



(12) 发明专利申请

(10) 申请公布号 CN 114154428 A

(43) 申请公布日 2022.03.08

(21) 申请号 202210119592.3

(22) 申请日 2022.02.09

(71) 申请人 中国科学院地理科学与资源研究所  
地址 100101 北京市朝阳区大屯路甲11号

(72) 发明人 江东 丁方宇 郝蒙蒙 付晶莹

(74) 专利代理机构 北京中和立达知识产权代理  
有限公司 11756

代理人 张可

(51) Int. Cl.

G06F 30/27 (2020.01)

G06Q 10/06 (2012.01)

G06Q 50/26 (2012.01)

G06N 20/00 (2019.01)

G06F 111/08 (2020.01)

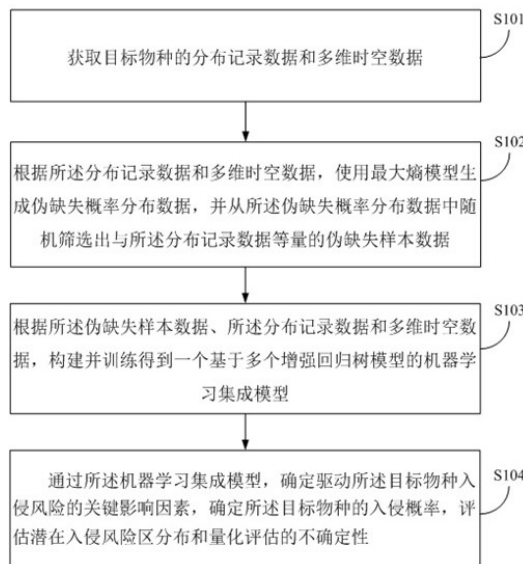
权利要求书2页 说明书12页 附图6页

(54) 发明名称

目标物种潜在入侵风险评估方法及装置

(57) 摘要

本发明是关于一种目标物种潜在入侵风险评估方法及装置,其中,方法包括:获取目标物种的分布记录数据和多维时空数据;根据所述分布记录数据和多维时空数据,使用最大熵模型生成伪缺失概率分布数据,并从所述伪缺失概率分布数据中随机筛选出与所述分布记录数据等量的伪缺失样本数据;根据所述伪缺失样本数据、所述分布记录数据和多维时空数据,构建并训练得到一个基于多个增强回归树模型的机器学习集成模型;通过所述机器学习集成模型,确定驱动所述目标物种入侵风险的关键影响因素,确定所述目标物种的入侵概率,评估潜在入侵风险区分布和量化评估的不确定性。



1. 一种目标物种潜在入侵风险评估方法,其特征在于,所述方法包括:

获取目标物种的分布记录数据和多维时空数据;

根据所述分布记录数据和多维时空数据,使用最大熵模型生成伪缺失概率分布数据,并从所述伪缺失概率分布数据中随机筛选出与所述分布记录数据等量的伪缺失样本数据;

根据所述伪缺失样本数据、所述分布记录数据和多维时空数据,构建并训练得到一个基于多个增强回归树模型的机器学习集成模型;

通过所述机器学习集成模型,确定驱动所述目标物种入侵风险的关键影响因素,确定所述目标物种的入侵概率,评估潜在入侵风险区分布和量化评估的不确定性。

2. 根据权利要求1所述的方法,其特征在于,所述分布记录数据包括所述目标物种分布的经纬度信息,所述多维时空数据包括气候因素信息、地理因素信息和社会经济因素信息,其中,所述气候因素信息包括年平均气温、平均昼夜温度差、等温性、气温季节性变动系数、最热月份最高温度、最冷月份最低温度、最高温度与最低温度差、最湿季度平均温度、最干季度平均温度、最暖季度平均温度、最冷季度平均温度、年降雨量、最湿月份降水量、最干月份降水量、降水量季节性变化、最干季度降水量、最湿季度降水量、最暖季度降水量和最冷季度降水量,所述地理因素信息包括海拔、归一化植被指数和土地利用,所述社会经济因素信息包括城市可达性和道路密度。

3. 根据权利要求1所述的方法,其特征在于,根据所述分布记录数据和多维时空数据,使用最大熵模型生成伪缺失概率分布数据,并从所述伪缺失概率分布数据中随机筛选出与所述分布记录数据等量的伪缺失样本数据,包括:

将所述分布记录数据和多维时空数据作为输入数据,运行最大熵模型,以生成介于0-1之间的环境适宜性概率;

将所述环境适宜性概率小于0.5作为所述伪缺失样本数据的筛选依据。

4. 根据权利要求1所述的方法,其特征在于,根据所述伪缺失样本数据、所述分布记录数据和多维时空数据,构建并训练得到一个基于多个增强回归树模型的机器学习集成模型,包括:

将随机筛选出与所述分布记录数据等量的伪缺失样本数据的过程执行多次;

利用每次获得的伪缺失样本数据、所述分布记录数据和多维时空数据构建增强回归树模型,进而得到一个基于多个增强回归树模型的机器学习集成模型。

5. 根据权利要求1所述的方法,其特征在于,通过所述机器学习集成模型,确定驱动所述目标物种入侵风险的关键影响因素,确定所述目标物种的入侵概率,评估潜在入侵风险区分布和量化评估的不确定性,包括:

通过所述机器学习集成模型确定所述多维时空数据中各类数据的相对贡献率;

利用所述机器学习集成模型在预设面积的格网预测得到多个入侵概率结果,其中,一个增强回归树模型对应一个入侵概率结果,逐栅格单元计算平均值,最终获取所述目标物种的全球入侵概率;

将入侵概率结果大于预设概率值的区域确定为潜在入侵风险区;

基于预测得到的多个入侵概率结果,逐栅格计算标准差,以分析预测结果的不确定性。

6. 一种目标物种潜在入侵风险评估装置,其特征在于,所述装置包括:

获取模块,用于获取目标物种的分布记录数据和多维时空数据;

筛选模块,用于根据所述分布记录数据和多维时空数据,使用最大熵模型生成伪缺失概率分布数据,并从所述伪缺失概率分布数据中随机筛选出与所述分布记录数据等量的伪缺失样本数据;

模型构建模块,用于根据所述伪缺失样本数据、所述分布记录数据和多维时空数据,构建并训练得到一个基于多个增强回归树模型的机器学习集成模型;

确定模块,用于通过所述机器学习集成模型,确定驱动所述目标物种入侵风险的关键影响因素,确定所述目标物种的入侵概率,评估潜在入侵风险区分布和量化评估的不确定性。

7. 根据权利要求6所述的装置,其特征在于,所述分布记录数据包括所述目标物种分布的经纬度信息,所述多维时空数据包括气候因素信息、地理因素信息和社会经济因素信息,其中,所述气候因素信息包括年平均气温、平均昼夜温度差、等温性、气温季节性变动系数、最热月份最高温度、最冷月份最低温度、最高温度与最低温度差、最湿季度平均温度、最干季度平均温度、最暖季度平均温度、最冷季度平均温度、年降雨量、最湿月份降水量、最干月份降水量、降水量季节性变化、最干季度降水量、最湿季度降水量、最暖季度降水量和最冷季度降水量,所述地理因素信息包括海拔、归一化植被指数和土地利用,所述社会经济因素信息包括城市可达性和道路密度。

