Based on linear regression, BP neural network, CNN convolution neural network, the Asian giant hornet prevention and treatment suggestions

Xiaochen Xiao, Jiepu Cui and Qishan Cai

University of Electronic Science and Technology of China, Zhongshan, Guangdong, 528402

Keywords: univariate linear regression, BP neural network, convolution neural network, CNN, image recognition

Abstract: Through the preprocessing of the data set, the data set is divided according to the laboratory status, and the positive ID is selected. We use linear regression to make regression analysis on the time series of the selected data set, and further build BP neural network algorithm model to analyze the relationship between the longitude and latitude of the Bumblebee witness report and the detection date in the data set, and finally predict and analyze the model. The results showed that the longitude and latitude of the witness report and the detection date could predict the propagation of bumblebee.

1. Introduction

Vespa mandarinia is the largest wasp species in the world. The emergence of its nest is shocking. In addition, bumblebees are predators of European bees, invading and destroying their nests. A small number of wasps can destroy the entire European bee community in a short time. At the same time, they are considered to be greedy predators of other pests in agriculture.The life cycle of bumblebees is similar to that of many other wasps. The fertilized Queen appears in spring and starts a new colony.In autumn, the new queen leaves her nest and winters in the soil, waiting for the arrival of spring. 50% of the original population will go out to build nests, and the distance of a new queen is estimated to be 30 km.

Due to the potential serious impact of the Asian giant hornet on the local bee population, the existence of the Asian giant hornet will cause great anxiety. Washington state has set up a help line and a website for people to report sightings of these x's. Based on these reports from the public, the state must decide how to prioritize the allocation of limited resources for follow-up investigations. Although some reports have been identified as the Asian giant hornet, many other eyewitnesses have identified other types of insects.

2. Model establishment and solution

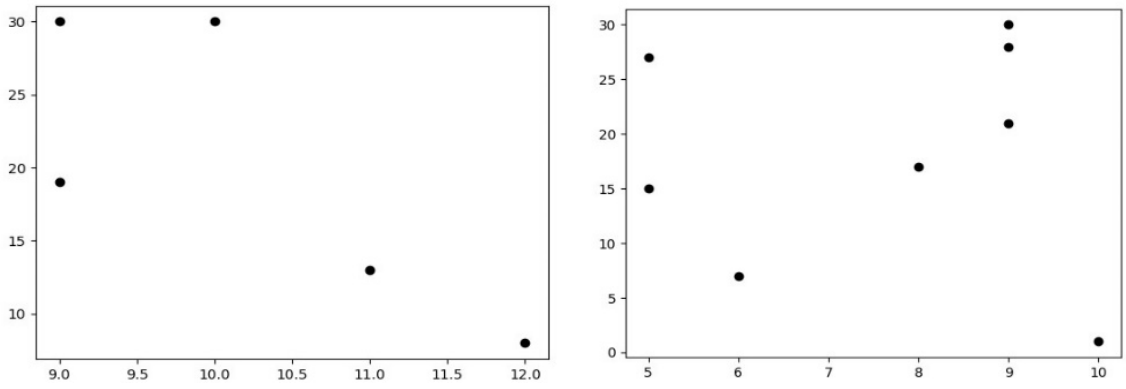

Figure 1. Scatter plot of time for positive ID in 2019 and 2020

The Asian giant hornet 2019, 2020 years of time to make two scatter plot arrangement respectively, we found that point distribution is linear, so we use linear regression algorithm to predict the time (The

X-axis is the month, and the Y-axis is the day of the month), using the BP neural network to predict location reported (Because the distribution pattern of locations we get according to the report is not obvious)

To predict the possible range of the Asian giant Hornet, it is necessary to divide the date and longitude and latitude provided by Negative ID into two groups from the dataset. .The time, longitude and latitude are used to predict the location.

