用基于距离和密度特征两种形式证据的组合模型等方法。但是已有方法存在无法在个体层面上发现异常行为、对轨迹距离范围有特定要求等问题。而且由于轨迹数据的隐私性和复杂性等问题，目前可行的船舶轨迹异常检测方法仍然相对匮乏。

4、**发明目的**

基于以上所描述的背景和现状，本发明的目的在于提出一种切实可行的基于机器学习的船舶航行异常和工作流程异常检测方法，以一定程度上弥补当前该领域的理论空缺。

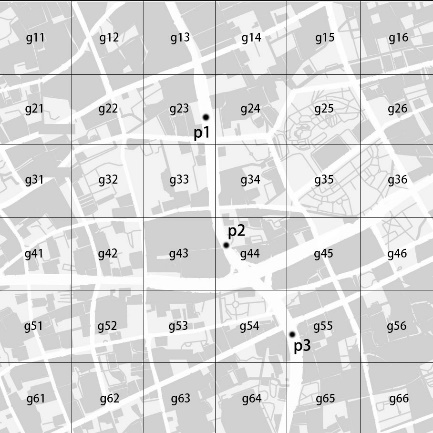
5、**发明的内容**

本发明要解决的技术问题在于，提出一种对船舶轨迹数据的有效建模方式，并在其基础之上通过应用机器学习的相关算法，来检测船舶航行异常和工作流程异常。

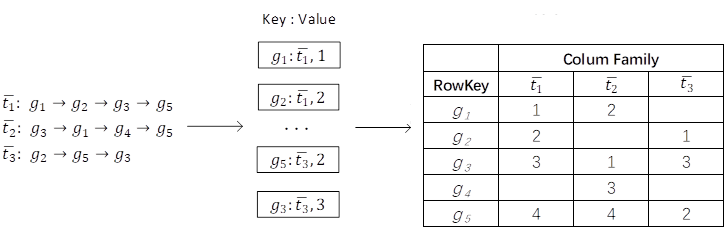
本发明的技术方案，其核心思想在于：1）将船舶轨迹原始历史数据进行分析，提取其中关键特征，并通过网格化、倒排索引等手段对提取出的轨迹特征进行存储及索引，然后根据异常点易于隔离特性对轨迹进行分类从而检测船舶航行异常。2）通过将完整工作流程划分为事件流，依据AIS坐标相对位置和预设区域的关系将轨迹数据处理划分，识别其中关键事件，检测流程从而进行异常分析。

首先对本发明的船舶轨迹倒排索引数据库的建立进行简要说明。由于AIS点坐标是以经纬度的形式进行定位的，而经度和纬度的值都属于连续域。这里定义一个映射函数:

函数将属于连续域的AIS坐标映射到网格矩阵内的一个网格中，网格矩阵是将地理区域分解为若干个网格得到的，表示矩阵中第行、第列的网格。将轨迹中的所有点全部通过映射后得到网格轨迹,（对任意）。如在图中，，。



倒排索引是一种索引方法，被用来存储在全文搜索下某个单词在一个文档或者一组文档中的存储位置的映射。它是文档检索系统中最常用的数据结构。本发明将轨迹作为文档，单元格作为单词，构建对每个单元格的索引。将单元格作为行键，轨迹编号作为列修饰符，单元格位置作为该列的值写入数据库。该过程如下图所示：



在建立倒排索引之后，我们若要查找通过某个网格的所有轨迹，则直接将对应行的所有列返回即可得到结果。

然后对本发明的船舶工作流程检测进行说明。对于有具体规定起始工作区域和目标工作区域的工程，船舶的轨迹点可以用一些状态进行划分，如：S1(在起始区域内),S2（在目标区域内）,S3（在起始区域和目标区域以外的区域）。那么船舶轨迹就可以转化成状态序列进而可以从状态序列中提取出标志性事件。如：E1（进入起始区域）,E2（出起始区域）,E3（进入目标区域）,E4（出目标区域）。

6、**发明的效果**

本发明提出了一种基于船舶轨迹倒排索引和工作流程检测的数据建模方法，极大简化了原本复杂的船舶轨迹数据表示，也使得复杂轨迹点数据之间隐含的特征得以便于分析。不同于传统全局时间异常的定义和检测，本发明通过对单个坐标点的位置属性进行分析实现对单个点的实时异常检测。从而可以将异常发生的具体时间更精确的识别出来，便于对异常信息进行提取和记录。同时针对明确具体起始区域和目标区域的任务，对船舶轨迹转化成状态序列并进行事件检测，进而提取流程异常记录。

1. **附图**

图1为坐标点异常检测的整体流程

图2为坐标点异常检测算法

图3为异常坐标点检测实例

**8、具体实施方式**

本发明中的船舶航行异常检测方法为一般性的理论框架，

下面结合具体实例来对方法进行详细说明。

本发明实施方案将分为坐标点异常检测和流程异常检测两个部分，首先对坐标点异常检测进行说明，其按照以下步骤进行：

步骤1：离线处理阶段操作的主要目的是从原始的AIS记录当中提取出有效的船舶轨迹。需要使用的原始AIS信息中数据为：坐标记录时间，坐标点经度，坐标点纬度。对原始数据进行按照船只编号进行分组并进行降噪处理，得到对应每条船只的历史轨迹序列。

步骤2：在原始数据集合的基础上生成历史轨迹网格的倒排索引表。首先将轨迹中的所有点全部通过映射后得到网格轨迹,然后将网格轨迹映射到（网格：轨迹编号，网格位置）的键值对上，通过对相同键的规约将结果写入数据库。