8. 根据权利要求6所述的装置,其特征在于,所述筛选模块用于:

将所述分布记录数据和多维时空数据作为输入数据,运行最大熵模型,以生成介于0-1之间的环境适宜性概率;

将所述环境适宜性概率小于0.5作为所述伪缺失样本数据的筛选依据。

9. 根据权利要求6所述的装置,其特征在于,所述模型构建模块包括:

执行单元,用于将随机筛选出与所述分布记录数据等量的伪缺失样本数据的过程执行多次;

模型构建单元,用于利用每次获得的伪缺失样本数据、所述分布记录数据和多维时空数据构建增强回归树模型,进而得到一个基于多个增强回归树模型的机器学习集成模型。

10. 根据权利要求6所述的装置,其特征在于,所述确定模块包括:

第一确定单元,用于通过所述机器学习集成模型确定所述多维时空数据中各类数据的相对贡献率;

第二确定单元,用于利用所述机器学习集成模型在预设面积的格网预测得到多个入侵概率结果,其中,一个增强回归树模型对应一个入侵概率结果,逐栅格单元计算平均值,最终获取所述目标物种的全球入侵概率;

第三确定单元,用于将入侵概率结果大于预设概率值的区域确定为潜在入侵风险区;

第四确定单元,用于基于预测得到的多个入侵概率结果,逐栅格计算标准差,以分析预测结果的不确定性。

## 目标物种潜在入侵风险评估方法及装置

### 技术领域

[0001] 本发明涉及风险评估技术领域,尤其涉及一种目标物种潜在入侵风险评估方法及装置。

### 背景技术

[0002] 棕榈树的经济效益在世界范围内得到充分证明,包括提供食物、饮料、衣物、燃料、照明、药品、扫帚、篮子、垫子、建筑用木材、渔网和屋顶小屋材料。犀牛甲虫作为一种食腐鞘翅目类的入侵物种,成虫能够钻入棕榈树的顶端,穿过叶片的基部,进入展开叶片的心脏,对受感染棕榈树的生长点造成物理伤害,这种伤害可能导致棕榈树生长迟缓、幼叶变小或变形,如果攻击严重,则会导致棕榈树死亡。成虫通常攻击幼嫩的棕榈树,但在严重侵扰或没有幼嫩的棕榈树的情况下,成熟的棕榈树便会受到攻击。近些年,犀牛甲虫已入侵多个国家和地区,对全球棕榈业产业产生了重要的威胁。据估计,犀牛甲虫在马来西亚对当地棕榈树产业造成的经济损失达到约28.537亿美元。犀牛甲虫正在全球范围内导致其原生范围之外前所未有的经济损失。如果管理不当,可能会对全球棕榈树产业造成毁灭性打击。精准评估犀牛甲虫在全球的潜在入侵风险区是制定一套长期生态友好的主动管理策略的先决条件。

[0003] 以往有关犀牛甲虫的研究主要集中在管理策略、生物学特性和环境压力研究领域,主要侧重于被动纠正措施。与被动纠正措施相比,早期监测预警等主动管理技术在应对入侵物种挑战方面具有成本更低且更有效的优势。由于犀牛甲虫在全球的潜在入侵风险区缺乏科学的评估,导致针对犀牛甲虫的主动管理技术难以实施,无法为制定生态友好的主动管理策略提供支撑。目前,有关犀牛甲虫入侵风险区评估高度依赖于专家的先验知识构成的规则集,无法厘定犀牛甲虫入侵风险的关键影响因素,且无法绘制高空间分辨率的入侵概率以及量化评估不确定性。

### 发明内容

[0004] 为克服相关技术中存在的问题,本发明提供一种目标物种潜在入侵风险评估方法及装置。

[0005] 根据本发明实施例的第一方面,提供一种目标物种潜在入侵风险评估方法,所述方法包括:

获取目标物种的分布记录数据和多维时空数据;

根据所述分布记录数据和多维时空数据,使用最大熵模型生成伪缺失概率分布数据,并从所述伪缺失概率分布数据中随机筛选出与所述分布记录数据等量的伪缺失样本数据;

根据所述伪缺失样本数据、所述分布记录数据和多维时空数据,构建并训练得到一个基于多个增强回归树模型的机器学习集成模型;

通过所述机器学习集成模型,确定驱动所述目标物种入侵风险的关键影响因素,

确定所述目标物种的入侵概率,评估潜在入侵风险区分布和量化评估的不确定性。

[0006] 在一个实施例中,优选地,所述分布记录数据包括所述目标物种分布的经纬度信息,所述多维时空数据包括气候因素信息、地理因素信息和社会经济因素信息,其中,所述气候因素信息包括年平均气温、平均昼夜温度差、等温性、气温季节性变动系数、最热月份最高温度、最冷月份最低温度、最高温度与最低温度差、最湿季度平均温度、最干季度平均温度、最暖季度平均温度、最冷季度平均温度、年降雨量、最湿月份降水量、最干月份降水量、降水量季节性变化、最干季度降水量、最湿季度降水量、最暖季度降水量和最冷季度降水量,所述地理因素信息包括海拔、归一化植被指数和土地利用,所述社会经济因素信息包括城市可达性和道路密度。

[0007] 在一个实施例中,优选地,根据所述分布记录数据和多维时空数据,使用最大熵模型生成伪缺失概率分布数据,并从所述伪缺失概率分布数据中随机筛选出与所述分布记录数据等量的伪缺失样本数据,包括:

将所述分布记录数据和多维时空数据作为输入数据,运行最大熵模型,以生成介于0-1之间的环境适宜性概率;

将所述环境适宜性概率小于0.5作为所述伪缺失样本数据的筛选依据。

[0008] 在一个实施例中,优选地,根据所述伪缺失样本数据、所述分布记录数据和多维时空数据,构建并训练得到一个基于多个增强回归树模型的机器学习集成模型,包括:

将随机筛选出与所述分布记录数据等量的伪缺失样本数据的过程执行多次;

利用每次获得的伪缺失样本数据、所述分布记录数据和多维时空数据构建增强回归树模型,进而得到一个基于多个增强回归树模型的机器学习集成模型。

[0009] 在一个实施例中,优选地,通过所述机器学习集成模型,确定驱动所述目标物种入侵风险的关键影响因素,确定所述目标物种的入侵概率,评估潜在入侵风险区分布和量化评估的不确定性,包括:

通过所述机器学习集成模型确定所述多维时空数据中各类数据的相对贡献率;

