



(21) 申请号 202011601369.X

(22) 申请日 2020.12.29

(65) 同一申请的已公布的文献号

申请公布号 CN 112699467 A

(43) 申请公布日 2021.04.23

(73) 专利权人 香港理工大学深圳研究院

地址 518057 广东省深圳市南山区粤海街  
道高新技术产业园南区粤兴一道18号  
香港理工大学产学研大楼205室

(72) 发明人 王帅安 鄢然

(74) 专利代理机构 深圳市君胜知识产权代理事

务所(普通合伙) 44268

专利代理师 徐凯凯

(51) Int. Cl.

G06F 30/15 (2020.01)

G06F 30/27 (2020.01)

G06Q 10/04 (2023.01)

G06Q 10/0631 (2023.01)

G06Q 10/0635 (2023.01)

G06F 18/214 (2023.01)

G06F 18/2431 (2023.01)

G06F 111/04 (2020.01)

(56) 对比文件

CN 109754132 A, 2019.05.14

CN 110309967 A, 2019.10.08

审查员 胡瑞娟

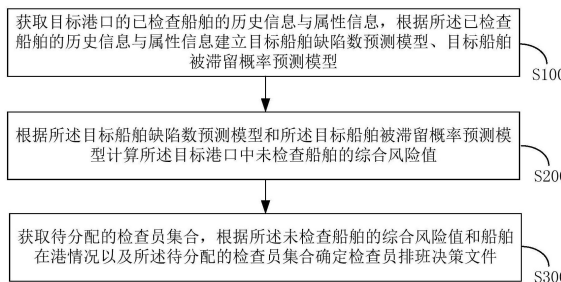
权利要求书2页 说明书10页 附图2页

(54) 发明名称

一种港口国监督检查船舶和检查员分配的方法

(57) 摘要

本发明公开了一种港口国监督检查船舶和检查员分配的方法,所述方法包括:获取目标港口的已检查船舶的历史信息与属性信息,根据所述已检查船舶的历史信息与属性信息建立目标船舶缺陷数预测模型、目标船舶被滞留概率预测模型;根据所述目标船舶缺陷数预测模型和所述目标船舶被滞留概率预测模型计算所述目标港口中未检查船舶的综合风险值;获取待分配的检查员集合,根据所述未检查船舶的综合风险值和船舶在港情况以及所述待分配的检查员集合确定检查员排班决策文件。解决了现有技术中缺乏船舶综合风险预测模型的问题,以及缺乏同时考虑检查资源以及船舶风险和停泊情况的检查员排班系统的问题。



1. 一种港口国监督检查船舶和检查员分配的方法,其特征在于,所述方法包括:

获取目标港口的历史港口国监督检查数据,根据所述历史港口国监督检查数据得到目标港口内的历史检查记录以及已检查船舶的编码数据;

基于所述编码数据获取所述已检查船舶的属性信息以及船舶历史检查信息;

将所述历史检查记录以及所述船舶历史检查信息作为目标港口的已检查船舶的历史信息;

对所述已检查船舶的历史信息与属性信息进行预处理操作,将预处理操作后得到的数据作为模型相关数据;

将所述模型相关数据分为训练数据集以及验证数据集;

根据所述训练数据集建立船舶缺陷数预测模型;

根据所述验证数据集对所述船舶缺陷数预测模型进行超参数调整,将超参数调整后得到的船舶缺陷数预测模型作为目标船舶缺陷数预测模型;

将所述训练数据集分为被滞留样本集以及未被滞留样本集;

根据所有所述被滞留样本集中的样本数据与等量随机抽取的所述未被滞留样本集中的样本数据生成平衡样本集;

对所述平衡样本集内的样本数据进行有放回抽样操作,抽取与所述平衡样本集的样本量等量的样本数据作为输入训练集;

根据所述输入训练集建立船舶被滞留概率预测模型;

根据所述验证数据集对所述船舶被滞留概率预测模型进行超参数调整,将超参数调整后得到的船舶被滞留概率预测模型作为目标船舶被滞留概率预测模型;

根据所述目标船舶缺陷数预测模型和所述目标船舶被滞留概率预测模型计算所述目标港口中未检查船舶的综合风险值;

获取待分配的检查员集合,根据所述未检查船舶的综合风险值和船舶在港情况以及所述待分配的检查员集合确定检查员排班决策文件。

2. 根据权利要求1所述的一种港口国监督检查船舶和检查员分配的方法,其特征在于,所述根据所述目标船舶缺陷数预测模型和所述目标船舶被滞留概率预测模型计算所述目标港口中未检查船舶的综合风险值包括:

计算所述训练数据集中船舶的缺陷数的平均值,将计算出的船舶的缺陷数的平均值作为船舶平均缺陷数;

计算所述训练数据集中船舶的被滞留概率的平均值,将计算出的船舶的被滞留概率的平均值作为船舶平均被滞留概率;

根据所述目标船舶缺陷数预测模型以及所述目标船舶被滞留概率预测模型分别计算所述未检查船舶的缺陷数和被滞留概率;

根据所述船舶平均缺陷数以及所述船舶平均被滞留概率分别对所述未检查船舶的缺陷数和被滞留概率进行归一化处理,并根据归一化处理后得到的数据计算所述未检查船舶的综合风险值。

3. 根据权利要求2所述的一种港口国监督检查船舶和检查员分配的方法,其特征在于,所述根据所述船舶平均缺陷数以及所述船舶平均被滞留概率分别对所述未检查船舶的缺陷数和被滞留概率进行归一化处理,并根据归一化处理后得到的数据计算所述未检查船舶

的综合风险值包括：

将所述未检查船舶的缺陷数与所述船舶平均缺陷数的比值作为归一化处理后的未检查船舶的缺陷数；

将所述未检查船舶的被滞留概率与所述船舶平均被滞留概率的比值作为归一化处理后的未检查船舶的被滞留概率；

根据所述归一化处理后的未检查船舶的缺陷数与所述归一化处理后的未检查船舶的被滞留概率之和得到所述未检查船舶的综合风险值。

4. 根据权利要求1所述的一种港口国监督检查船舶和检查员分配的方法,其特征在于,所述获取待分配的检查员集合,根据所述未检查船舶的综合风险值和船舶在港情况以及所述待分配的检查员集合确定检查员排班决策文件包括：