Distribution of Negative ID on the map:

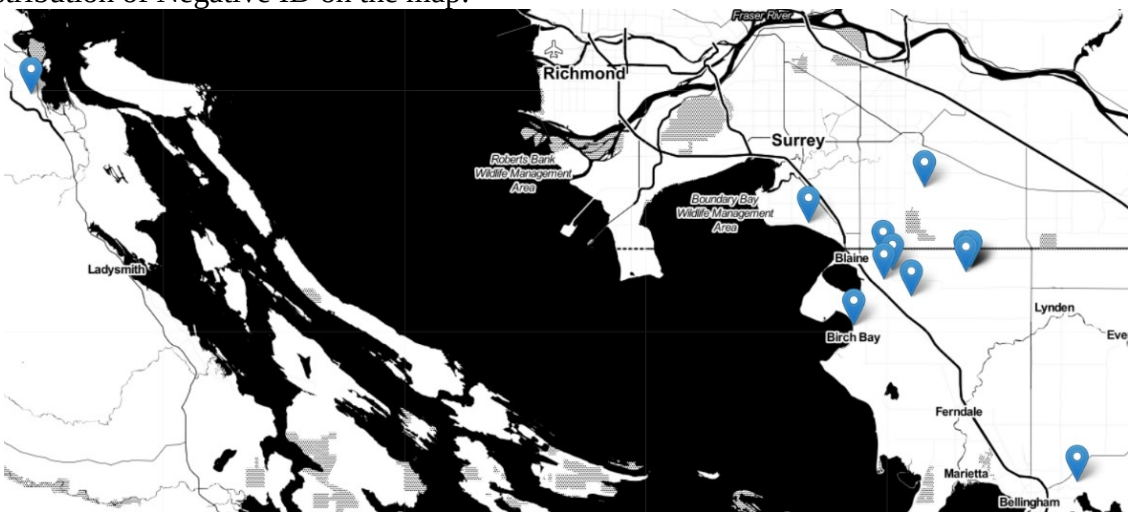

Figure 2. The specific display of positive ID in the map

Through the data obtained by linear regression and BP neural network algorithm, the occurrence time and location of the Asian giant hornet are predicted, as shown in the figure below:

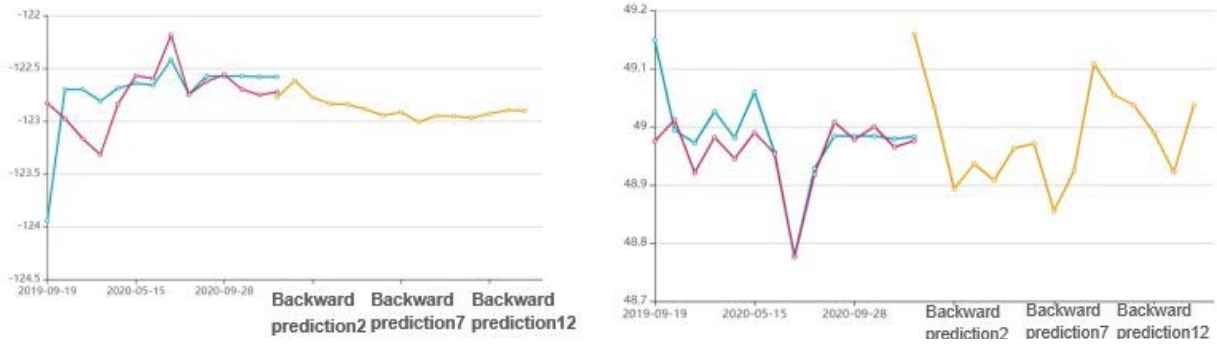

Figure 3. The prediction chart of time and place based on regression

The longitude and latitude of 15 predicted values are obtained and visualized:

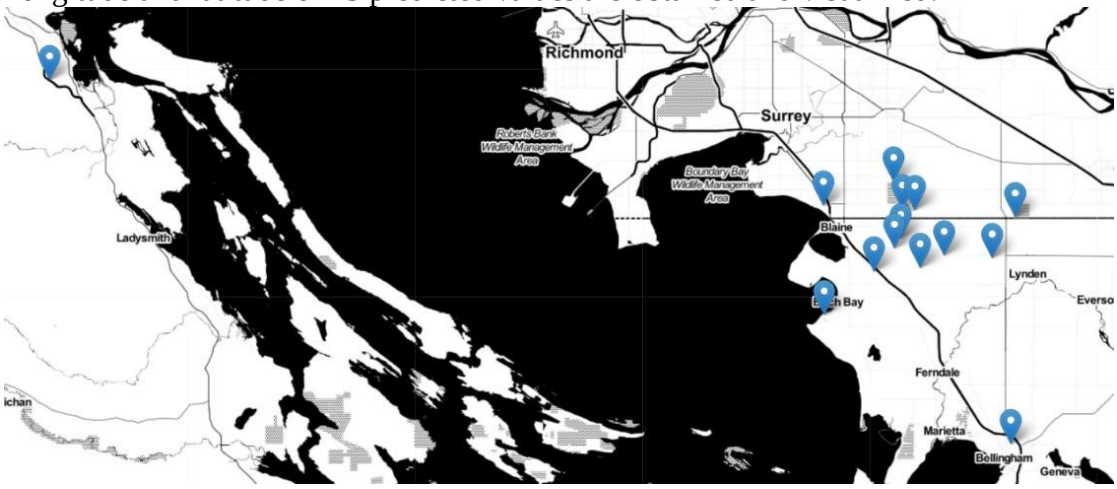

Figure 4. The specific display diagram of the map that predicts the appearance of positive ID in the future

BP model algorithm steps and results.

ata is put into the input layer, and a weight value from 0 to 1 is randomly generated by random library:

$$f = \sum_i W_i X + b$$

Compared with the expected output, the loss function is obtained

$$E = \frac{1}{2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{2}\sum_{i=1}^{n}(y_i - (WX_i + b))^2$$

We need to get the minimum loss function, so we need to iterate to change the weight

$$\Delta W = -\eta E'$$

$$\Delta W = -\eta \frac{\partial E'}{\partial W'}$$

The changed weights are calculated with the training samples and compared with the expected values. If the loss function is not the minimum, the iteration is continued. After several iterations, the minimum loss function is obtained, and the iteration is stopped for data verification. Put the predicted value into the output layer to get the result. 0 means that it does not exist and 1 means that it exists.

From the above analysis, it can be concluded that the spread of this pest in a period of time is predictable

By applying linear regression and BP neural network algorithm to predict the diffusion of the Asian giant hornet in a period of time, and then analyzing the results, the accuracy of the model is about 80%.

```
[49.1409, -123.9647, 1] -> [0.7146574440229292 33]
[49.0146, -122.7683, 1] -> [0.9865744331831903 09]
[49.0388, -122.6600, 1] -> [0.8944571438909715 91]
[49.0109, -122.6462, 1] -> [0.9874874640086126 77]
[48.9809, -122.6490, 1] -> [0.9764573440156971 87]
[48.9710, -122.6586, 1] -> [0.9845728152906220 62]
[48.9478, -122.6902, 1] -> [0.9345722450000301 98]
[48.9514, -122.6181, 1] -> [0.9845686045585519 22]
[48.9032, -122.7671, 1] -> [0.9345719637604211 95]
[48.9641, -122.5808, 1] -> [0.7845742764167351 88]
[49.0096, -122.6261, 1] -> [0.6945678748997754 33]
[48.9613, -122.5076, 1] -> [0.9745714676483234 82]
[49.0030, -122.4711, 1] -> [0.8845718611637204 68]
[48.7723, -122.4787, 1] -> [0.9845710164736318 08]
```

Figure 5. The validity of each forecast location obtained by BP neural network algorithm

We think that the possibility above 0.8 is true, there is possibility between 0.7 and 0.8, and the possibility below 0.7 is false. The validity of the model is 78.57%.

From the above analysis, it can be concluded that the spread of this pest in a period of time is predictable

By applying linear regression and BP neural network algorithm to predict the diffusion of the Asian giant hornet in a period of time, and then analyzing the results, the accuracy of the model is about 80%.

Using a large number of data image data convolution neural network, the image is set to 64*64*3 and put into the input layer for feature extraction.