利用所述机器学习集成模型在预设面积的格网预测得到多个入侵概率结果,其中,一个增强回归树模型对应一个入侵概率结果,逐栅格单元计算平均值,最终获取所述目标物种的全球入侵概率;

将入侵概率结果大于预设概率值的区域确定为潜在入侵风险区;

基于预测得到的多个入侵概率结果,逐栅格计算标准差,以分析预测结果的不确定性。

[0010] 根据本发明实施例的第二方面,提供一种目标物种潜在入侵风险评估装置,所述装置包括:

获取模块,用于获取目标物种的分布记录数据和多维时空数据;

筛选模块,用于根据所述分布记录数据和多维时空数据,使用最大熵模型生成伪缺失概率分布数据,并从所述伪缺失概率分布数据中随机筛选出与所述分布记录数据等量的伪缺失样本数据;

模型构建模块,用于根据所述伪缺失样本数据、所述分布记录数据和多维时空数据,构建并训练得到一个基于多个增强回归树模型的机器学习集成模型;

确定模块,用于通过所述机器学习集成模型,确定驱动所述目标物种入侵风险的

关键影响因素,确定所述目标物种的入侵概率,评估潜在入侵风险区分布和量化评估的不确定性。

[0011] 在一个实施例中,优选地,所述分布记录数据包括所述目标物种分布的经纬度信息,所述多维时空数据包括气候因素信息、地理因素信息和社会经济因素信息,其中,所述气候因素信息包括年平均气温、平均昼夜温度差、等温性、气温季节性变动系数、最热月份最高温度、最冷月份最低温度、最高温度与最低温度差、最湿季度平均温度、最干季度平均温度、最暖季度平均温度、最冷季度平均温度、年降雨量、最湿月份降水量、最干月份降水量、降水量季节性变化、最干季度降水量、最湿季度降水量、最暖季度降水量和最冷季度降水量,所述地理因素信息包括海拔、归一化植被指数和土地利用,所述社会经济因素信息包括城市可达性和道路密度。

[0012] 在一个实施例中,优选地,所述筛选模块用于:

将所述分布记录数据和多维时空数据作为输入数据,运行最大熵模型,以生成介于0-1之间的环境适宜性概率;

将所述环境适宜性概率小于0.5作为所述伪缺失样本数据的筛选依据。

[0013] 在一个实施例中,优选地,所述模型构建模块包括:

执行单元,用于将随机筛选出与所述分布记录数据等量的伪缺失样本数据的过程执行多次;

模型构建单元,用于利用每次获得的伪缺失样本数据、所述分布记录数据和多维时空数据构建增强回归树模型,进而得到一个基于多个增强回归树模型的机器学习集成模型。

[0014] 在一个实施例中,优选地,所述确定模块包括:

第一确定单元,用于通过所述机器学习集成模型确定所述多维时空数据中各类数据的相对贡献率;

第二确定单元,用于利用所述机器学习集成模型在预设面积的格网预测得到多个入侵概率结果,其中,一个增强回归树模型对应一个入侵概率结果,逐栅格单元计算平均值,最终获取所述目标物种的全球入侵概率;

第三确定单元,用于将入侵概率结果大于预设概率值的区域确定为潜在入侵风险区;

第四确定单元,用于基于预测得到的多个入侵概率结果,逐栅格计算标准差,以分析预测结果的不确定性。

[0015] 根据本发明实施例的第三方面,提供一种计算机可读存储介质,其上存储有计算机指令,该指令被处理器执行时实现第一方面中任一项方法的步骤。

[0016] 本发明的实施例提供的技术方案可以包括以下有益效果:

本发明实施例中,提出了一种基于多维时空数据和机器学习的目标物种潜在入侵风险评估方法,其基于目标物种的已知分布记录数据和多维时空数据,通过最大熵模型生成伪缺失概率分布数据;并随机筛选与已知分布记录数据等量的伪缺失样本数据,结合分布记录数据和多维时空数据,训练增强回归树模型,进而构建一个基于多个增强回归树模型的机器学习集成模型;通过机器学习集成模型,厘定驱动犀牛甲虫入侵风险的关键影响因素,绘制高空间分辨率的目标物种入侵概率,评估潜在入侵风险区分布,量化评估的不

确定性。

[0017] 应当理解的是,以上的一般描述和后文的细节描述仅是示例性和解释性的,并不能限制本发明。

### 附图说明

[0018] 此处的附图被并入说明书中并构成本说明书的一部分,示出了符合本发明的实施例,并与说明书一起用于解释本发明的原理。

[0019] 图1是根据一示例性实施例示出的一种目标物种潜在入侵风险评估方法的流程图。

[0020] 图2是根据一示例性实施例示出的一种目标物种潜在入侵风险评估方法中步骤S102的流程图。

[0021] 图3是根据一示例性实施例示出的一种目标物种潜在入侵风险评估方法中步骤S103的流程图。

[0022] 图4是根据一示例性实施例示出的一种目标物种潜在入侵风险评估方法中步骤S104的流程图。

[0023] 图5是根据一示例性实施例示出的一种犀牛甲虫潜在入侵风险评估方法的具体流程图。

[0024] 图6是根据一示例性实施例示出的一种目标物种潜在入侵风险评估装置的框图。

[0025] 图7是根据一示例性实施例示出的一种目标物种潜在入侵风险评估装置中模型构建模块的框图。

[0026] 图8是根据一示例性实施例示出的一种目标物种潜在入侵风险评估装置中模型构建模块的框图。

### 具体实施方式

[0027] 这里将详细地对示例性实施例进行说明,其示例表示在附图中。下面的描述涉及附图时,除非另有表示,不同附图中的相同数字表示相同或相似的要素。以下示例性实施例中所描述的实施方式并不代表与本发明相一致的所有实施方式。相反,它们仅是与如所附权利要求书中所详述的、本发明的一些方面相一致的装置和方法的例子。

[0028] 图1是根据一示例性实施例示出的一种目标物种潜在入侵风险评估方法的流程图,如图1所示,该方法包括:

步骤S101,获取目标物种的分布记录数据和多维时空数据;

步骤S102,根据所述分布记录数据和多维时空数据,使用最大熵模型生成伪缺失概率分布数据,并从所述伪缺失概率分布数据中随机筛选出与所述分布记录数据等量的伪缺失样本数据;

步骤S103,根据所述伪缺失样本数据、所述分布记录数据和多维时空数据,构建并训练得到一个基于多个增强回归树模型的机器学习集成模型;