获取所有未检查船舶的综合风险值和船舶在港情况以及待分配的检查员集合,根据所述所有未检查船舶的综合风险值和船舶在港情况以及所述待分配的检查员集合中所有决策变量建立检查员分配模型；

获取预设的约束条件,根据满足所述预设的约束条件,且所述检查员分配模型的目标函数的数值最大时对应的决策变量的取值,生成检查员排班决策文件;所述决策变量包括:所述检查员的决策变量、所述未检查船舶的决策变量、所述未检查船舶被所述检查员在子时间段检查的决策变量。

5. 根据权利要求4所述的一种港口国监督检查船舶和检查员分配的方法,其特征在于,所述预设的约束条件包括下述的一种或者多种：

所述未检查船舶的检查时间满足在所述未检查船舶在所述目标港口的停留时间段内；

所述未检查船舶的检查时长与所述目标港口的所需检查时长相等；

所述未检查船舶的检查时间是一段连续的时间；

所述未检查船舶的检查开始时间是唯一的；

所述未检查船舶的检查次数小于或者等于1；

所述检查员在一个时间段检查的船舶数量小于或者等于1；

决策变量的取值为0或者1。

## 一种港口国监督检查船舶和检查员分配的方法

### 技术领域

[0001] 本发明涉及港口国监督检查领域,尤其涉及的是一种港口国监督检查船舶和检查员分配的方法。

### 背景技术

[0002] 现有的检查员排班系统采用一种贪心算法,即不考虑船舶的缺陷和滞留状况,仅根据可用检查员人数和船舶停靠时间随机地将需要检查的船舶分配给检查员。然而船舶缺陷和滞留作为港口国监督检查的主要且互相不可替代的结果,若缺乏对船舶缺陷和滞留的考量,则无法令港口国监督部门进行一个全面且直观的船舶风险评估,这样得到的排班策略也不能保证尽可能多的高风险船舶被检查到,进而导致无法保证有限的检查资源被最大化利用,从而限制了港口国监督的效用。

[0003] 因此,现有技术还有待改进和发展。

### 发明内容

[0004] 本发明要解决的技术问题在于,针对现有技术的上述缺陷,提供一种港口国监督检查船舶和检查员分配的方法,旨在解决现有技术中缺乏船舶综合风险预测模型的问题,以及缺乏同时考虑检查资源以及船舶风险和停泊情况的检查员排班系统,导致无法保证有限的检查资源被最大化利用,从而限制了港口国监督的效用问题。

[0005] 本发明解决问题所采用的技术方案如下:

[0006] 第一方面,本发明实施例提供一种港口国监督检查船舶和检查员分配的方法,其中,所述方法包括:

[0007] 获取目标港口的已检查船舶的历史信息与属性信息,根据所述已检查船舶的历史信息与属性信息建立目标船舶缺陷数预测模型、目标船舶被滞留概率预测模型;

[0008] 根据所述目标船舶缺陷数预测模型和所述目标船舶被滞留概率预测模型计算所述目标港口中未检查船舶的综合风险值;

[0009] 获取待分配的检查员集合,根据所述未检查船舶的综合风险值和船舶在港情况以及所述待分配的检查员集合确定检查员排班决策文件。

[0010] 在一种实施方式中,所述获取目标港口的已检查船舶的历史信息与属性信息,根据所述已检查船舶的历史信息与属性信息建立目标船舶缺陷数预测模型、目标船舶被滞留概率预测模型包括:

[0011] 获取目标港口的已检查船舶的历史信息与属性信息;

[0012] 对所述已检查船舶的历史信息与属性信息进行预处理操作,将预处理操作后得到的数据作为模型相关数据;

[0013] 将所述模型相关数据分为训练数据集以及验证数据集;

[0014] 根据所述训练数据集以及所述验证数据集确定目标船舶缺陷数预测模型;

[0015] 根据所述训练数据集以及所述验证数据集确定目标船舶被滞留概率预测模型。

[0016] 在一种实施方式中,所述获取目标港口的已检查船舶的历史信息与属性信息包括:

[0017] 获取目标港口的历史港口国监督检查数据,根据所述历史港口国监督检查数据得到目标港口内的已检查船舶的编码数据以及历史检查记录;

[0018] 基于所述编码数据获取所述已检查船舶的属性信息以及船舶历史检查信息;

[0019] 将所述历史检查记录以及所述船舶历史检查信息作为目标港口的已检查船舶的历史信息。

[0020] 在一种实施方式中,所述根据所述训练数据集以及所述验证数据集确定目标船舶缺陷数预测模型包括:

[0021] 根据所述训练数据集建立船舶缺陷数预测模型;

[0022] 根据所述验证数据集对所述船舶缺陷数预测模型进行超参数调整,将超参数调整后得到的船舶缺陷数预测模型作为目标船舶缺陷数预测模型。

[0023] 在一种实施方式中,所述根据所述训练数据集以及所述验证数据集确定目标船舶被滞留概率预测模型包括:

[0024] 将所述训练数据集分为被滞留样本集以及未被滞留样本集;

[0025] 根据所述被滞留样本集以及所述未被滞留样本集建立船舶被滞留概率预测模型;

[0026] 根据所述验证数据集对所述船舶被滞留概率预测模型进行超参数调整,将超参数调整后得到的船舶被滞留概率预测模型作为目标船舶被滞留概率预测模型。

[0027] 在一种实施方式中,所述根据所述被滞留样本集以及所述未被滞留样本集建立船舶被滞留概率预测模型包括:

[0028] 根据所有所述被滞留样本集中的样本数据与等量随机抽取的所述未被滞留样本集中的样本数据生成平衡样本集;

[0029] 对所述平衡样本集内的样本数据进行有放回抽样操作,抽取与所述平衡样本集的样本量等量的样本数据作为输入训练集;

[0030] 根据所述输入训练集建立船舶被滞留概率预测模型。

[0031] 在一种实施方式中,所述根据所述目标船舶缺陷数预测模型和所述目标船舶被滞留概率预测模型计算所述目标港口中未检查船舶的综合风险值包括:

[0032] 计算所述训练数据集中船舶的缺陷数的平均值,将计算出的船舶的缺陷数的平均值作为船舶平均缺陷数;

[0033] 计算所述训练数据集中船舶的被滞留概率的平均值,将计算出的船舶的被滞留概率的平均值作为船舶平均被滞留概率;

[0034] 根据所述目标船舶缺陷数预测模型以及所述目标船舶被滞留概率预测模型分别计算所述未检查船舶的缺陷数和被滞留概率;

