Problem Chosen	2022 MCM/ICM Summary Sheet	Team Control Number
C		2019057552

Fight for Washington State: Can Artificial Intelligence Beat the Asian Giant Hornet?

Summary

Recently, the Asian giant hornet has been observed in Washington State, which may cause damage to the ecosystem in the future. Therefore, the Washington State Department of Agri- culture (WSDA) has provided large amounts of observations on the species, hoping to get our assistance.

For problem 1, we propose two metrics: **the resource competition coefficient** and **the environmental friendliness** to construct **a time-step difference equation**, and simulate the population dispersal of the Asian giant hornet. We predict the distribution of nests in Washington State within 10 years, and gain the range of activities of an Asian giant hornet by adding noise. The results show that if no measures are taken against the spread of Asian giant hornets, the number of the species will show an approximate **exponential growth** at the initial stage in Washington State. To evaluate the accuracy of the model, we use **Logistic Growth** Model to test the accuracy of the model. The loss is 0.076, and the fastest growing year is the seventh year.These results indicate that we need to control the number of nests early.

In problem 2, we divide it into feature extraction and image classification. For the former, we establish a model based on **auto-encoder** to condense images information, of which the minimum testing loss dips to 0.0274. For the latter, we build three models, the Binary Logistic Regression, the Support Vector Machine, and **the Convolutional Neural Network (CNN)**. The highest accuracy of the three models on the testing set is 0.5303, 0.5758, and 0.8030 respectively, so we choose CNN as our classification model. Finally, we summarize the features of negative images from three aspects: **species characteristics, subject definition and background softness**.

In terms of problem 3, we conduct **Agglomerative Clustering** according to the latitude and longitude of each sighting with unverified or unprocessed status, and divide them into **5 classes**. Then we define the priority of a region based on the positive probabilities of input images in this area. Through analysis, we find that **Seattle is located in the center of the highest probability area**.

With regard to problem 4, we update the model with diffi erent intervals to find optimal update time interval. Selected indicators include the testing loss of auto-encoder and the accuracy of CNN in testing dataset. The optimal time interval for updating models is defined as **the abscissa corresponding to the extremum of the numerical derivative of the time-varying indicators**. Finally, we get the following conclusions: the auto-encoder needs to be updated **every four months**, and the CNN needs to be updated **every three months**.

As for problem 5, we design a number of variables based on observational data to characterize changes in the number of nests. We eliminate the influence of the image recognition model based on **Bayesian inference**. When K **converges to 0**, we think that the pest has been eradicate. Last but not least, we summarize the suggestions and write a memorandum for the WSDA, to assist relevant departments in biological control.

Keywords: Time-Step Diffierence Equation, Auto-Encoder, Convolutional Neural Network, Agglomerative Clustering

Contents

1 Introduction

1.1 Problem Background

The Asian giant hornet is the largest hornet in the world, which is native to East Asia, South Asia, mainland Southeast Asia and some far east parts of the Russian. Recently, it was discovered in the American Northwest at the end of 2019, and there had been lots of sightings in 2020. In fact, the invasion of the Asian giant hornet is not an isolated incident. In 2004, it first appeared in Europe, and then began to spread to Spain, Belgium, Portugal and Italy, rapidly. According to European studies, the propagation speed of Asian giant hornet can reach 49.5 kilometers per year [2, 11].

Even worse, the Asian giant hornet attacks and hunts variety of insects such as bees, bum-blebees, praying mantises and other insects. Moreover, if the Asian giant hornet reach all suitable habitats in North America, the cost for dealing with this disaster in America will ex-ceed $113.7 millions [1]. Therefore, it is very necessary to find and process the Asian giant hornet effiectively.