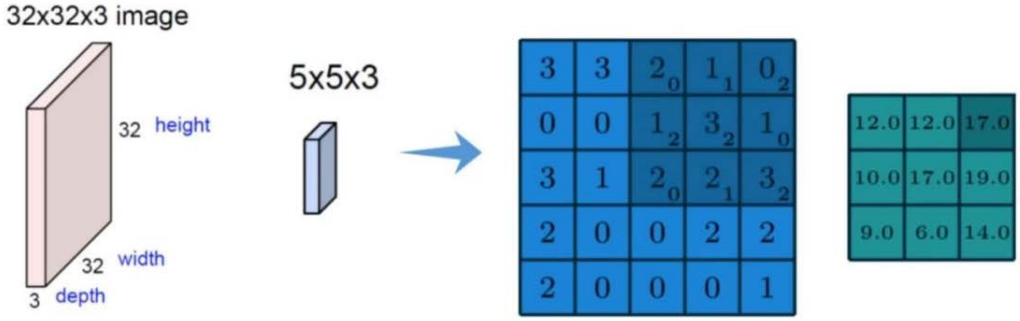

Figure 6. Convolution process diagram of convolution neural network

Random library is used to randomize the weight parameter matrix of 3 * 3. The image is divided into three parts according to RGB. The boundary is filled once with the value of 0. Different weight parameter matrixes are added to the rgb3 layer. The weight parameter matrixes in the same layer are the same. The inner product of each layer is calculated, and the bias term is added. Then add the data to get a feature map. After the activation function, the eigenvalues are compressed by pooling layer.

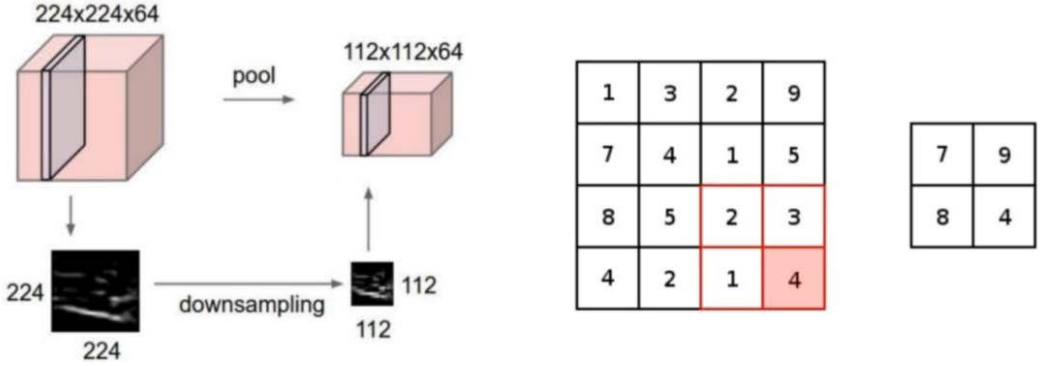

Figure 7. Pooling process of convolution neural network

The feature graph is transformed into a long feature vector through the full link layer. [2]

