Journal of Electronics and Information Science (2021) 6: 27-31
Clausius Scientific Press, Canada

DOI: 10.23977/jeis.2021.61004
ISSN 2371-9524

Analysis of Hornet Forecast Model based on Fuzzy Theory

Qian Chen

School of Statistics and Mathematics, Guangdong University of Finance and Economics, Guangzhou, Guangdong 510320

Keywords: SMOTE, GM Model, hornets, Positive ID

Abstract: The 14 Positive ID the paperre arranged in time order and the GM Model was used to predict the range of propagation, and the results of the prediction the paper re obtained as follows: from 48.92 to 49.05 in length and from -122.47 to -122.55 in latitude in 2021. There is a distance of 30 km betthe paperen the predicted results and the initial point where the presence of hornets was confirmed. The average relative error is less than 0.01, so the model prediction accuracy is good. Since the life cycle of hornets is very related to seasons, the time is converted into seasons and then One-Hot-Encoding of seasons; the TFIDF Algorithm is used to calculate the importance of each Note to replace the original Notes. The SMOTE Method used in this paper to fill the Positive ID minority class sample leads to the proliferation of Vespa mandarinia seriously endangering the local ecology, so the SMOTE Method used in this paper to fill the Positive ID minority class sample. The models all seek to maximize the recall of a few classes of Positive ID. After model testing our models are all excellent in identifying pests accurately, as evidenced by the ROC (with Positive ID as a positive example) curve and AUC =0.99.

1. Introduction

There have been many reported pest sightings in neighboring Washington State and Vancouver Island since the discovery of a colony of Vespa mandarinia (also known as the Asian giant hornet) on Vancouver Island, Canada, in September 2019, as the paperll as a large number of mistaken sightings [1]. Vespa mandarinia is a predator of European honeybees and one of the largest species of hornet in the world. People are concerned about the significant effect on local bees that the arrival of this bee would have. The state of Washington, USA, collects reports of sightings of these hornets. Of these reports, there have been only very few identified as such. To properly observe and interpret Vespa mandarinia [2], the paper will analyze the public records and provide recommendations to the Washington State Department of Agriculture.

2. Gray prediction models to predict

Gray Time Series forecasting was chosen, i.e. the time series observed reflecting the characteristics of the predicted object was used to construct a gray forecasting model to predict the number of characteristics at a future time or time to achieve a certain number of characteristics. The paper have

drawn a graph (Figure 1) to show the trend [3]:

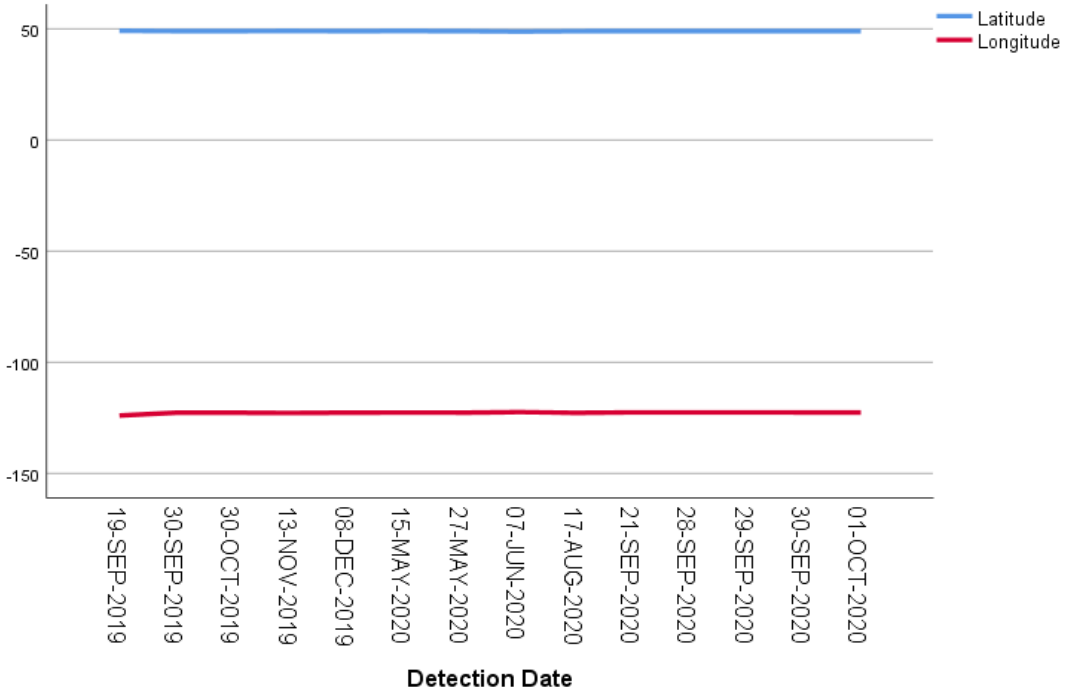

Figure 1: Latitude and longitude Change over time

Predict the next points where the correct sighting may occur.