[0035] 根据所述船舶平均缺陷数以及所述船舶平均被滞留概率分别对所述未检查船舶的缺陷数和被滞留概率进行归一化处理,并根据归一化处理后得到的数据计算所述未检查船舶的综合风险值。

[0036] 在一种实施方式中,所述根据所述船舶平均缺陷数以及所述船舶平均被滞留概率分别对所述未检查船舶的缺陷数和被滞留概率进行归一化处理,并根据归一化处理后得到的数据计算所述未检查船舶的综合风险值包括:

[0037] 将所述未检查船舶的缺陷数与所述船舶平均缺陷数的比值作为归一化处理后的未检查船舶的缺陷数；

[0038] 将所述未检查船舶的被滞留概率与所述船舶平均被滞留概率的比值作为归一化处理后的未检查船舶的被滞留概率；

[0039] 根据所述归一化处理后的未检查船舶的缺陷数与所述归一化处理后的未检查船舶的被滞留概率之和得到所述未检查船舶的综合风险值。

[0040] 在一种实施方式中,所述获取待分配的检查员集合,根据所述未检查船舶的综合风险值和船舶在港情况以及所述待分配的检查员集合确定检查员排班决策文件包括:

[0041] 获取所有未检查船舶的综合风险值和船舶在港情况以及待分配的检查员集合,根据所述所有未检查船舶的综合风险值和船舶在港情况以及所述待分配的检查员集合中所有决策变量建立检查员分配模型的函数;

[0042] 获取预设的约束条件,根据满足所述预设的约束条件,且所述检查员分配模型的目标函数的数值最大时对应的决策变量的取值,生成检查员排班决策文件。所述决策变量包括:所述检查员的决策变量、所述未检查船舶的决策变量、所述未检查船舶被所述检查员在子时间段检查的决策变量。

[0043] 在一种实施方式中,所述预设的约束条件包括下述的一种或者多种:

[0044] 所述未检查船舶的检查时间满足在所述未检查船舶在所述目标港口的停留时间段内;

[0045] 所述未检查船舶的检查时长与所述目标港口的所需检查时长相等;

[0046] 所述未检查船舶的检查时间是一段连续的时间;

[0047] 所述未检查船舶的检查开始时间是唯一的;

[0048] 所述未检查船舶的检查次数小于或者等于1;

[0049] 所述检查员在一个时间段检查的船舶数量小于或者等于1;

[0050] 决策变量的取值为0或者1。

[0051] 本发明的有益效果:本发明实施例通过获取目标港口的已检查船舶的历史信息与属性信息,根据所述已检查船舶的历史信息与属性信息建立目标船舶缺陷数预测模型、目标船舶被滞留概率预测模型;根据所述目标船舶缺陷数预测模型和所述目标船舶被滞留概率预测模型计算所述目标港口中未检查船舶的综合风险值;获取待分配的检查员集合,根据所述未检查船舶的综合风险值和船舶在港情况以及所述待分配的检查员集合确定检查员排班决策文件。解决了现有技术中缺乏船舶综合风险预测模型的问题,以及缺乏同时考虑检查资源以及船舶风险和停泊情况的检查员排班系统的问题。

## 附图说明

[0052] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍。显而易见地,下面描述中的附图仅仅是本发明中记载的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0053] 图1是本发明实施例提供的一种港口国监督检查船舶和检查员分配的方法的流程示意图。

[0054] 图2是本发明实施例提供的建立目标船舶缺陷数预测模型、目标船舶被滞留概率预测模型的流程示意图。

[0055] 图3是本发明实施例提供的计算所述目标港口中未检查船舶的综合风险值的流程示意图。

[0056] 图4是本发明实施例提供的确定检查员排班决策文件的流程示意图。

## 具体实施方式

[0057] 为使本发明的目的、技术方案及优点更加清楚、明确,以下参照附图并举实施例对本发明进一步详细说明。应当理解,此处所描述的具体实施例仅仅用以解释本发明,并不用于限定本发明。

[0058] 需要说明,若本发明实施例中有涉及方向性指示(诸如上、下、左、右、前、后……),则该方向性指示仅用于解释在某一特定姿态(如附图所示)下各部件之间的相对位置关系、运动情况等,如果该特定姿态发生改变时,则该方向性指示也相应地随之改变。

[0059] 安全是航运业永恒的主题,由于远洋船舶体积大,航运事故一旦发生将会对人力、物力和财力造成重大损失。港口所在地方的海事局有权力对外籍到访船舶进行检查,确保到访外籍船舶符合涉及船舶安全和船舶排放等方面的国际海事公约和国内法律法规,这种检查被称为港口国监督。港口国监督能有效遏制低标准船进入国家水域,对保障水上人命、财产安全以及防止船舶造成水域污染起到至关重要的作用。

[0060] 由于到访港口的外籍船舶数量多、情况复杂,且检查资源有限,设计切实高效的选船方案至关重要。目前,一种基于经验的船舶风险评估法是世界各地港口广泛采用的一种选船机制。该方法通过对船舶自身属性数据和历史检查数据的粗略评估,将船舶划分为三个风险等级,分别为高风险船舶、中风险船舶、低风险船舶。同时,各风险等级有不同长度的时间窗,港口国通过上次检查时间与时间窗的相对关系对外籍船舶进行抽检。通过对近十年来的历史数据分析,发明人发现只有60%左右的被检船舶存在缺陷,而大量不合格外籍船舶没有被检查,这些不合格船将对航运安全和海洋环境造成威胁。造成基于经验的选船方法效率不高的原因主要有以下三方面。第一,该基于经验的选船方法考虑的指标及其权重主要来源于主观经验,缺乏实证数据的支撑;第二,对不同指标赋予的权重值的准确性和一致性未得到验证,且最后计算出的船舶风险等级依赖于各指标权重值的简单加和,未考虑各指标之间的相互作用;第三,最终的风险等级将所有外籍船舶划分为三个大类,但对同一个大类中的船舶风险状况未提供更详细的信息帮助海事部门准确选船。这些不足极大限制了现有选船方案的检查效率和能检查到的船舶缺陷及船舶滞留。在挑选出需要检查的船舶后,海事管理部门需要分配港口国监督检查员登船检查。现有的检查员排班系统采用一种贪心算法,即不考虑船舶的缺陷和滞留状况,仅根据可用检查员人数和船舶停靠时间随机地将需要检查的船舶分配给检查员。这样的排班策略不能保证尽可能多的高风险船舶被检查到,从而大大减弱了港口国监督检查消除不合格船舶的效用。