步骤S104,通过所述机器学习集成模型,确定驱动所述目标物种入侵风险的关键影响因素,确定所述目标物种的入侵概率,评估潜在入侵风险区分布和量化评估的不确定性。

[0029] 在一个实施例中,优选地,所述分布记录数据包括所述目标物种分布的经纬度信息,所述多维时空数据包括气候因素信息、地理因素信息和社会经济因素信息,其中,所述气候因素信息包括年平均气温、平均昼夜温度差、等温性、气温季节性变动系数、最热月份最高温度、最冷月份最低温度、最高温度与最低温度差、最湿季度平均温度、最干季度平均温度、最暖季度平均温度、最冷季度平均温度、年降雨量、最湿月份降水量、最干月份降水量、降水量季节性变化、最干季度降水量、最湿季度降水量、最暖季度降水量和最冷季度降水量,所述地理因素信息包括海拔、归一化植被指数和土地利用,所述社会经济因素信息包括城市可达性和道路密度。

[0030] 图2是根据一示例性实施例示出的一种目标物种潜在入侵风险评估方法中步骤S102的流程图。

[0031] 如图2所示,在一个实施例中,优选地,上述步骤S102包括:

步骤S201,将所述分布记录数据和多维时空数据作为输入数据,运行最大熵模型,以生成介于0-1之间的环境适宜性概率;

步骤S202,将所述环境适宜性概率小于0.5作为所述伪缺失样本数据的筛选依据。

[0032] 图3是根据一示例性实施例示出的一种目标物种潜在入侵风险评估方法中步骤S103的流程图。

[0033] 如图3所示,在一个实施例中,优选地,上述步骤S103包括:

步骤S301,将随机筛选出与所述分布记录数据等量的伪缺失样本数据的过程执行多次;

步骤S302,利用每次获得的伪缺失样本数据、所述分布记录数据和多维时空数据构建增强回归树模型,进而得到一个基于多个增强回归树模型的机器学习集成模型。

[0034] 图4是根据一示例性实施例示出的一种目标物种潜在入侵风险评估方法中步骤S104的流程图。

[0035] 如图4所示,在一个实施例中,优选地,上述步骤S104包括:

步骤S401,通过所述机器学习集成模型确定所述多维时空数据中各类数据的相对贡献率;

步骤S402,利用所述机器学习集成模型在预设面积的格网预测得到多个入侵概率结果,其中,一个增强回归树模型对应一个入侵概率结果,逐栅格单元计算平均值,最终获取所述目标物种的全球入侵概率;

步骤S403,将入侵概率结果大于预设概率值的区域确定为潜在入侵风险区;

步骤S404,基于预测得到的多个入侵概率结果,逐栅格计算标准差,以分析预测结果的不确定性。

[0036] 下面以目标物种为犀牛甲虫为例,详细说明本发明的上述技术方案。

[0037] 如图5所示,犀牛甲虫的潜在入侵风险评估方法的具体流程包括:

(1) 获取全球已知犀牛甲虫分布记录数据

从全球生物多样性信息机构(GBIF,<http://www.gbif.org/>),欧洲和地中海植物保护组织(EPP0,<https://gd.eppo.int>)以及国际农业和生物科学中心(CABI,<http://www.cabi.org>),获取了犀牛甲虫的已知分布记录,并整理在线搜索获得的有关犀牛甲虫的已发表文献,最终形成一套全球已知犀牛甲虫分布记录数据,分布记录数据量达到295条。



[0038] (2) 多维时空数据获取及预处理

1) 多维时空数据获取:通过WorldClim获取了高质量高精度的气候因素;从航天飞机雷达地形测绘任务(<https://eosps0.gsfc.nasa.gov/missions/shuttle-radar-topography-mission>)、全球库存建模和绘图研究组(<https://iridl.ldeo.columbia.edu/>)和联合国粮食及农业组织(<https://www.fao.org/>)获得了三种地理因素,包括海拔、植被覆盖度(归一化植被指数)以及土地利用数据;从欧洲委员会联合研究中心(<https://forobs.jrc.ec.europa.eu/>)、开放式街道地图(<https://wordpress.org/>),下载得到两种社会经济因素,即城市可达性和道路分布数据,并通过核密度方法将道路分布数据转化为道路密度数据。

[0039] 2) 多维时空数据预处理:由于气候、地理和社会经济因素属于多源异构的时空数据,空间分辨率不一致。本发明选取WGS-84坐标系,以0.05度(约5公里)栅格单元为基准,将多维时空数据转化成统一的坐标系的基准栅格单元上,形成24个栅格图层,图层数据详见表1。

[0040] 表1多维时空数据概要

分类	环境变量
气候因素	年平均气温
	平均昼夜温度差
	等温性
	气温季节性变动系数
	最热月份最高温度
	最冷月份最低温度
	最高温度与最低温度差
	最湿季度平均温度
	最干季度平均温度
	最暖季度平均温度
	最冷季度平均温度
	年降雨量
	最湿月份降水量
	最干月份降水量
	降水量季节性变化
	最干季度降水量
	最湿季度降水量
	最暖季度降水量
	最冷季度降水量
	地理因素
归一化植被指数	
土地利用	
社会经济因素	城市可达性
	道路密度

### (3) 通过最大熵模型生成伪缺失样本

将犀牛甲虫分布记录数据和多维时空数据作为输入数据,运行最大熵模型,以0.5作为阈值,将小于该阈值作为犀牛甲虫的伪缺失样本筛选依据。

### [0041] (4) 构建增强回归树集成模型

随机筛选与犀牛甲虫发生记录等量的伪缺失样本数据(295个),结合犀牛甲虫分布记录数据(295个)和多维时空数据(如表1所示),构建增强回归树模型。在构建模型过程中,使用10折交叉验证方法以避免模型的过拟合。为了提升稳健性,将随机筛选与犀牛甲虫

发生记录等量的伪缺失样本数据的过程执行300次,进而获得了一个基于300个增强回归树模型的机器学习集成模型。本发明中,以R Version 3.3.3统计编程环境为基础,采用“dismo”扩展包搭建建模分析流程。使用曲线下面积(Area under the curve, AUC)作为集成模型的精度评价指标,10折交叉验证过程中的AUC数值达到 $0.980 \pm 0.008$ 。

[0042] (5) 使用机器学习集成模型输出结果

1) 厘定关键因素:多维时空数据的相对贡献率详见表2。表2表明,城市可达性、最高温度与最低温度差、道路密度、气温季节性变动系数、最冷月份最低温度、海拔、土地利用、最暖季度降水量、等温性和归一化植被指数是驱动犀牛甲虫入侵的关键因素。

[0043] 表2多维时空数据相对贡献率大小

分类	环境变量	平均贡献率 (%)
气候因素	年平均气温	——
	平均昼夜温度差	——
	等温性	3.071
	气温季节性变动系数	13.008
	最热月份最高温度	——
	最冷月份最低温度	8.756
	最高温度与最低温度差	19.445
	最湿季度平均温度	——
	最干季度平均温度	——
	最暖季度平均温度	——
	最冷季度平均温度	——
	年降雨量	——
	最湿月份降水量	——
	最干月份降水量	——
	降水量季节性变化	——
	最干季度降水量	——
	最湿季度降水量	——
	最暖季度降水量	3.271
	最冷季度降水量	——
地理因素	海拔	6.565
	归一化植被指数	2.165
	土地利用	3.873
社会经济因素	城市可达性	21.025
	道路密度	15.516

