

Design and Optimization of Alien Pest Intelligent Identification System Based on Artificial Intelligence Technology

Jiandong Li Min Liu Xiuhuan Zhu Fei Wang Zhiping Su*

Jinan Customs Technology Center, Jinan, Shandong, 250014, China

Abstract

Establishing an intelligent identification system and optimizing its performance is crucial for foreign pests that pose a serious threat to the ecosystem and economy. The intelligent identification system of alien pests is optimized and is based on artificial intelligence technology. The rationale for the identification system and the current technical challenges are outlined and existing solutions evaluated. The key steps such as data acquisition and preprocessing, feature extraction and selection, and the design and optimization of intelligent identification algorithm are introduced in detail, and these key steps are put forward as the system performance assessment indicators, the real-time and scalability of the assessment system.

Keywords

foreign pests; AI; intelligent identification system; feature extraction

基于人工智能技术的外来有害生物智能识别系统设计与优化

李建勇 刘敏 朱秀焕 王飞 栗智平*

济南海关技术中心, 中国·山东 济南 250014

摘要

建立智能识别系统, 优化其性能, 对生态系统和经济构成严重威胁的外来有害生物至关重要。设计优化了外来有害生物智能识别系统, 基于人工智能技术。对识别系统和当前技术挑战的基本原理进行了概述, 对已有的解决方案进行了评估。对数据采集与预处理、特征提取与选择、智能识别算法的设计与优化等关键步骤进行了详细地介绍, 并对这些关键步骤进行了提出制度绩效考核指标, 考核制度的实时性和可扩展性。

关键词

外来有害生物; 人工智能; 智能识别系统; 特征提取

1 引言

外来有害生物对生态环境和经济发展构成了严重威胁, 其入侵往往导致生物多样性丧失、农作物减产、生态系统功能紊乱等问题。在面对这一挑战时, 传统的监测与识别方法往往存在效率低下、准确率不高的问题。而基于人工智能技术的智能识别系统则具有很大潜力, 能够提高外来有害生物

的识别准确性和效率。论文旨在设计并优化一种基于人工智能技术的外来有害生物智能识别系统, 以提高对外来有害生物的管理水平。通过综合运用数据采集、特征提取、智能算法等关键技术, 本研究旨在为应对外来有害生物的挑战提供有效解决方案, 推动生态保护和可持续发展。

2 外来有害生物智能识别系统概述

2.1 识别系统基本原理

外来有害生物智能识别系统的基本原理是通过人工智能技术对采集到的外来有害生物数据进行分析 and 处理, 尤其是机器学习和深度学习算法, 实现智能识别不同种类的外来有害生物。该系统对数据进行采集和预处理, 包括以多种形式获取图像、声音和外来有害生物的生物特征等数据, 并对数据进行清洗、标记和规范处理, 确保数据的质量和一致性。一个常见的特征提取方法, 离散傅立叶变换 (Discrete Fourier Transform, DFT) 的数学表达式:

【基金项目】山东省重点研发计划(重大科技创新工程)(项目编号: 2021CXGC011306)。

【作者简介】李建勇(1970-), 男, 中国山东济南人, 博士, 研究员, 从事植物检疫及实验室质量管理研究。

【通讯作者】栗智平(1971-), 男, 中国山东济南人, 硕士, 正高级农艺师, 从事生物物种鉴定、植物检疫、转基因检测等研究。

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-i2\pi nk/N} \quad (1)$$

其中, $x(n)$ 为输入信号的离散样本; N 为信号的长度; $x(k)$ 为傅立叶变换后得到的频域表示。通过这样的公式, 可以将信号从时域转换到频域, 提取出不同频率成分的特征, 为后续的分类识别提供依据。该系统从海量数据中提取有效的特征表示, 采用传统的特征提取方法和基于深度学习的特征提取方法, 为后续的分类识别提供依据。为了建立准确的分类器或识别模型, 实现对外来有害生物的智能识别和分类, 系统采用机器学习算法或深度学习模型进行训练和优化。

2.2 技术挑战与现有解决方案评估

外来有害生物智能识别系统面临着诸多技术挑战, 外来有害生物的种类繁多, 形态各异, 加之受到环境因素的影响, 使得其特征表达复杂多样, 如何准确提取有效特征是一个关键问题。外来有害生物数据的获取和标注需要大量的人力和时间成本, 而且数据标注的准确性对系统性能有着直接影响。外来有害生物的智能识别系统需要具备良好的实时性和可扩展性, 能够适应不同环境和场景的应用需求。针对这些挑战, 现有的解决方案主要包括基于传统机器学习的方法和基于深度学习的方法。主要依靠手工设计的特征和经典的分类器来识别, 对复杂多变的外来有害生物识别任务效果有限, 尽管在一些简单的场景中表现不错。则可以更好地适应复杂环境下的识别任务, 并通过深度神经网络自动学习特征表示和分类模型取得一定效果。深度学习模型需要大量的标记数据和计算资源, 模型的可解释性较差, 这也使其在实际应用中的推广和应用受到一定的限制, 深度学习模型在应用过程中未来国外有害生物智能识别系统研究的一个重要方向是如何综合利用传统方法和深度学习方法, 克服数据标注费用高、模型可解释性差等问题, 提高系统的性能和实用性。