1.2 Clarifications and Restatements

In this problem, we are given the data of observations on the Asian giant hornet, which is perceived by local citizens, including the time, location, latitude and longitude, and the cor- responding observation photos. Washington State Department of Agriculture divided observa- tions into four status based on photos, namely: positive, negative, unverified, and unprocessed. We will solve the following problems based on the historial observations:

1. Based on positive observations and corresponding latitudes and longitudes, combined with the biological characteristics of the Asian giant hornet, forecast the short-term spread of the species and analyze the accuracy of the prediction;

2. Establish a model for judging whether a photo contains at least an Asian giant hornet, and analyze the image features that make the model output “Negative” results;

3. Based on the above classification model, define the government’s priority in handling citizen observations, for the reason that some unprocessed or unverified observations are most likely to be positive and require the government to explore;

4. Consider that more sightings have been observed, determine the update method and update frequency of the model;

5. According to the model above, analyze what indicators of Washington State has achieved, it can be said that the pest has been “eliminated”;

6. Submit a memorandum to to the Washington State Department of Agriculture to com- prehensively supplement our research results and take corresponding protective measures at the same time.

1.3 Our Work

In problem 1, we construct and simulate a time-step diffierence equation with two metrics: **the resource competition coefficient** and **the environmental friendliness**. We use this model to predict the distribution of nests in Washington State. Then, we gain the range of activities of a single Asian giant hornet by adding noise. After that, we use the Logistic Growth Model to test the accuracy of the model.

We divide problem 2 into two subproblems: **feature extraction**, **and image classification**. In order to reduce the impact of sample imbalance, we apply image flipping and Borderline-SMOTE methods for **data augmentation** first, and then divide the data into the training set and the validation set (testing set). Next, we **utilize auto-encoder to collect key features of images**. To prevent over-fitting, we store **the best performance model** in the testing set for subsequent use. Next, we establish three models, the Binary Logistic Regression, the Support Vector Machine, and the Convolutional Neural Network. After com- paring their accuracy on the testing set, we choose CNN as our image classification model. Finally, we summarize the main features of having negative labels from three aspects: **species characteristics, subject definition and background softness**.

As for problem 3, we take the scope of activities of one-time inspection by the government into account, and conduct Agglomerative Clustering according to the latitude and longitude of each sighting. Then we define the priority of a region based on the positive probabilities of input images.

With regard to problem 4, we discuss the effect of updating the model with different time intervals on the testing results, including the loss of auto-encoder and the accuracy of CNN. **The optimal time interval of updating models is defined as the abscissa corresponding to the extremum of the numerical derivative of the function**, which is the maximum value for auto-encoder, and the opposite for CNN.

In terms of problem 5, the number of nests, the range of spatial distribution and the positive ratio in observations can all reflect the growth of pests. We design a number of variables to characterize changes in the number of nests. Since the image recognition model has a certain error probability, we have made a correction based on Bayesian inference.

2 Reasonable Assumptions

Assumptions about the data provided.

1. The uploaded location (latitude and longitude) has high accuracy and no deviation from the map on Google[1]

2. The judgments made by the WSDA based on photos and the corresponding comments

[1] `https://www.google.com`

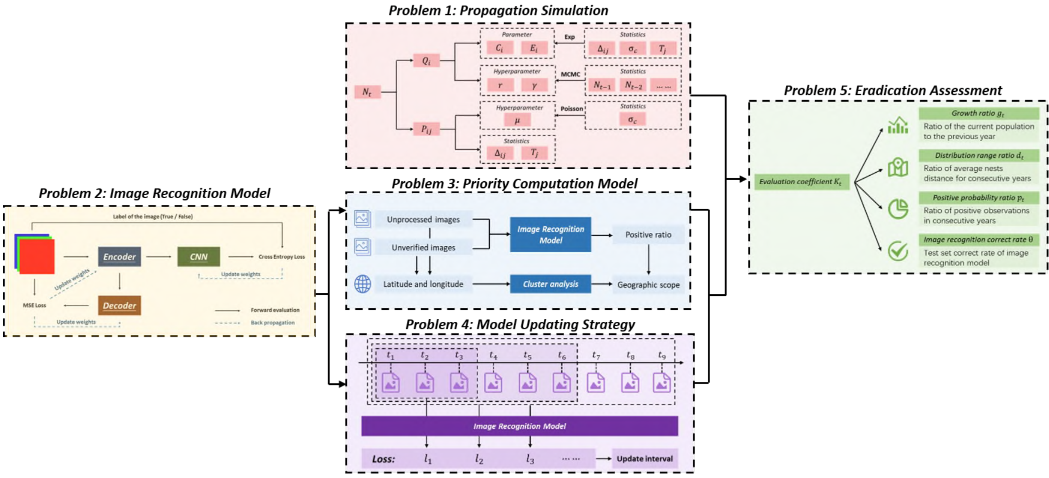


Figure 1: The workflow

is accurate, and there is no error in recognition;