其中,贡献率大于1%的要素是关键因素,贡献率等于或小于1%代表不显著,用“——”表示。

[0044] 2) 预测入侵概率:利用集成模型,在约5公里格网上生成犀牛甲虫在全球的入侵概率,预测得到300个入侵概率结果,逐栅格单元计算平均值,最终获得犀牛甲虫的全球入侵概率,概率值范围介于[0,1]。

[0045] 3) 评估入侵风险:以0.5作为阈值,判定犀牛甲虫在全球的入侵风险区,结果表明,

犀牛甲虫的入侵风险区主要集中在南美洲北部、非洲的中西部及东部部分区域、亚洲的南部和东南部区域。

[0046] 4) 量化不确定性: 基于预测的300个入侵概率结果, 逐栅格单元计算标准差, 分析评估的不确定性总体来看, 评估不确定性较低。

[0047] 图6是根据一示例性实施例示出的一种目标物种潜在入侵风险评估装置的框图。

[0048] 如图6所示, 根据本发明实施例的第二方面, 提供一种目标物种潜在入侵风险评估装置, 所述装置包括:

获取模块61, 用于获取目标物种的分布记录数据和多维时空数据;

筛选模块62, 用于根据所述分布记录数据和多维时空数据, 使用最大熵模型生成伪缺失概率分布数据, 并从所述伪缺失概率分布数据中随机筛选出与所述分布记录数据等量的伪缺失样本数据;

模型构建模块63, 用于根据所述伪缺失样本数据、所述分布记录数据和多维时空数据, 构建并训练得到一个基于多个增强回归树模型的机器学习集成模型;

确定模块64, 用于通过所述机器学习集成模型, 确定驱动所述目标物种入侵风险的关键影响因素, 确定所述目标物种的入侵概率, 评估潜在入侵风险区分布和量化评估的不确定性。

[0049] 在一个实施例中, 优选地, 所述分布记录数据包括所述目标物种分布的经纬度信息, 所述多维时空数据包括气候因素信息、地理因素信息和社会经济因素信息, 其中, 所述气候因素信息包括年平均气温、平均昼夜温度差、等温性、气温季节性变动系数、最热月份最高温度、最冷月份最低温度、最高温度与最低温度差、最湿季度平均温度、最干季度平均温度、最暖季度平均温度、最冷季度平均温度、年降雨量、最湿月份降水量、最干月份降水量、降水量季节性变化、最干季度降水量、最湿季度降水量、最暖季度降水量和最冷季度降水量, 所述地理因素信息包括海拔、归一化植被指数和土地利用, 所述社会经济因素信息包括城市可达性和道路密度。

[0050] 在一个实施例中, 优选地, 所述筛选模块62用于:

将所述分布记录数据和多维时空数据作为输入数据, 运行最大熵模型, 以生成介于0-1之间的环境适宜性概率;

将所述环境适宜性概率小于0.5作为所述伪缺失样本数据的筛选依据。

[0051] 图7是根据一示例性实施例示出的一种目标物种潜在入侵风险评估装置中模型构建模块的框图。

[0052] 如图7所示, 在一个实施例中, 优选地, 所述模型构建模块63包括:

执行单元71, 用于将随机筛选出与所述分布记录数据等量的伪缺失样本数据的过程执行多次;

模型构建单元72, 用于利用每次获得的伪缺失样本数据、所述分布记录数据和多维时空数据构建增强回归树模型, 进而得到一个基于多个增强回归树模型的机器学习集成模型。

[0053] 图8是根据一示例性实施例示出的一种目标物种潜在入侵风险评估装置中模型构建模块的框图。

[0054] 如图8所示, 在一个实施例中, 优选地, 所述确定模块64包括:

第一确定单元81,用于通过所述机器学习集成模型确定所述多维时空数据中各类数据的相对贡献率;

第二确定单元82,用于利用所述机器学习集成模型在预设面积的格网预测得到多个入侵概率结果,其中,一个增强回归树模型对应一个入侵概率结果,逐栅格单元计算平均值,最终获取所述目标物种的全球入侵概率;

第三确定单元83,用于将入侵概率结果大于预设概率值的区域确定为潜在入侵风险区;

第四确定单元84,用于基于预测得到的多个入侵概率结果,逐栅格计算标准差,以分析预测结果的不确定性。

[0055] 根据本发明实施例的第三方面,提供一种计算机可读存储介质,其上存储有计算机指令,该指令被处理器执行时实现第一方面中任一项方法的步骤。

[0056] 本发明由于采取以上技术方案,具有以下优点:

(1)以往有关犀牛甲虫入侵风险的评估高度依赖于专家的先验知识构成的规则集,不能识别影响犀牛甲虫入侵的关键因素,且无法在全球尺度上绘制犀牛甲虫入侵风险分布状况。本方案相较于基于先验知识的方法,厘定出城市可达性、最高温度与最低温度差、道路密度、气温季节性变动系数、海拔、土地利用和归一化植被指数等要素是驱动犀牛甲虫入侵的关键因素,与此同时,能够在高空间分辨率格网上绘制犀牛甲虫的全球入侵风险。

[0057] (2)以往基于先验知识的犀牛甲虫入侵风险评估方法,缺少准确性和不确定性的评估,本方案采用AUC指标和标准差计算方法评价入侵风险评估的准确性和不确定性,使得评估结果更加科学合理。

[0058] 进一步可以理解的是,本发明中“多个”是指两个或两个以上,其它量词与之类似。“和/或”,描述关联对象的关联关系,表示可以存在三种关系,例如,A和/或B,可以表示:单独存在A,同时存在A和B,单独存在B这三种情况。字符“/”一般表示前后关联对象是一种“或”的关系。单数形式的“一种”、“”和“该”也旨在包括多数形式,除非上下文清楚地表示其他含义。

[0059] 进一步可以理解的是,术语“第一”、“第二”等用于描述各种信息,但这些信息不应限于这些术语。这些术语仅用来将同一类型的信息彼此区分开,并不表示特定的顺序或者重要程度。实际上,“第一”、“第二”等表述完全可以互换使用。例如,在不脱离本发明范围的情况下,第一信息也可以被称为第二信息,类似地,第二信息也可以被称为第一信息。

[0060] 进一步可以理解的是,本发明实施例中尽管在附图中以特定的顺序描述操作,但是不应将其理解为要求按照所示的特定顺序或是串行顺序来执行这些操作,或是要求执行全部所示的操作以得到期望的结果。在特定环境中,多任务和并行处理可能是有利的。

[0061] 本领域技术人员在考虑说明书及实践这里公开的发明后,将容易想到本发明的其它实施方案。本申请旨在涵盖本发明的任何变型、用途或者适应性变化,这些变型、用途或者适应性变化遵循本发明的一般性原理并包括本发明未公开的本技术领域中的公知常识或惯用技术手段。说明书和实施例仅被视为示例性的,本发明的真正范围和精神由下面的权利要求指出。