[0061] 现有的港口国监督相关研究提出了一些识别高风险到访船舶的模型。这些模型主要通过分析船舶自身属性数据(例如船龄、船型、船旗、船舶管理公司、船舶认证机构)、船舶历史检查数据(例如上次港口国监督检查时间、历史被滞留情况)以及船舶涉及的航运事故数据来量化船舶的风险值。但是这些模型存在一些问题,并且很难被海事部门直接使用。最

主要原因是绝大多数研究旨在通过预测模型生成船舶的风险值分数,但是风险值分数无法直接观测或衡量,导致这些模型的准确性难以度量,实用性被大大降低。其次,船舶缺陷和滞留作为港口国监督检查的主要且互相不可替代的结果,在现有的模型中很少被同时考虑,因此无法给港口国监督部门提供一个全面且直观的船舶风险评估。同时,现有的模型考虑的因素多直接来自于港口国监督数据库,其他数据库提供的船舶相关数据较少被考虑。另外,由于现有的模型大多为静态的理论模型,无法根据实时更新的船舶检查数据和船舶到访数据实现在线决策支持系统,供港口国监督部门下载可视化的在港船舶位置信息及风险预测结果,从而进一步限制了它们的实用价值。概括地讲,现有的港口国监督相关研究没有建立考虑检查资源以及船舶风险和停泊情况的检查员排班系统,因而无法保证有限的检查资源被最大化利用,从而限制了港口国监督的效用。

[0062] 针对现有技术的上述缺陷,本发明提供了一种港口国监督检查船舶和检查员分配的方法,通过建立船舶缺陷数预测模型和船舶被滞留概率预测模型计算港口中未检查船舶的综合风险值,并根据计算出的未检查船舶的综合风险值、船舶停留情况以及可分配的检查员集合,确定检查员能检查到的所有未检查船舶的综合风险值之和最大时的检查员排班决策文件,使有限的检查资源被最大化利用。本发明在优化检查员排班系统时充分考虑了港口的检查资源、船舶风险和停泊情况,可以保证有限的检查资源被最大化利用,从而提高了港口国监督的效用。

[0063] 如图1所示,本实施例提供一种港口国监督检查船舶和检查员分配的方法,所述方法包括如下步骤:

[0064] 步骤S100、获取目标港口的已检查船舶的历史信息与属性信息,根据所述已检查船舶的历史信息与属性信息建立目标船舶缺陷数预测模型、目标船舶被滞留概率预测模型。

[0065] 具体地,为了实现后续对未检查船舶的风险进行评估,本实施例需要先建立用于预测船舶的缺陷数和被滞留概率的模型。建立模型之前需要获取用于建模的数据,本实施例将目标港口的已检查船舶的历史信息与属性信息作为建模数据,根据所述已检查船舶的历史信息与属性信息建立目标船舶缺陷数预测模型、目标船舶被滞留概率预测模型,从而实现后续使用目标船舶缺陷数预测模型、目标船舶被滞留概率预测模型综合评估船舶风险。

[0066] 在一种实现方式中,如图2所示,所述步骤S100具体地包括如下步骤:

[0067] 步骤S110、获取目标港口的已检查船舶的历史信息与属性信息;

[0068] 步骤S120、对所述已检查船舶的历史信息与属性信息进行预处理操作,将预处理操作后得到的数据作为模型相关数据;

[0069] 步骤S130、将所述模型相关数据分为训练数据集以及验证数据集;

[0070] 步骤S140、根据所述训练数据集以及所述验证数据集确定目标船舶缺陷数预测模型;

[0071] 步骤S150、根据所述训练数据集以及所述验证数据集确定目标船舶被滞留概率预测模型。

[0072] 首先,本实施例需要获取历史到港并进行过检查的船舶的历史信息与属性信息作为构建预测模型的基础数据。在一种实现方式中,可以通过首先获取目标港口的历史港口



国监督检查数据,根据所述历史港口国监督检查数据得到目标港口内的已检查船舶的编码数据以及历史检查记录,然后根据所述编码数据获取所述已检查船舶的属性信息以及船舶历史检查信息,最后将所述历史检查记录以及所述船舶历史检查信息作为目标港口的已检查船舶的历史信息。

[0073] 具体地,由于港口国监督检查数据通常都有存储记录,因此本实施例可以通过预设的计算机程序实现访问存储港口国监督检查数据的网站,并提取需要的港口国监督检查数据作为目标港口的已检查船舶的历史信息。举例说明,在亚太地区,存储港口国监督检查数据的是东京备忘录网站,本实施例可以通过自行设计的高效的网络爬虫程序采集东京备忘录网站上亚太地区各港口近二十年的港口国监督检查记录,并从中获取与目标港口相关的历史港口国监督检查数据。历史港口国监督检查数据中的每条检查记录包含了船舶IMO编码和船舶历史检查信息,例如检查日期,缺陷数,滞留情况等。然后可以根据船舶IMO编码数据,从本地World Shipping Register数据库中爬取更详细的船舶的属性信息,例如船型,船龄,船舶容积总吨,船长,船宽,船深,历史更换船旗次数,历史被滞留次数,五年内事故发生情况。以及从东京备忘录网站中爬取更详细的船舶历史检查信息,例如上次检查到的缺陷数,船旗表现,认证机构表现,船公司表现等历史检查数据。

[0074] 获取到目标港口的已检查船舶的历史信息以后,需要对所述已检查船舶的历史信息与属性信息进行预处理操作。在一种实现方式中,所述预处理操作包括:数据清洗、数据集成、数据规约、数据变换等操作中的一种或者多种。然后,将预处理操作后得到的数据作为模型相关数据,并将所述模型相关数据分为训练数据集以及验证数据集。具体地,可以将所述模型相关数据平均分为五组,记做 $g_1$ 到 $g_5$ 。每次实验中,分别选择 $g_1$ 到 $g_5$ 中的一组作为验证数据集,剩余四组作为训练数据集,例如第一次采用 $g_1$ 作为验证数据集, $g_2$ 到 $g_5$ 作为训练数据集,第二次采用 $g_2$ 作为验证数据集,其他四组作为训练数据集,以此类推,一共进行五次实验即五折交叉验证。所述训练数据集作为船舶缺陷数预测模型和船舶被滞留概率预测模型的输入数据来构建模型,例如所述训练数据集可以包含船型,船龄,船舶容积总吨,船长,船宽,船深,历史更换船旗次数,历史被滞留次数,五年内事故发生情况,上次港口国监督检查时间,上次检查到的缺陷数,船旗表现,认证机构表现,船公司表现等十四个变量。并通过所述验证数据集以及训练数据集确定目标船舶缺陷数预测模型和目标船舶被滞留概率预测模型。具体地,可以根据所述训练数据集建立船舶缺陷数预测模型,然后根据所述验证数据集对所述船舶缺陷数预测模型进行超参数调整,将超参数调整后得到的船舶缺陷数预测模型作为目标船舶缺陷数预测模型。