步骤3：初始化阶段是根据输入的起点和终点，依据离线阶段生成的轨迹数据库和倒排索引数据库，对所有从起点到终点的轨迹进行提取。通过已经建立的倒排索引数据库，分别提取经过起点和终点的所有轨迹编号，并且除去从终点到起点的轨迹，从而得到所有有效轨迹。

以5中倒排索引数据为例，若要获得从到的所有轨迹，首先从倒排索引数据库获得经过的所有轨迹以及经过的轨迹，又对于轨迹，网格在轨迹中的位置为3，网格在轨迹中的位置为2，则可以断定是从到的轨迹，因此到的所有轨迹。

步骤4：异常检测阶段是利用坐标点异常检测算法对船舶实时坐标进行处理，检测其是否存在异常并进行处理和记录。坐标点异常检测算法的主要思想是：维持一个包含最新传入GPS点的自适应窗口，并将窗口中的子轨迹与历史轨迹集合进行比较。随着新的GPS点被添加到自适应窗口中，逐步删除不包含自适应窗口中子轨迹的历史轨迹。只要在自适应窗口中的子轨迹的支持度高于，新的点将继续添加到窗口中。如果支持度低于，则自适应工作窗口减小为仅包含最新的GPS点。算法伪代码如图2所示。

下面用图3中的例子对算法进行简单说明：

图中有三条常走的轨迹route1、route2、route3，分别是100、200、150条。我们对待测轨迹进行检测，开始的时候适应窗口不断增加直到，到的时候由于窗口中的子轨迹没有轨迹支持，因此被加入异常点集合，接着子轨迹也没有轨迹支持，因此被加入异常点集合，接着子轨迹有150条route3轨迹支持，因此点正常，接着、、又被加入异常点集合，直到最后轨迹完成，异常点集合。

然后对船舶工作流程异常检测进行说明：

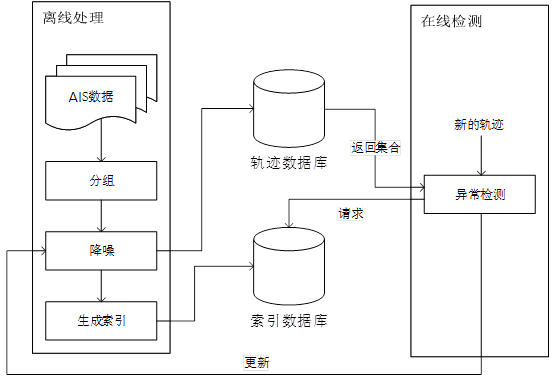
步骤1：对每条监测船舶指定具体的起始区域和目标区域，并明确区域具体坐标信息。

步骤2：将历史AIS船舶坐标序列转化成包含地理位置信息的状态序列和距离序列。状态值集合为：{ S1(在起始区域内),S2（在目标区域内）,S3（在起始区域和目标区域以外的区域）}。例如将轨迹中的所有点转化为状态序列：{S1,S1,S1,…,S3,S3,..,S3,S2,…,S3,…,S1,…}，距离序列：{d1,d2,…dn}。进而依据状态转化从状态序列中提取出标志性事件序列及其对应时间，并将历史工作流程信息记录到数据库中。如：S3->S1:E1（进入起始区域）,S1->S3:E2（出起始区域）,S3->S2:E3（进入目标区域）,S2->S3:E4（出目标区域）。标志性事件序列：{E1,E2,E3,E4,E1,E2,E3,E4,E1,…}。一个完整的工作流程为：E1(t1), E2(t2), E3(t3), E4(t4)。

步骤3：识别接近目标区域的坐标点集合：P={pn|dn<D},其中dn是距离序列中的值，对应坐标序列中的点到目标区域的距离。D是距离常量。

步骤4：根据数据库中历史工作流程计算每个工程标志性事件之间的平均时间间隔：E1->E2:T1,E2->E3:T2,E3->E4:T3。

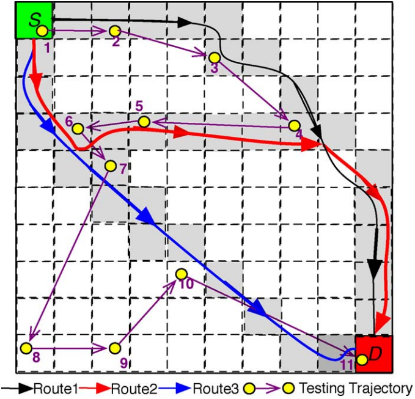
步骤5：将要进行判断的AIS坐标序列按照前述方法转化成状态序列，距离序列和标志性事件序列。提取标志性事件序列中的完整工作流程，计算工作流程相关时间间隔信息（t1,t2,t3,t4）。如果某工作流程Rn={E1,E2,E3,E4},(t1,t2,t3,t4),存在t1>|T1+x%|或t2>|T2+x%|或t3>|T3+x%|或t4>|T4+x%|, 则Rn为异常工作流程。



**图 1**

|  |
| --- |
| **坐标点异常检测算法** |
| 输入： - 待检测的轨迹流  - 起点到终点之间的所有网格轨迹集合  – 支持度的阈值 |
| 输出：轨迹的异常点集合 |
| //初始化  //待测点在轨迹中的位置  //自适应窗口初始化  **while** 待测轨迹未完成 **do**    //GPS点映射到网格  //将当前点加入  //计算窗口内子轨迹的支持度  //删除历史轨迹集合中不包括子轨迹的轨迹  **if** **then**  //将异常点加入异常点集合  //重置历史轨迹集合  //将自适应窗口修改为当前点  **end if**  **end while** |

**图 2**



**图3**