[0062] 应当理解的是,本发明并不局限于上面已经描述并在附图中示出的精确结构,并

且可以在不脱离其范围进行各种修改和改变。本发明的范围仅由所附的权利要求来限制。

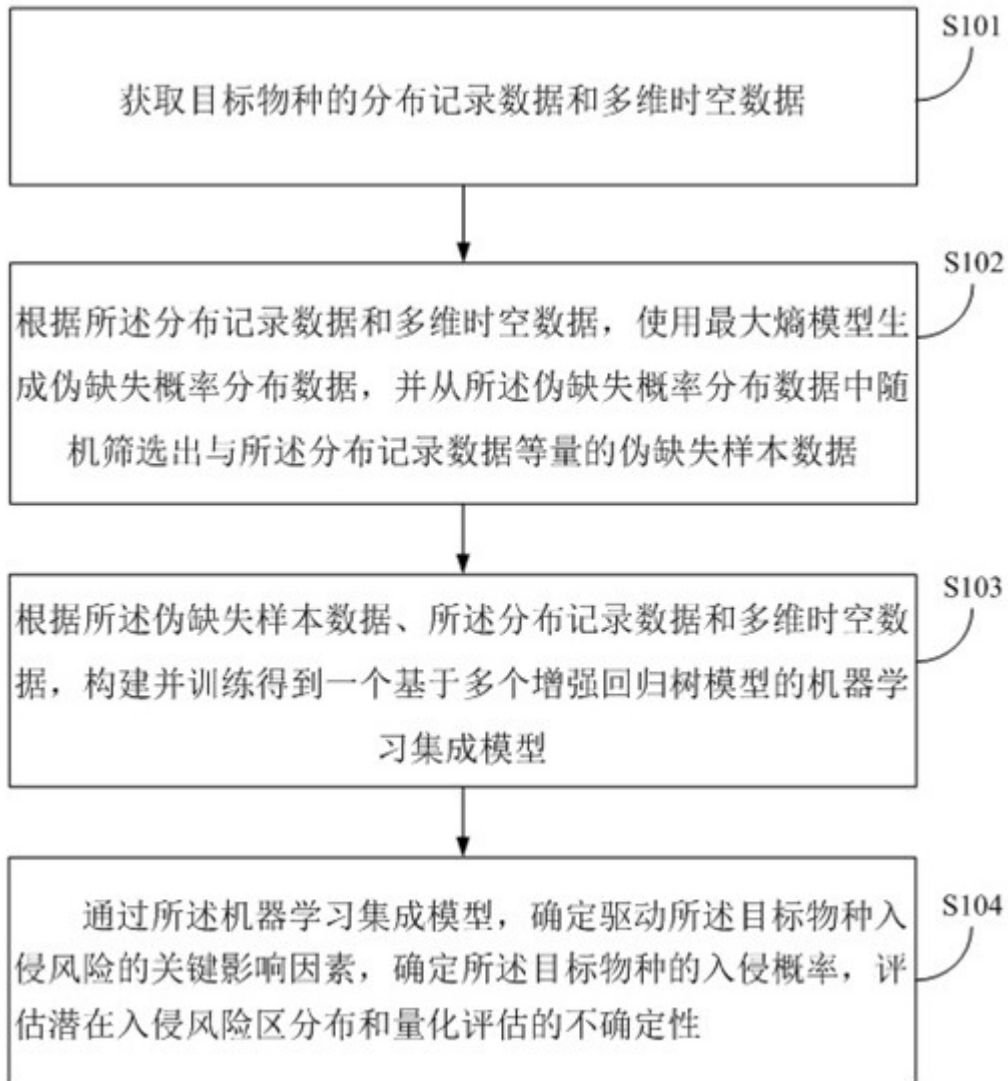


图1