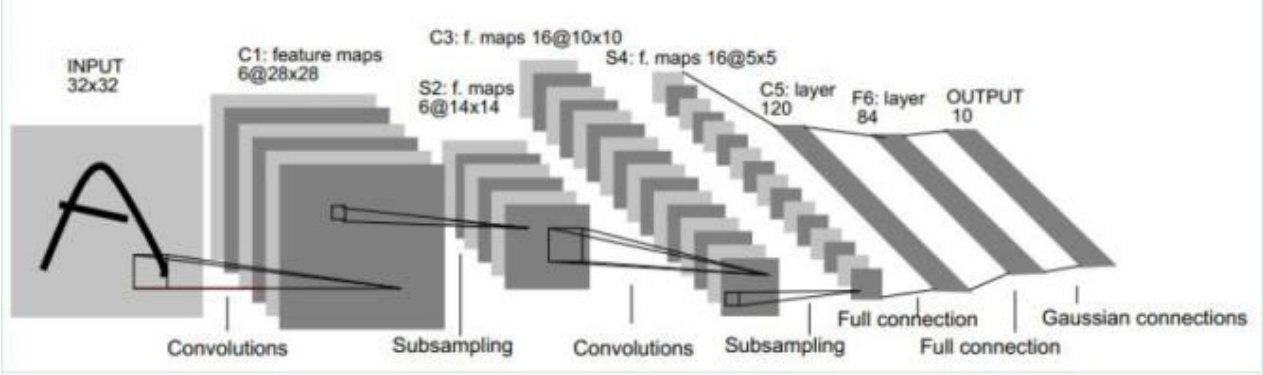

Figure 8. Convolution process diagram of convolution neural network

We use the GPU version of tensorflow 2.3.0 developed by Google. Because of the traditional computer vision to identify an item, we need to outline the location of the item in the picture and save the relevant files. There are more than 3000 pictures in this competition. Considering the competition time, we identified the classic cat and dog binary classification models as Vespa Mandarina A and Vespa Mandarina B respectively. In this way, the binary classification model is changed into a single item recognition model, which takes into account both the time and the accuracy of model recognition.

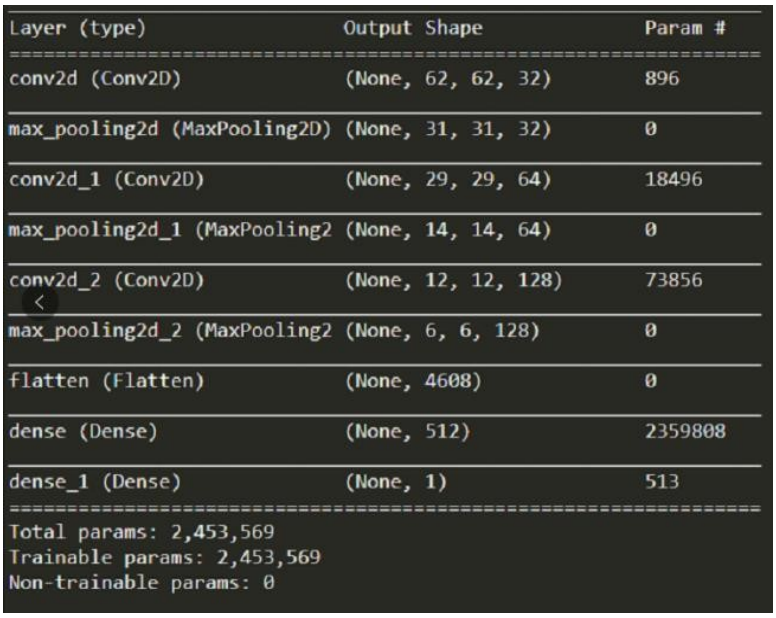

Figure 9. The structure diagram of the CNN model we used

3. Conclusion

In other words, autumn is the most likely time to witness the Asian giant hornet [4]. As long as the witness report of the Asian giant hornet within 30 km of Washington state and its vicinity determined in the three months of September, October and November is 0, we believe that this pest has been eliminated in Washington state

It is feasible to qualitatively analyze the occurrence of pests by linear regression analysis. In addition, we can find out the main factors that affect the occurrence of crop pests, and analyze the propagation trend of Bumblebee to a certain extent;

However, the effect of quantitative prediction on the transmission of bumblebee is not ideal, and there are some limitations as a quantitative prediction method of Bumblebee transmission.

References

[1] Fan Gaofeng, Wang Weili, Liu Chun, Dai Huizhu, wind power prediction based on artificial neural network [J]. Chinese Journal of electrical engineering, 2008.12 (34): 118-123

[2] Zhao Xinqiu, he Hailong, Yang Dongdong, Duan Siyu. Application of improved convolution neural network in image classification [J]. High tech communication, 2018, (11-12): 930-936

[3] Zhou Jianliang, some understandings of fenfeng [J], Chinese bee industry, 2014 (7): 15-16

[4] Hong Wei, Xu Baohua, Liu Shengping, design and Experimental Research on long-term monitoring system of bee colony multi characteristics [J]. Smart agriculture 2020.6 (65): 105-114