3. The date of the photo was taken is always equal to the corresponding detection date, i.e., there is no delay in uploading photos;

4. Whether a photo contains an Asian giant hornet, is independent of the status of other photos.

Assumptions about the behavior of Asian giant hornet.

1. When the surrounding geographical environment is similar, the Asian giant hornet preys randomly, as well as the population dispersal. At the same time, the species tends to prey and migrate nearby.

2. The Asian giant hornet cannot build nests or prey in the water;

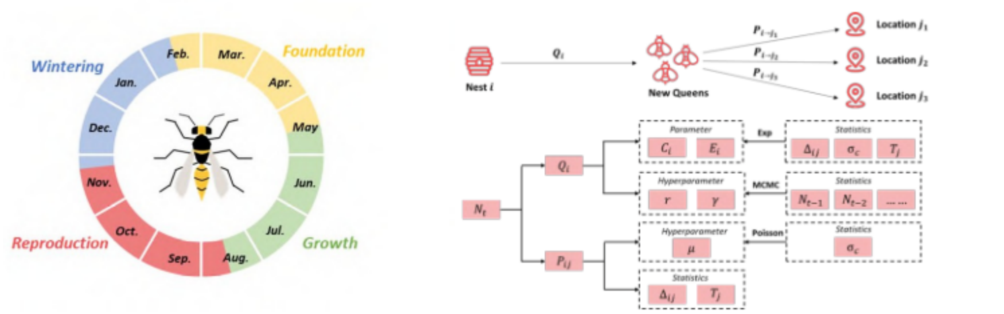
3. The environment is similar in different regions, so fitted hyperparameters in a certain small area can be regarded as constants. Therefore, hyperparameters can be promoted throughout Washington State.

3 Problem 1: Propagation Simulation Based on Yearly Time-Step Diffierence Equation

The parameters used in our analysis in this section are as the table(1)

Table 1: Key parameters for problem 1.

a	b	c
d	e	f



(a) The life cycle of the Asian giant hornet

(b) Parameter relationship in propagation simulation.

Figure 2: Introduction to pest extinction assessment model.

3.1 Method Overview

By clustering and plotting the existing positive data, we get that there are three nests near the State of Washington currently. Therefore, we set the initial number of nests to 3, and their location coordinates are the center of the latitude and longitude observed by citizens. We build a time-step difference equation [8] model in years to predict propagation. Based on the probability distribution of each parameter, we utilize Monte Carlo simulation to solve the model. The workflow is shown in Figure 2(b).

3.2 Model Construction

Like other wasps, Asian giant hornets are an annual species, and the time points when queens produce male peak females are concentrated in September to November. Therefore, the life of the species has a strong periodicity, which belongs to rigid seasonal behaviour. This type of rigid seasonal behaviour is suitable for constructing a time-step difference equation, and iterating in units of years. In the annual Monte Carlo simulation, we conduct propagation on a single nest, and then calculate the number of nests each year.

Parameters Selection. When building the model, we select several parameters with practical significance such as biological populations and geographic conditions, which are introduced below.