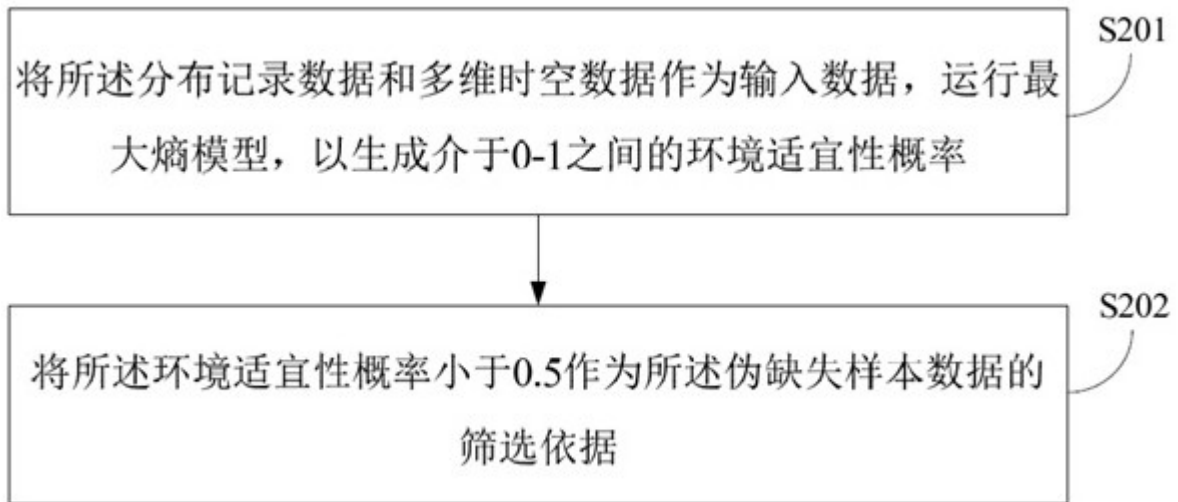


图2

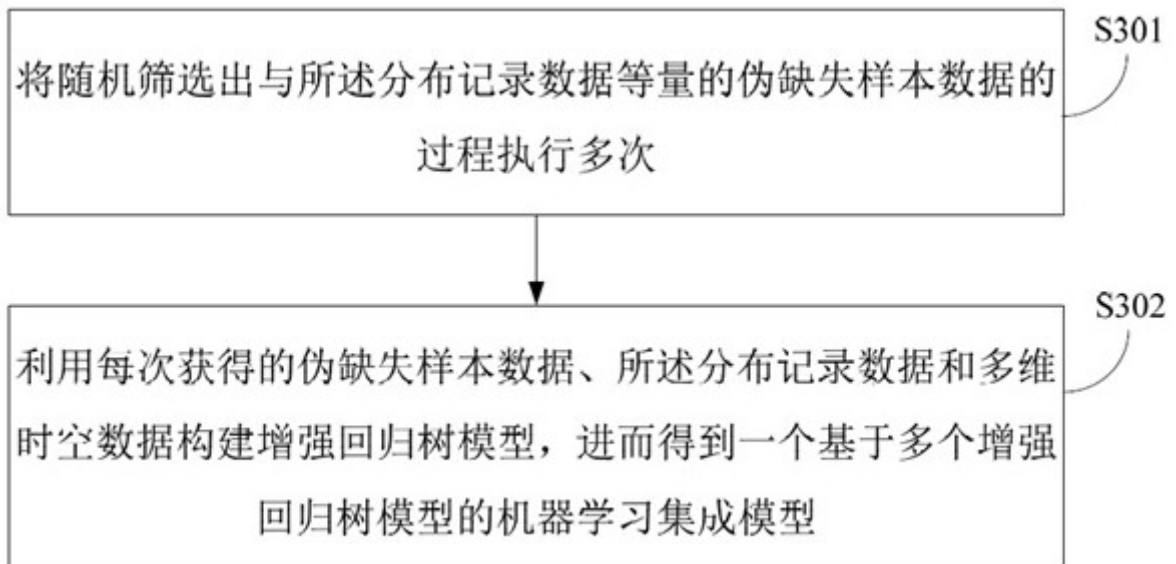


图3

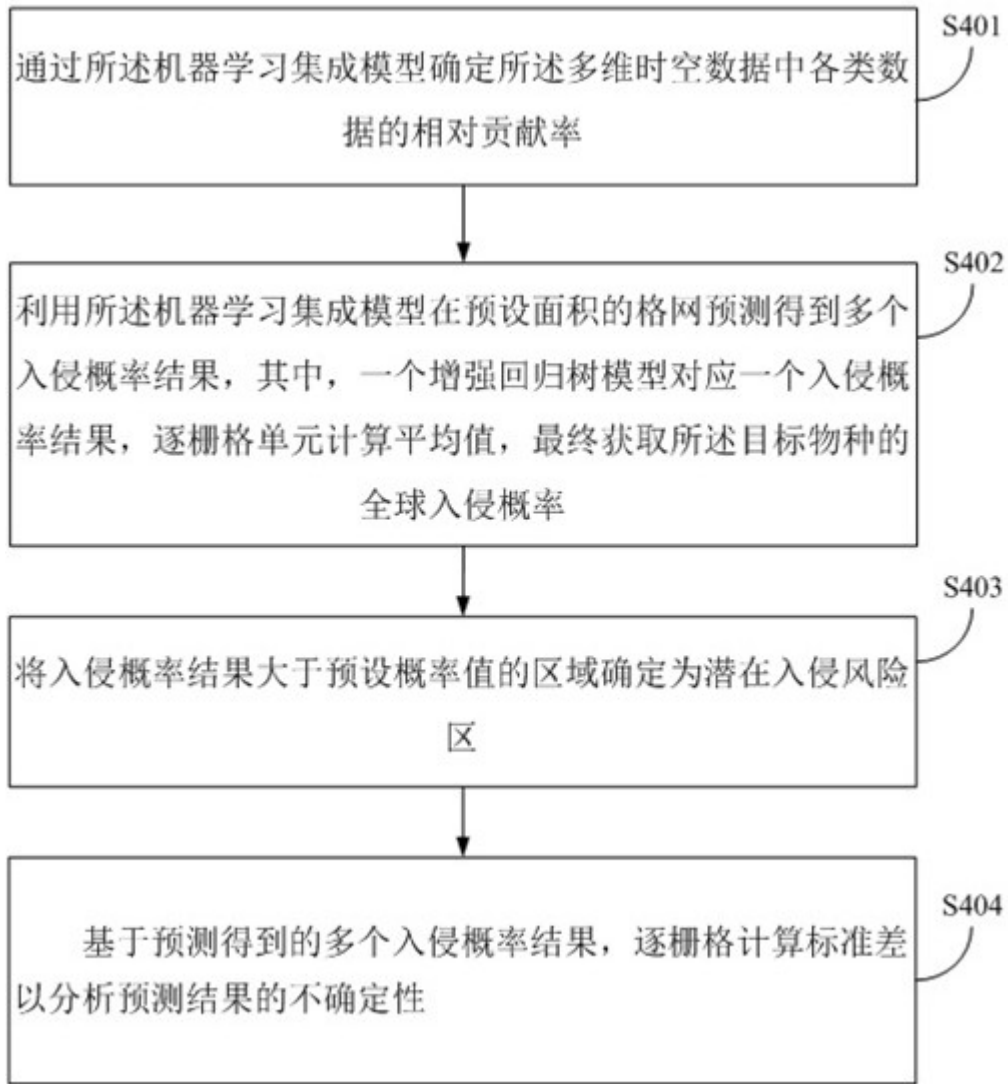


图4

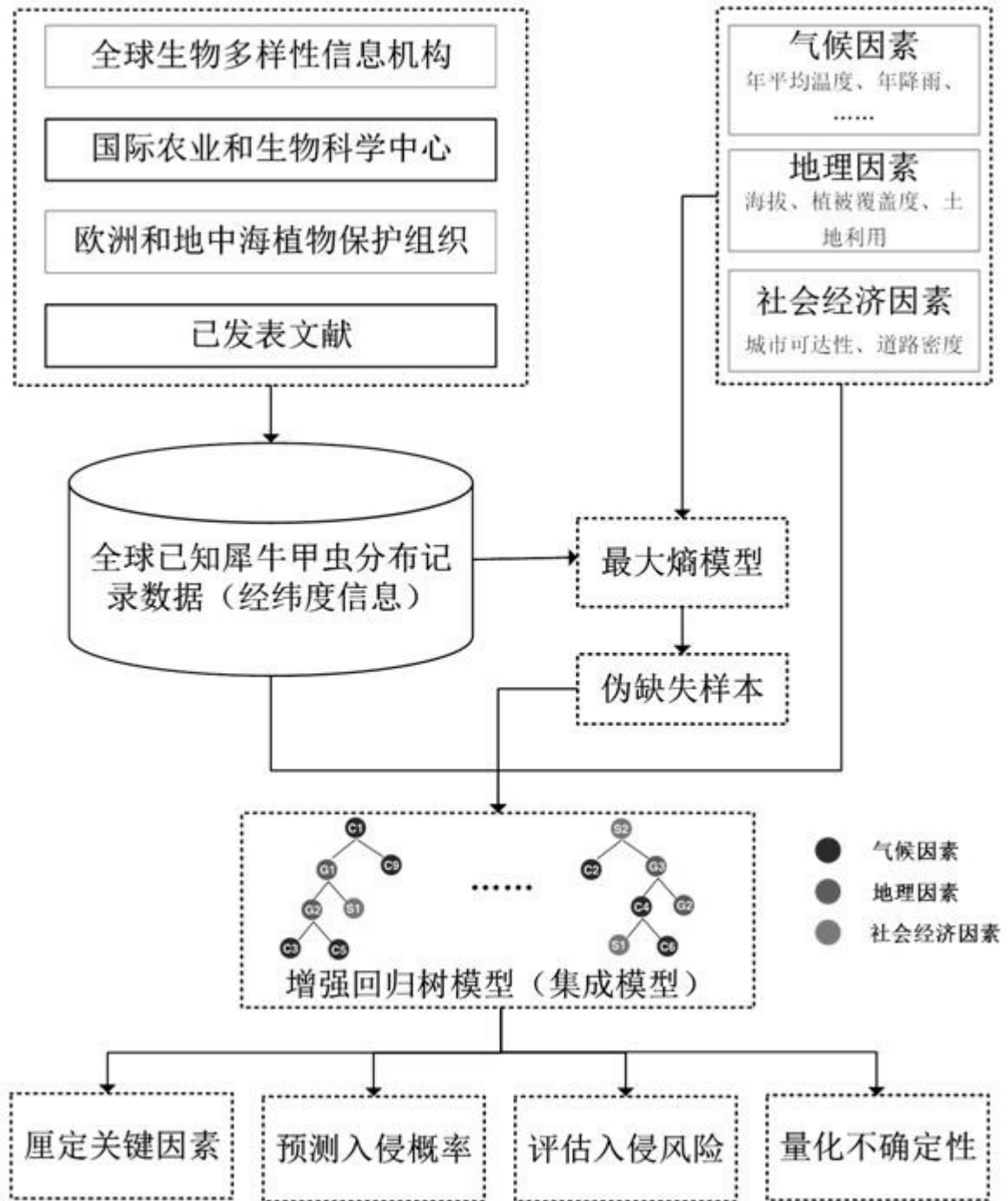