[0075] 在一种实现方式中,所述船舶缺陷数预测模型可以为机器学习中基于CART回归树的随机森林模型,即本实施例可以根据所述训练数据集建立基于CART回归树的随机森林模型,得到待调整的船舶缺陷数预测模型。然后利用所述验证数据集对所述船舶缺陷数预测模型进行超参数调整。在一种实现方式中,所述超参数调整可以通过网格搜索对所述船舶缺陷数预测模型中的以下超参数进行调整:每颗决策树的最大深度max\_depth,每个叶子节点中需要包含的最少样本数min\_samples\_leaf,随机森林中包含的决策树数量n\_estimators,每颗决策树考虑的特征个数max\_features。然后将验证结果中模型表现最好的情况下对应的那组超参数作为最优超参数,以完成船舶缺陷数预测模型的超参数调整。在一种实现方式中,所述船舶缺陷数预测模型的结点的分裂目标和模型训练目标为最小化

均方误差 (mean squared error, 简称MSE)。例如, 当将模型相关数据平均分成五份以后, 每次采用其中一份作为验证数据集, 另外四份作为训练数据集, 一共进行五次试验, 则五个训练集对应的五个验证集上平均MSE最小的那一组超参数为最优的超参数, 根据最优的超参数对船舶缺陷数预测模型进行超参数调整。

[0076] 为了得到目标船舶被滞留概率预测模型, 在一种实现方式中, 首先将所述训练数据集分为被滞留样本集以及未被滞留样本集, 然后根据所述被滞留样本集以及所述未被滞留样本集建立船舶被滞留概率预测模型。具体地, 可以先根据所有所述被滞留样本集中的样本数据与等量随机抽取的所述未被滞留样本集中的样本数据生成平衡样本集, 所述平衡样本集是对随机森林中的每棵决策树进行的, 所述随机森林中包含多颗决策树。换言之, 所述平衡样本集在每建立一颗树的时候都会生成一次。然后对所述平衡样本集内的样本数据进行有放回抽样操作, 抽取与所述平衡样本集的样本量等量的样本数据作为输入训练集, 最后根据所述输入训练集建立船舶被滞留概率预测模型。例如训练数据集中有十个被滞留样本和九十十个未被滞留样本, 先抽取所有的被滞留样本和随机的十个未被滞留样本生成平衡样本集, 在这包含二十个样本的平衡样本集中有放回抽样二十个样本作为输入训练集。需要说明的是, 上述步骤均是建立一颗决策树时的操作, 对于随机森林中的每棵决策树, 都应进行上述操作。本实施例中的船舶被滞留概率预测模型为平衡随机森林模型, 通过该模型可以着重处理训练数据集中被滞留船舶检查记录和未被滞留船舶检查记录分布严重不均的问题。

[0077] 同样地, 所述船舶被滞留概率预测模型也需要根据所述验证数据集进行超参数调整, 将超参数调整后得到的船舶被滞留概率预测模型作为目标船舶被滞留概率预测模型。在一种实现方式中, 所述船舶被滞留概率预测模型的超参数调整可以通过网格搜索对所述船舶被滞留概率预测模型中的以下超参数进行调整: 每颗决策树的最大深度max\_depth, 每个叶子节点中需要包含的最少样本数min\_samples\_leaf, 随机森林中包含的决策树数量n\_estimators, 每颗决策树考虑的特征个数max\_features。然后将验证结果中模型表现最好的情况下对应的那组超参数作为最优超参数, 以完成船舶被滞留概率预测模型的超参数调整。在一种实现方式中, 所述船舶被滞留概率预测模型的结点的分裂目标为最小化Gini系数 (Gini index), 模型的训练目标为最大化ROC曲线下方的面积大小 (ROC AUC)。例如, 当将模型相关数据平均分成五份以后, 每次采用其中一份作为验证数据集, 另外四份作为训练数据集, 一共进行五次试验, 则五个训练集对应的五个验证集上ROC曲线下方的平均面积最大的那一组超参数为最优的超参数, 根据最优的超参数对船舶被滞留概率预测模型进行超参数调整。

[0078] 确定目标船舶缺陷数预测模型和所述目标船舶被滞留概率预测模型以后, 如图1所示, 所述方法还包括如下步骤:

[0079] 步骤S200、根据所述目标船舶缺陷数预测模型和所述目标船舶被滞留概率预测模型计算所述目标港口中未检查船舶的综合风险值。

[0080] 具体地, 由于本实施例的目标是希望能最大化检查到的船舶风险, 因此首先需要对每一艘未检查船舶的风险值进行评估。所述目标船舶缺陷数预测模型可以用于预测未检查船舶的缺陷数, 所述目标船舶被滞留概率预测模型可以用于预测未检查船舶的被滞留概率, 并将缺陷数和被滞留概率作为未检查船舶的风险值的评价指标。

[0081] 在一种实现方式中,如图3所示,所述方法包括如下步骤:

[0082] 步骤S210、计算所述训练数据集中船舶的缺陷数的平均值,将计算出的船舶的缺陷数的平均值作为船舶平均缺陷数;

[0083] 步骤S220、计算所述训练数据集中船舶的被滞留概率的平均值,将计算出的船舶的被滞留概率的平均值作为船舶平均被滞留概率;

[0084] 步骤S230、根据所述目标船舶缺陷数预测模型以及所述目标船舶被滞留概率预测模型分别计算所述未检查船舶的缺陷数和被滞留概率;

[0085] 步骤S240、根据所述船舶平均缺陷数以及所述船舶平均被滞留概率分别对所述未检查船舶的缺陷数和被滞留概率进行归一化处理,并根据归一化处理后得到的数据计算所述未检查船舶的综合风险值。