(i) The time was pre-processed to set the date of the first bumblebee sighting as the first day and then in increasing order for each day thereafter. Set the time as X. Use the time as X [4].

(ii) The time series model was constructed using latitude as y1and longitude as y2.

(iii) Assuming serial data of time and latitude as (The same for longitude):

$$y_1^{(0)} = \left\{ y_1^{(0)}(1), y_1^{(0)}(2) \cdots y_1^{(0)}(n) \right\} \tag{1}$$

(iv) Generated by first-order accumulation[5]:

$$Y_1^{(k)} = \sum_{m=1}^{k} Y_0(m), (k = 1, 2, \cdots, n) \tag{2}$$

$$Y_1^{(1)} = \left\{ Y_1^{(1)}(1), Y_1^{(1)}(2) \cdots Y_1^{(1)}(n) \right\} \tag{3}$$

(v) The corresponding whitening differential equation is obtained as:

$$\frac{\mathrm{d}Y_1^1}{\mathrm{d}t} + aY^1 = u \tag{4}$$

(vi) Introducing the notation:

$$\hat{a} = \begin{bmatrix} a \\ u \end{bmatrix}, x_1 = \begin{bmatrix} Y_0(2) \\ \vdots \\ Y_0(n) \end{bmatrix}, B = \begin{bmatrix} -\frac{1}{2}\left(Y_1^1(2) + Y_1^1(1)\right) & 1 \\ \vdots & \\ -\frac{1}{2}\left(Y_1^1(n) + Y_1^1(n-11)\right) & 1 \end{bmatrix} \tag{5}$$

(vii) Using the least squares method to solve for:

$$\hat{a} = (a,u)^T = \left[B^T \cdot B\right]^{-1} \cdot B^T \cdot x_1 \tag{6}$$

(viii) Use the form of least squares to solve for:

$$\tilde{Y}^1(k) = \left(Y_0(1) - \frac{u}{a}\right)e^{-a(k-1)} + \frac{u}{a} \tag{7}$$

(ix) Reduction of predicted cumulative values to predicted values:

$$x^{(0)}(k+1) = x^{(1)}(k+1) - x^{(1)}(k) \tag{8}$$

3. Level of grey model precision

According to the model, the measurement of $\hat{X}^{(1)}(k)$ and cumulative subtraction to produce $\hat{X}^{(0)}(k)$ is carried out and the absolute and relative error series of the initial $\hat{X}^{(0)}(k)$ and $\hat{X}^{(0)}(k)$ series are determined as follows:

$$\Phi(i) = \frac{\Delta^{(0)}(i)}{X^{(0)}(k)} \times 100\backslash\%, i = (1,2,\cdots,n) \tag{9}$$

y1: The average relative residuals is 0.0008085 and the average eta is 0.0013443.
y2: The average relative residuals is 0.00039723 and the average eta is 0.0014598.

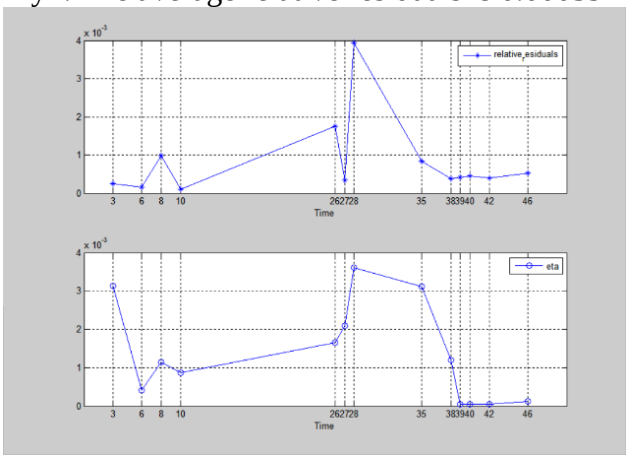

Figure 2: y1model testing

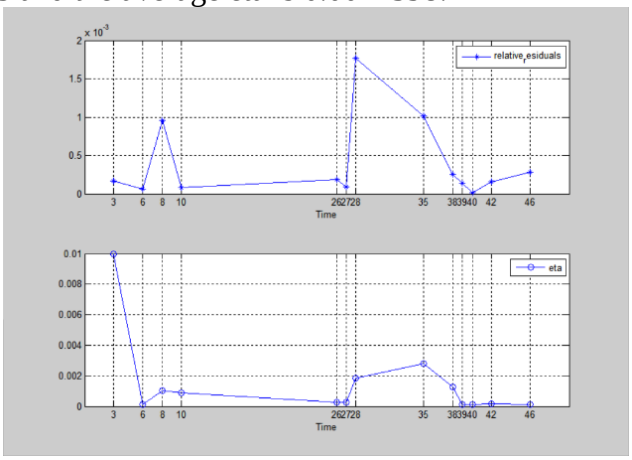

Figure 3: y2model testing

4. A possible range of activities to validate

The paper draw possible swarm scattering routes, as judged by time.