3 数据采集与预处理

3.1 数据来源与获取

包括但不限于传感器监测、图像采集、声音录制等生物特征的采集, 外来有害生物智能识别系统的数据来源是多样化的。通过布设传感器网络实时监测外来有害生物生态环境, 获取环境参数变化信息, 是传感器监测常用的数据来源方式。图像采集则是将外来有害生物的外观特征通过照相机等设备拍摄下来并记录下来。声音录音则是利用话筒等设备, 捕捉外来有害生物的声音特征, 特别是对鸣叫声特殊的鸟类、昆虫等生物, 效果更好^[1]。外来有害生物的位置信息和迁徙轨迹也可通过 GPS 追踪和生物标记等手段获取。

3.2 数据质量评估与处理

数据缺失、噪声干扰、数据偏差等数据采集过程中可能面临的各种质量问题, 评估和处理数据质量是数据预处理的重要环节。主要包括对数据完整性、一致性、准确性、及

时性等方面的评估, 可以通过统计分析、可视化分析等方式进行数据质量评估。针对数据质量问题, 可采用数据清洗、异常值处理、数据填充等方法处理, 使数据的质量得到提高, 可用性得到提高, 也可以扩充数据集, 增加数据的多样性与丰富性, 并利用数据强化技术与合成数据等方法提升模型的泛化能力。

3.3 数据标注与标准化

数据标注是指人工标注采集到的数据, 并添加相应的标签或标注信息, 以便进行后续的模式训练与评价。数据标注的好坏直接影响到车型的性能和概括能力, 所以标注的精确性和连贯性都是需要严格把控的。在进行数据标注时, 为了确保标注结果的可信度和可用性, 通常需要制定标注规范和标准, 明确标注对象和方式, 并对标注人员进行培训和质量把关。为了提高数据的可比性和通用性, 还需要在不同系统和平台之间进行数据交互和共享, 对数据进行包括统一规范的数据格式、数据单位、数据命名等方面的标准化处理。

4 特征提取与选择

4.1 传统特征提取方法

传统特征提取方法是指将能够描述外来有害生物特征的数值或向量表示, 通过对数据的预处理和特征工程, 基于领域知识和手工设计的方法提取出来^[1]。常见的传统特色有造型特色, 纹路特色, 色彩特色等。其中形态特征主要描述大小、形态等外来有害生物的外在形态特征; 肌理特征描述肌理分布及外来有害生物表面特征; 颜色特征是对外来有害生物表面颜色分布及其特征的描述。传统的特征提取方法通常需要选择合适的特征和特征表现方式, 结合专家在领域的经验和知识, 通过特征选择、降维等方法进一步加工, 使特征的分度度和分类性能得到提高。

4.2 基于深度学习的特征提取方法

基于深度学习的特征提取方法是指利用深度神经网络自动学习数据中的高级特征表示的方法, 深度学习模型通常包括多层神经网络, 能够从原始数据中学习不同抽象层次的特征表示, 具有很强的表征学习能力和泛化能力。常用的深度学习模型包括卷积神经网络 (CNN)、循环神经网络 (RNN) 等。一个常用于描述卷积神经网络 (CNN) 中的卷积操作的数学公式:

$$y_j = \sigma \left(\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} x_i + m, j + n \cdot w_{m,n} + b \right) \quad (2)$$

其中, y_j 为卷积后的输出特征图中的一个元素; $x_{i+m, j+n}$ 为输入特征图中的一个元素; $w_{m,n}$ 为卷积核中的权重; b 为偏置项; σ 为激活函数。通过这样的卷积操作, CNN 可以从原始数据中提取出不同层次的特征表示, 用于后续的分类识别任务。在外来有害生物智能识别系统中, 可以利用预训练的深度学习模型, 如 ImageNet 上训练的 CNN 模型, 通过微调或迁移学习的方式, 对外来有害生物数据进行特征提取。

4.3 特征选择与降维策略

在特征提取过程中,经常会面临需要进行特征选择和降维处理以提高模型的效率和泛化能力的维度灾难和特征冗余问题。特征选择是指为了保留最重要的特征信息,减少特征维度和模型复杂度,从原始特征集中选择一些具有代表性的特征子集。常用的特征选择方式有滤镜、封装、嵌入法等几种。滤镜主要是根据相关系数、相互信息等特征与目标的关联性来选择^[2];包装法则是通过构建模型,如基于遗传算法的递归特征消除、特征选择等进行特征评估和选择;嵌入法则是在模型训练过程中嵌入特征选择,如拉索回归,决策树等。