图5

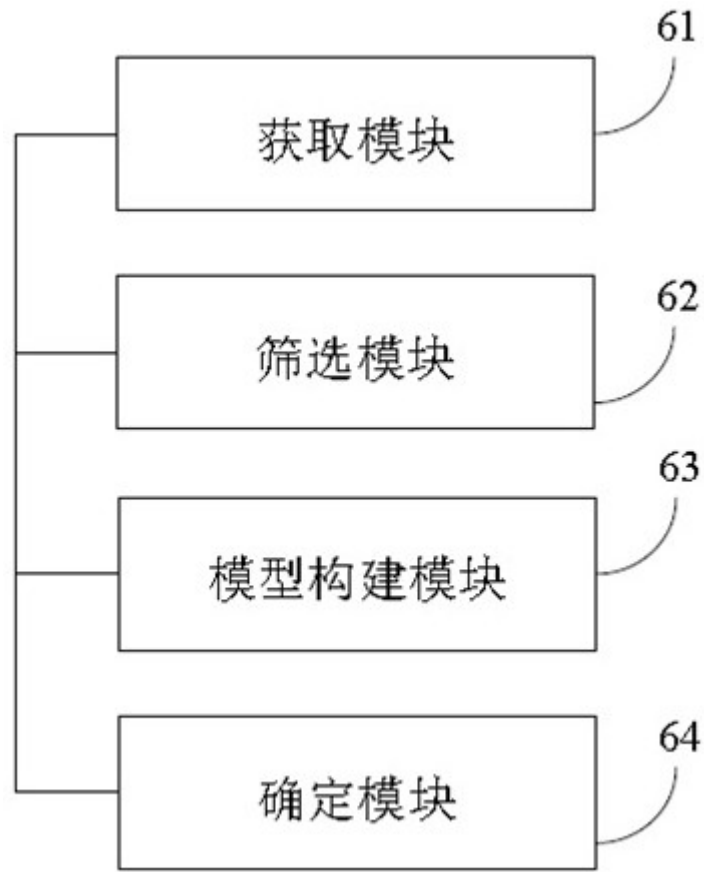


图6

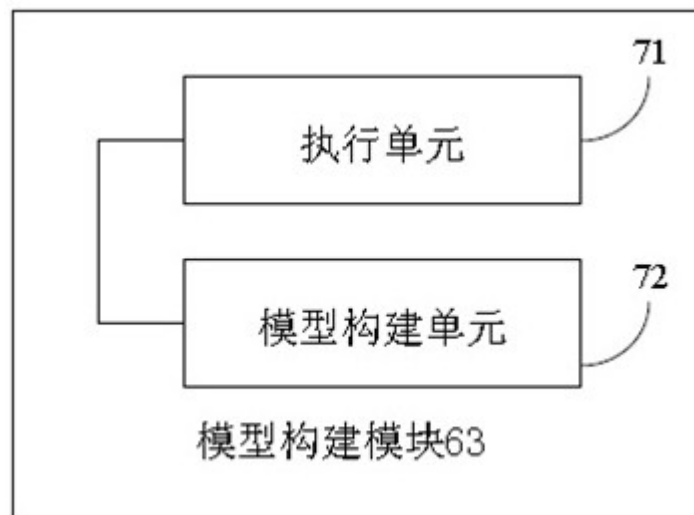


图7

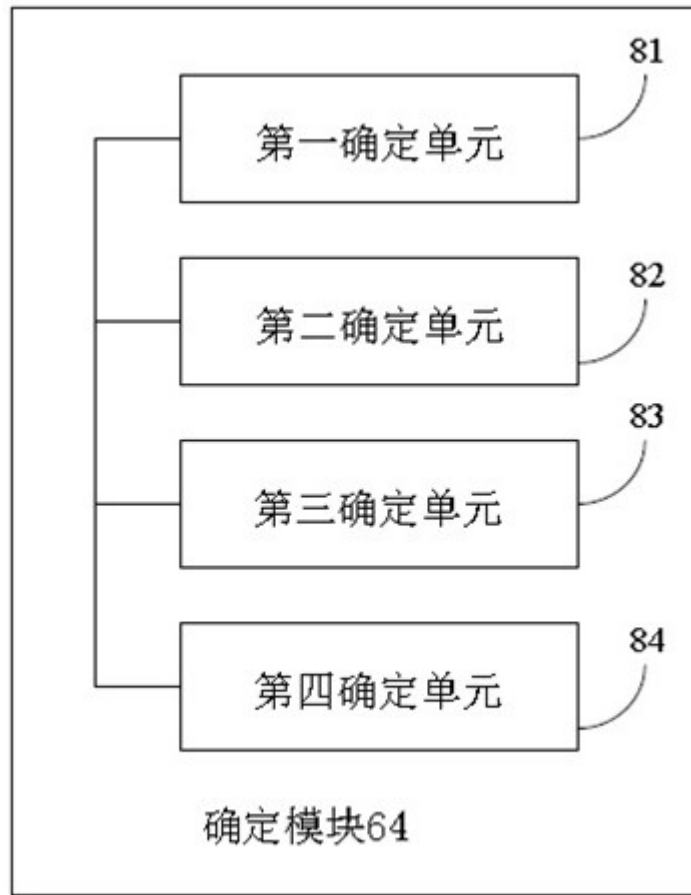


图8