First of all, we consider the influence between nests. Since the Asian giant hornet will inevitably produce more nests when spreads in Washington State, the resources suitable for the species in State of Washington will gradually become scarce, and intraspecffic struggle will be inevitable. Therefore, we constructed a parameter to measure the competition for resources of different nests: the intra-species resource competition coeflcient C_{ij}, the formula is shown as follows:

4 The Model Results

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

5 Validating the Model

Morbi luctus, wisi viverra faucibus pretium, nibh est placerat odio, nec commodo wisi enim eget quam. Quisque libero justo, consectetuer a, feugiat vitae, porttitor eu, libero. Suspendisse sed mauris vitae elit sollicitudin malesuada. Maecenas ultricies eros sit amet ante. Ut venenatis velit. Maecenas sed mi eget dui varius euismod. Phasellus aliquet volutpat odio. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Pellentesque sit amet pede ac sem eleifend consectetuer. Nullam elementum, urna vel imperdiet sodales, elit ipsum pharetra ligula, ac pretium ante justo a nulla. Curabitur tristique arcu eu metus. Vestibulum lectus. Proin mauris. Proin eu nunc eu urna hendrerit faucibus. Aliquam auctor, pede consequat laoreet varius, eros tellus scelerisque quam, pellentesque hendrerit ipsum dolor sed augue. Nulla nec lacus.

6 Conclusions

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

7 A Summary

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

8 Evaluate of the Mode

This mode is very good -:)

9 Strengths and weaknesses

Etiam euismod. Fusce facilisis lacinia dui. Suspendisse potenti. In mi erat, cursus id, nonummy sed, ullamcorper eget, sapien. Praesent pretium, magna in eleifend egestas, pede pede pretium lorem, quis consectetuer tortor sapien facilisis magna. Mauris quis magna varius nulla scelerisque imperdiet. Aliquam non quam. Aliquam porttitor quam a lacus. Praesent vel arcu ut tortor cursus volutpat. In vitae pede quis diam bibendum placerat. Fusce elementum convallis neque. Sed dolor orci, scelerisque ac, dapibus nec, ultricies ut, mi. Duis nec dui quis leo sagittis commodo.

9.1 Strengths

- **Applies widely**
  This system can be used for many types of airplanes, and it also solves the interference during the procedure of the boarding airplane,as described above we can get to the optimization boarding time.We also know that all the service is automate.

- **Improve the quality of the airport service**
  Balancing the cost of the cost and the benefit, it will bring in more convenient for airport and passengers.It also saves many human resources for the airline.

References

[1] D. E. KNUTH The TEXbook the American Mathematical Society and Addison-Wesley Publishing Company , 1984-1986.

[2] Lamport, Leslie, LaTeX: " A Document Preparation System ", Addison-Wesley Publishing Company, 1986.

[3] http://www.latexstudio.net/

Appendices

Appendix A First appendix

Aliquam lectus. Vivamus leo. Quisque ornare tellus ullamcorper nulla. Mauris porttitor pharetra tortor. Sed fringilla justo sed mauris. Mauris tellus. Sed non leo. Nullam elementum, magna in cursus sodales, augue est scelerisque sapien, venenatis congue nulla arcu et pede. Ut suscipit enim vel sapien. Donec congue. Maecenas urna mi, suscipit in, placerat ut, vestibulum ut, massa. Fusce ultrices nulla et nisl. Here are simulation programmes we used in our model as follow.
Input matlab source:

```
function [t,seat,aisle]=OI6Sim(n,target,seated)
pab=rand(1,n);
for i=1:n
    if pab(i)<0.4
       aisleTime(i)=0;
    else
        aisleTime(i)=trirnd(3.2,7.1,38.7);
    end
end
```

Appendix B Second appendix

some more text **Input C++ source:**

```
//===========================================================================
// Name         : Sudoku.cpp
// Author       : wzlf11
// Version      : a.0
// Copyright    : Your copyright notice
// Description : Sudoku in C++.
//===========================================================================

#include <iostream>
#include <cstdlib>
#include <ctime>

using namespace std;

int table[9][9];
```

```
int main() {

    for(int i = 0; i < 9; i++){
        table[0][i] = i + 1;
    }

    srand((unsigned int)time(NULL));

    shuffle((int *)&table[0], 9);

    while(!put_line(1))
    {
        shuffle((int *)&table[0], 9);
    }

    for(int x = 0; x < 9; x++){
        for(int y = 0; y < 9; y++){
            cout << table[x][y] << " ";
        }

        cout << endl;
    }

    return 0;
}
```