5 智能识别算法设计与优化

5.1 机器学习算法选择与比较

外来有害生物智能识别系统中,选择恰当的机器学习算法是关键的一步,目前比较常用的有支持向量机随机森林和K近邻算法(K-Nearest Neighbors)^[2]。以上三种算法各自有其特点和适用场景。例如,支持向量机适用于高维数据和线性不可分问题,通过核函数将数据映射到高维空间进行分类;随机森林属于集成学习算法,将多个决策树组合在一起来提高分类性能和鲁棒性;K近邻算法是一种以实例为基础的学习法,通过计算样本之间的距离来进行分类。在选择算法的时候,要结合实际问题进行选择和比较,以达到最好的分类效果。

5.2 深度学习模型设计与优化

深度学习模型是近年来在外来有害生物智能识别领域取得显著成效的关键技术之一,深度学习模型通常包括多层神经网络,能够自动学习数据中的高级特征表示,具有很强的表征学习能力和泛化能力。在设计深度学习模型时,需要考虑网络的结构、层数、激活函数等因素,并通过反向传播算法和优化器对模型参数进行更新和优化^[3]。还可以利用正则化技术、批归一化等方法来提高模型的泛化能力和抗过拟合能力。

5.3 模型融合与集成策略

模型融合与集成是一种常用的提高模型性能和鲁棒性的策略,常用的模型融合方法包括投票法、堆叠法等。投票法是指通过组合多个基础模型的预测结果来进行最终的分类决策,可以降低模型的方差和提高模型的稳定性;堆叠法则是通过训练一个元模型来组合多个基础模型,从而进一步提高模型的性能和泛化能力。在外来有害生物智能识别系统中,可以利用模型融合和集成策略来综合利用不同模型的优势,进一步提高识别准确率和鲁棒性。

6 系统性能评估与指标

6.1 识别准确率评估方法

识别准确率是评估外来有害生物智能识别系统性能的重要指标之一,通常通过混淆矩阵、准确率、召回率、F1

值等指标来进行评估。混淆矩阵是一种二维表格,用于比较分类模型预测结果和实际标签的一致性,其中包括真正例(True Positive)、真负例(True Negative)、假正例(False Positive)和假负例(False Negative)四个方面。准确率指标是指分类器正确分类的样本数占总样本数的比例,是评估分类模型整体性能的常用指标;召回率指标是指分类器正确识别正样本的能力,即正样本中被正确分类的比例;F1值则是准确率和召回率的调和平均值,综合考虑了分类器的精确度和召回率,适用于不平衡类别分布的情况。

6.2 系统响应速度与效率评估

通常通过系统的响应时间、处理速度、资源消耗等指标来评价外来有害生物智能识别系统的实用性和可操作性,系统响应速度和效率是其中一项重要指标。系统的响应时间直接影响到系统的实时性和用户体验,是指系统从接收到输入数据到输出识别结果的时间间隔;处理速度是指反映计算效率和系统性能的系统在单位时间内对数据进行处理的能力;资源消耗是指CPU占用、内存占用等系统运行过程中消耗的计算资源和存储资源。

6.3 实时性与可扩展性评估

实时性与可扩展性是评估外来有害生物智能识别系统在不同场景下适用性和可操作性的重要指标,实时性指系统对输入数据能够及时、快速地进行处理和识别的能力,尤其适用于对外来有害生物进行实时监测和预警的场景。在评估实时性时,需考虑系统的响应时间、处理速度和数据处理能力。可扩展性指系统能够灵活地扩展和适应不同规模和复杂度的数据和任务的能力,适用于不同环境和需求的应用场景。在评估可扩展性时,需考虑系统的并发处理能力、数据处理能力和分布式计算能力等因素。综合考虑实时性和可扩展性,需要进行系统性能测试和性能分析,以验证系统在实际应用中的性能表现。

7 结语

文章设计优化了外来有害生物智能识别系统,基于人工智能技术,深入探讨分析了关键步骤。通过对数据采集与预处理、特征提取与选择、智能识别算法设计与优化等关键技术的综合运用,建立了有效的外来有害生物智能识别系统,目前,国外有害生物智能识别系统已具备该系统对外来有害生物的监测和管理提供了重要的技术支持,提高了识别的准确性,也具有很高的实时性和可扩展性。

参考文献

- [1] 杨博森.计算机人工智能识别技术的应用瓶颈探析[J].数字通信世界,2024(1):110-112.
- [2] 周渊,张天驰.AI无处不在,促进产业改变生活[N].文汇报,2023-07-08(006).
- [3] 张宁.探究计算机信息技术在人工智能发展中的应用[J].信息记录材料,2024(2):165-167.