[0086] 具体地,所述未检查船舶包括实时在港但是未经过检查的船舶和即将到港船舶,未检查船舶的相关信息可以通过自行设计的高效网络爬虫系统从相关海事管理部门的官方网站获取。此外,由于缺陷数和被滞留概率的数量级不同,首先需要对未检查船舶的缺陷数和被滞留概率进行归一化处理。本实施例将训练数据集中的船舶平均缺陷数和船舶平均被滞留概率作为所述归一化处理中要用到的归一化因子。然后根据所述船舶平均缺陷数以及所述船舶平均被滞留概率分别对所述未检查船舶的缺陷数和被滞留概率进行归一化处理,并根据归一化处理后得到的数据计算所述未检查船舶的综合风险值。在一种实现方式中,所述归一化处理具体为:将所述未检查船舶的缺陷数与所述船舶平均缺陷数的比值作为归一化处理后的未检查船舶的缺陷数,将所述未检查船舶的被滞留概率与所述船舶平均被滞留概率的比值作为归一化处理后的未检查船舶的被滞留概率。举例说明,将训练数据集中船舶平均缺陷数记做 $\overline{def}$ ,船舶平均被滞留概率记做 $\overline{det}$ 。根据所述目标船舶缺陷数预测模型以及所述目标船舶被滞留概率预测模型分别计算所述未检查船舶的缺陷数 $def'$ 和被滞留概率 $det'$ ,该未检查船舶的综合风险值计算公式如下:

[0087]  $risk = (def' / \overline{def}) + (det' / \overline{det})$ 。

[0088] 确定了所有未检查船舶的综合风险值以后,如图1所示,所述方法还包括如下步骤:

[0089] 步骤S300、获取待分配的检查员集合,根据所述未检查船舶的综合风险值和船舶在港情况以及所述待分配的检查员集合确定检查员排班决策文件。

[0090] 具体地,为了保证有限的检查资源被最大化利用,从而提高港口国监督的效用,需要充分考虑目标港口的检查资源、船舶风险和停泊情况。因此本实施例还需要获取待分配的检查员集合,这里的待分配即为可分配的意思。然后根据所述未检查船舶的综合风险值和船舶在港情况以及所述待分配的检查员集合确定最大化检查到的船舶风险对应的检查员排班决策文件,以此实现对有限检查资源的合理调配。

[0091] 在一种实现方式中,如图4所示,所述步骤S300具体包括:

[0092] 步骤S310、获取所有未检查船舶的综合风险值和船舶在港情况以及待分配的检查员集合,根据所述所有未检查船舶的综合风险值和船舶在港情况以及所述待分配的检查员集合中所有决策变量建立检查员分配模型;

[0093] 步骤S320、获取预设的约束条件,将满足所述预设的约束条件,且所述检查员分配

模型的目标函数的数值最大时对应的决策变量的取值,生成检查员排班决策文件。所述决策变量包括:所述检查员的决策变量、所述未检查船舶的决策变量、所述未检查船舶被所述检查员在子时间段检查的决策变量。

[0094] 具体地,本实施例首先需要根据所有未检查船舶的综合风险值和船舶在港情况以及所述待分配的检查员集合中所有决策变量值建立检查员分配模型。然后将满足预设的约束条件,且所述检查员分配模型的目标函数的数值最大时对应的决策变量的取值,生成检查员排班决策文件。所述决策变量包括:所述检查员的决策变量、所述未检查船舶的决策变量、所述未检查船舶被所述检查员在子时间段检查的决策变量。举例说明,将某时刻在港且尚未被检查的船舶和可分配的检查员集合分别记做S和P,计算船s,  $s \in S$  的综合风险值记做  $risk_s$ ,检查员分配模型的目标函数是最大化能检查到的船舶风险值之和,检查员分配模型的目标函数如下所示:

$$[0095] \quad \max \sum_{s \in S} \sum_{p \in P} risk_s y_{sp}$$

[0096] 其中,  $risk_s$  是船s的综合风险值,  $y_{sp}$  是表示船s是否被检查员p检查的决策变量值。在一种实现方式中,所述预设的约束条件包括下述的一种或者多种:

[0097] (1) 所述未检查船舶的检查时间满足在所述未检查船舶在所述目标港口的停留时间段内。公式为:

$$[0098] \quad x_{spt} \leq e_t^s, \forall s \in S, \forall p \in P, t = 1, \dots, T$$

[0099] 其中,假设子时间段为 $\tau$ ,此时刻距检查员下班时间的的时间间隔与 $\tau$ 时间段长度的比值向上取整记做T,一次检查时长(港口国监督所需检查时长)与 $\tau$ 时间段长度的比值向上取整记做T'。对于子时间段 $\tau$ ,如果船s在整个时间段中均在港则令  $e_t^s = 1$ , 否则  $e_t^s = 0$ 。定义0-1决策变量  $x_{spt}$ , 在时间段 $\tau$ 中如果船s正在被检查员p检查则令  $x_{spt} = 1$ , 否则  $x_{spt} = 0$ 。即  $e_t^s$  表示船舶s在子时间段 $\tau$ 内的在港情况,  $x_{spt}$  表示船舶s在子时间段 $\tau$ 被检查员p检查的情况。

[0100] (2) 所述未检查船舶的检查时长与所述目标港口的所需检查时长相等。公式为:

$$[0101] \quad \sum_{\tau=1}^T x_{spt} = T' y_{sp}, \forall s \in S, \forall p \in P$$

[0102] 其中,定义0-1决策变量  $y_{sp}$ , 如果船s被检查员p检查则令  $y_{sp} = 1$ , 否则  $y_{sp} = 0$ 。即  $y_{sp}$  表示船s是否被检查员p检查。

[0103] (3) 所述未检查船舶的检查时间是一段连续的时间。公式为:

$$[0104] \quad \sum_{\tau=\bar{t}}^{\bar{t}+3} x_{spt} \geq T' \sigma_{sp}^{\bar{t}}, \forall s \in S, \forall p \in P, 1 \leq \bar{t} \leq T - T'$$

[0105] 其中,定义0-1决策变量  $\sigma_{sp}^{\bar{t}}$ , 如果船s被检查员p从时间段 $\tau$ 开始检查则令  $\sigma_{sp}^{\bar{t}} = 1$ , 否则  $\sigma_{sp}^{\bar{t}} = 0$ 。记此时刻所在的时间段为 $\bar{t}$ ,  $\bar{t} \leq T - T'$ 。即  $\sigma_{sp}^{\bar{t}}$  表示船s是否被检查员p从时间段 $\tau$ 开始检查。

