Population propagation model based on **Holt-Winters** and **Decision Tree Classification**

摘要：

2019年9月，于加拿大不列颠哥伦比亚省温哥华岛发现的野黄蜂引起邻近的华盛顿州发生害虫目击事件。现有公众提供的信息报告，及害虫的图片集。我们的任务是帮助当地政府分析和解释该信息集的数据，并挖掘出其背后所反映的野黄蜂的生物特征和基于时空的种群繁衍传播规律。近而为当地政府对信息集进行真伪判断、是否进行额外调查及其优先度，提供一套较为准确、合理的决策标准。此外，我们也要提出为保证该标准的适宜性的模型自我更新及更新频率的方式。

首先，我们基于报告中的害虫经纬度位置信息可判断蜂群小范围集中分布；利用简单的灰度预测模型对已知野黄蜂的经纬度位置进行分析；利用具有趋势性和季节性预测特征的Holt-Winters时序模型来建立对已知野黄蜂传播范围的分析策略。得出蜂群随时间向内陆传播。

其次，我们对报告中的‘Lab Status’为‘Positive ID’及‘Negative ID’的数据进行以下处理：先划分训练集和测试集，用来训练出基于支持向量机(SVM)多分类和决策树模型和基于GoogleNet的图像分类模型，至此，我们得到三个用于决策标准的分类模型，并在测试集中测试准确度和召回率。

Then,我们建立一套基于上述三个分类模型的评价标准,包括对报告中图片的蜂的尺寸、密度、完整度、清晰度的因素考虑。

最后,我们解释模型的更新机制、复杂度、时效性和不同地区适应性。大黄蜂被根除的直接表现是大黄蜂的数量减少至模型认定的安全范围甚至接近于零，并持续一个预测周期。

Population propagation model based on **Holt-Winters** and **Decision Tree Classification**

In September 2019, bumblebees found on Vancouver Island, British Columbia, Canada, caused a pest sighting incident in nearby Washington state. Information reports provided by the public and photo collections of pests are available. Our task is to help the local government to analyze and interpret the values of the information set, and to find out the biological characteristics of bumblebees and the laws of population reproduction and propagation based on time and space.Otherwise,We should also provides a set of more accurate and reasonable decision-making standards for the local government to judge the authenticity of the information set, whether to conduct additional investigation and its priority. In addition, it is also important to propose a self -renewal mode of models to ensure the suitability of the standards.

Firstly, based on the longitude and latitude position information of pests in the information set, we can judge the small-scale centralized distribution of bumblebees, using simple GM(1,1) model to analyze the longitude and latitude position, and using Holt-Winters time series model with trend and seasonal prediction characteristics to establish the analysis strategy of known bumblebees' transmission range. It is concluded that the colony spreads inland with time.

Secondly, we deal with the data with "lab status" as "positive ID" and "negative ID" in the report as follows:firstly, the training set and the test set are divided to train the multi classification and decision tree model based on support vector machine (**SVM**) and the image classification model based on **GoogleNet**. So far, we get three classification models for decision criteria, and test the accuracy in the test set.

Then, we set up a set of evaluation criteria based on the above three classification models, including the size, density, integrity and clarity of the bees in the report’s images.

Finally, we explain the updating mechanism, complexity, timeliness and adaptability of the model. The direct manifestation of the eradication of bumblebees is that the number of bumblebees is reduced to the safe range determined by the model, even close to zero, and lasts for a prediction cycle.

Keywords：GM(1,1) , Holt-Winters , SVM, Decision Tree Classification , GoogleNet , Difference equation