[0106] (4) 所述未检查船舶的检查开始时间是唯一的。公式为:

$$[0107] \quad \sum_{\tau=1}^T \sigma_{sp}^{\tau} = y_{sp}, \forall s \in S, \forall p \in P$$

[0108] (5) 所述未检查船舶的检查次数小于或者等于1。公式为:

$$[0109] \quad \sum_{p \in P} y_{sp} \leq 1, \forall s \in S$$

[0110] (6) 所述检查员在一个时间段检查的船舶数量小于或者等于1。公式为:

$$[0111] \quad \sum_{s \in S} x_{sp\tau} \leq 1, \forall p \in P, \tau = 1, \dots, T$$

[0112] (7) 决策变量的取值为0或者1。公式为:

$$[0113] \quad x_{sp\tau} \in \{0, 1\}, \forall s \in S, \forall p \in P, \tau = 1, \dots, T$$

$$[0114] \quad y_{sp} \in \{0, 1\}, \forall s \in S, \forall p \in P$$

$$[0115] \quad \sigma_{sp}^{\tau} \in \{0, 1\}, \forall s \in S, \forall p \in P, \tau = 1, \dots, T$$

[0116] 然后求解满足预设的约束条件,且当检查员分配模型的目标函数是最大化能检查到的船舶风险值之和时对应的决策变量的取值,所述决策变量包括:所述检查员的决策变量、所述未检查船舶的决策变量、所述未检查船舶被所述检查员在子时间段检查的决策变量,并生成检查员排班决策文件。在一种实现方式中,还可以对在港船舶所在位置和对其预测的综合风险值在地图上进行可视化,并且提供可供预览和下载的船舶的综合风险值预测文件和检查员排班决策文件,以实现提升港口国监督识别高风险船舶和分配检查员的效率。

[0117] 综上所述,本发明公开了一种港口国监督检查船舶和检查员分配的方法,通过获取目标港口的已检查船舶的历史信息与属性信息,根据所述已检查船舶的历史信息与属性信息建立目标船舶缺陷数预测模型、目标船舶被滞留概率预测模型;根据所述目标船舶缺陷数预测模型和所述目标船舶被滞留概率预测模型计算所述目标港口中未检查船舶的综合风险值;获取待分配的检查员集合,根据所述未检查船舶的综合风险值和船舶在港情况以及所述待分配的检查员集合确定检查员排班决策文件。解决了现有技术中缺乏船舶综合风险预测模型,以及缺乏同时考虑检查资源以及船舶风险和停泊情况的检查员排班系统,导致无法保证有限的检查资源被最大化利用,从而限制了港口国监督效用的问题。

[0118] 应当理解的是,本发明的应用不限于上述的举例,对本领域普通技术人员来说,可以根据上述说明加以改进或变换,所有这些改进和变换都应属于本发明所附权利要求的保护范围。

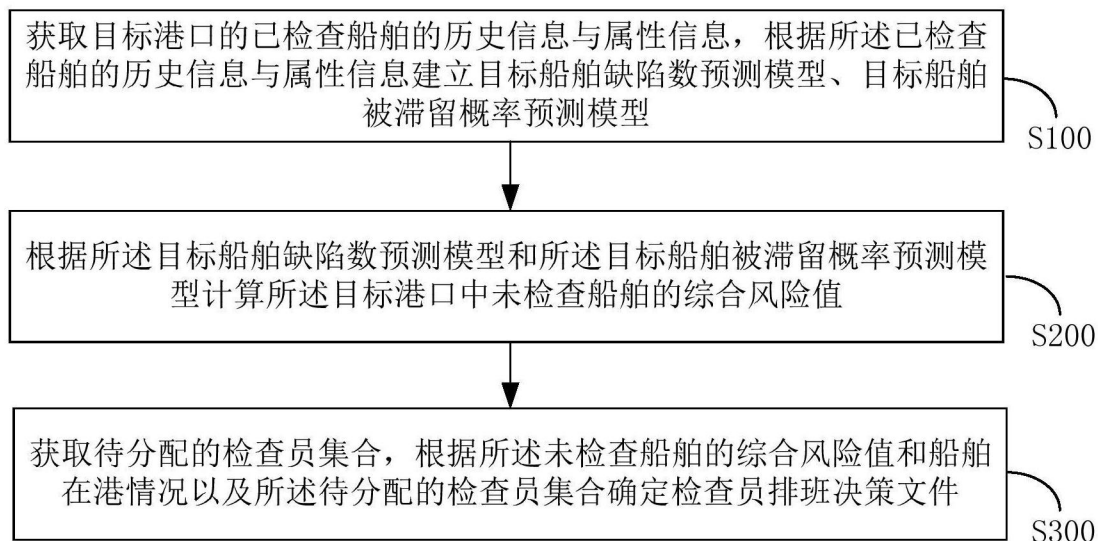


图1

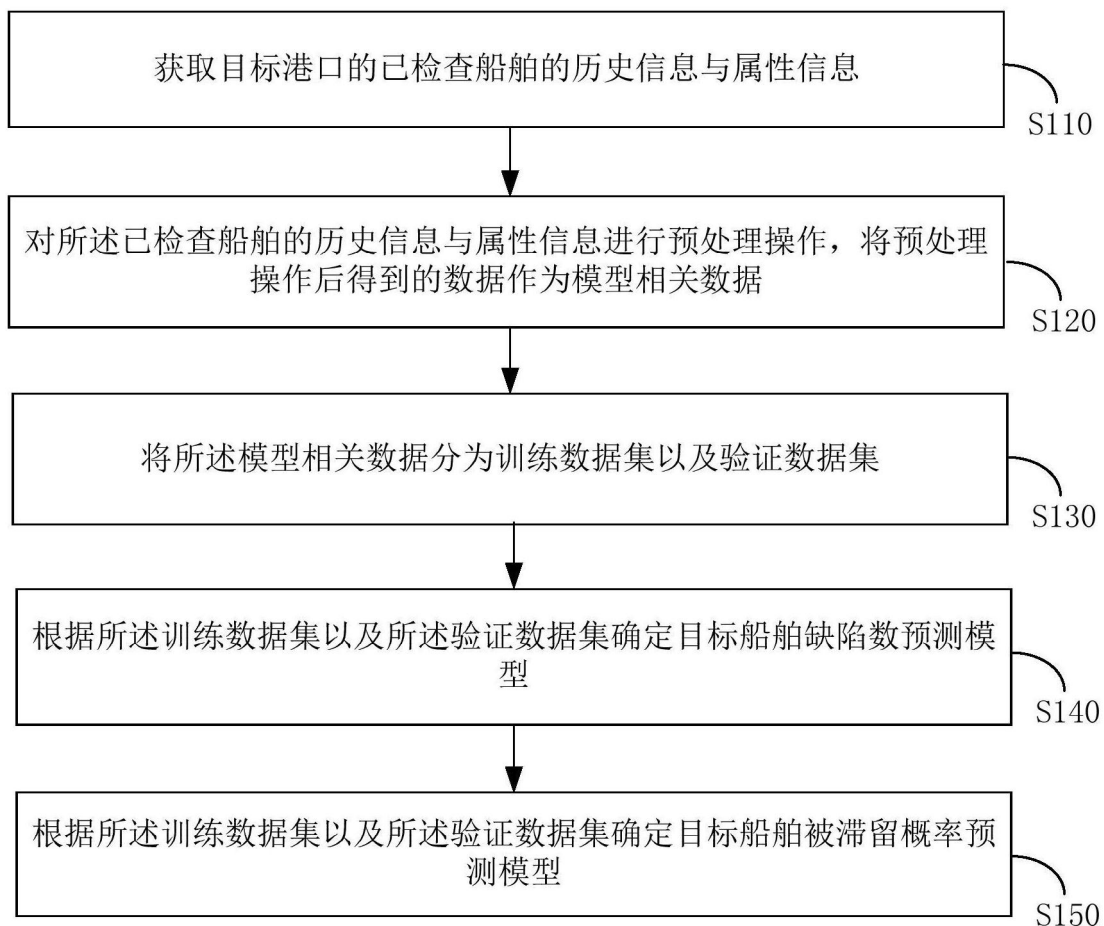


图2

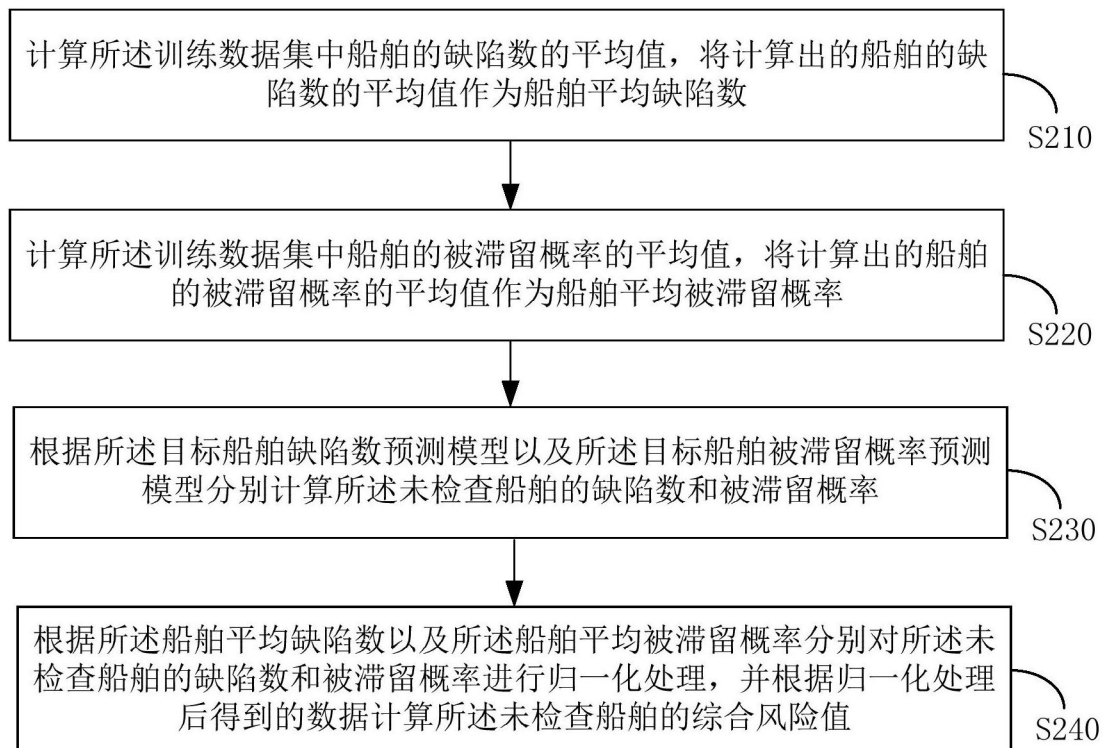


图3

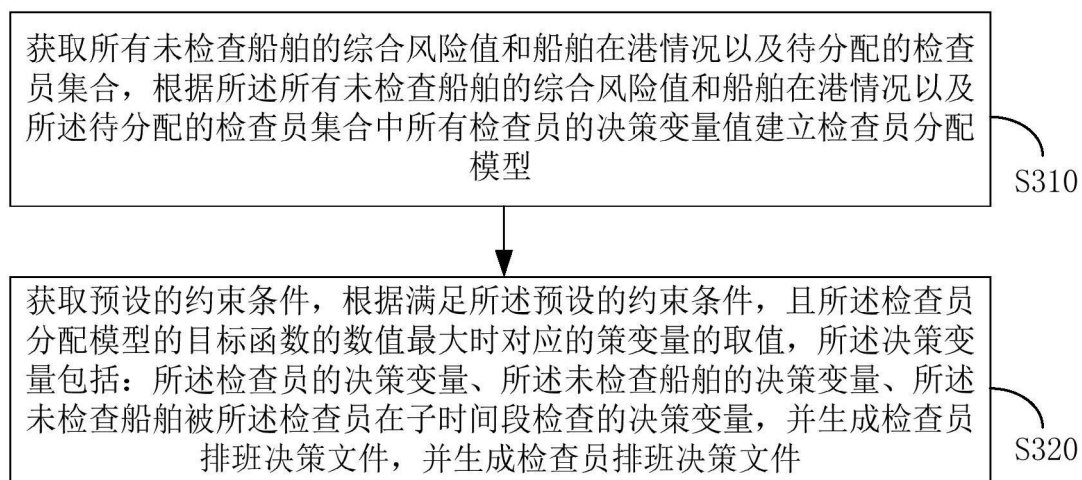


图4