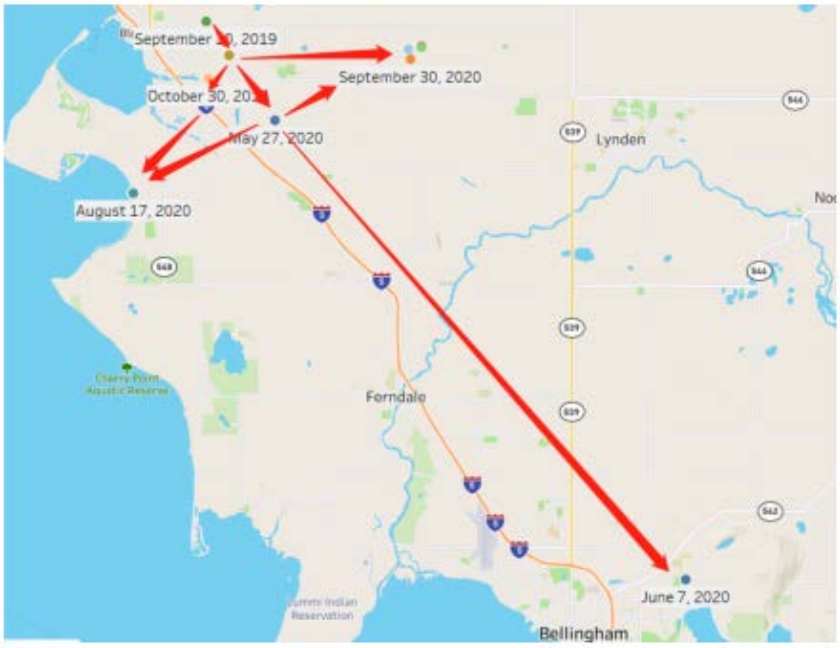

Figure 4: Possible dispersal route of V espa mandarinia

Second, the paper consulted the formula for converting latitude and longitude to distance, which is known from reference [4].

$$1^{\circ} \approx 111\text{km} \tag{10}$$

The more ellipses overlap, the higher the probability of seeing a V espa mandarinia swarm in the area.

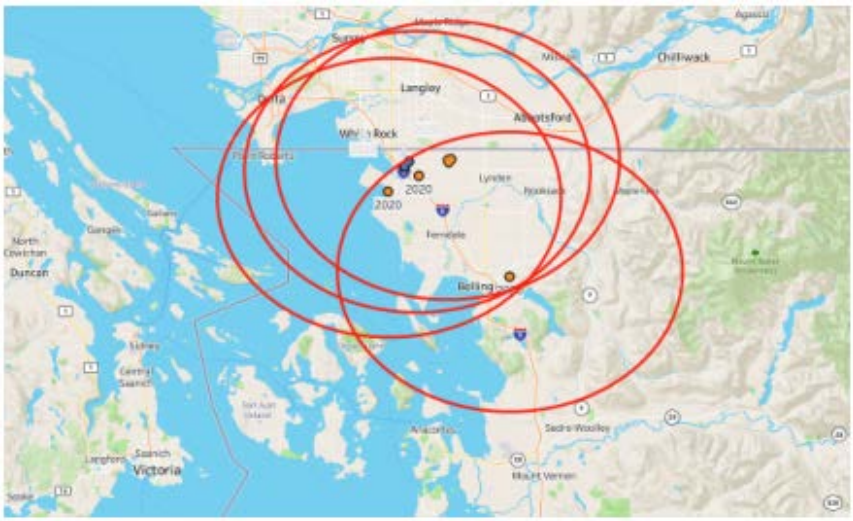

Figure 5: Mapping the range of possible swarm sightings in 2021

As the paper can see, the current trend, regardless of policy and other influences, is that the Vespa mandarinia range in northern Washington will expand even further, with nearby Lynden, Nooksack, Maple Falls, and even the small island of Drcas Island all likely to become new Vespa mandarinia colonies. Although there is no evidence of natural dispersal across the ocean yet, it is possible that

Hoopoe could reach Drcas Island by logistic means of transport.

To test the scientific effectiveness of this method, The paper use Grey Prediction Model to predict the position of the next foothold, judging if it is in the circle,. The result of 2021 proves that the grey prediction model has a certain reliability. So the paper explored the scope of possible eyewitness reports in 2022. f not prevented, the potential living space of Hu Peak will be further extended in two years to include Mount Baker Wilderness, Sedro-Wolley, Anacortes, etc. Five years from now, the surrounding Olympic National Park, E. C. Manning Provincial Park, Passyten Wilderness, North Cascades National Park, Glacier Peak Wilerness the surrounding Olympic National Park.

5. Conclusions

The training of Machine Learning model is usually carried out by learning the mapping betthe paperen a certain set of input features and output targets. Ideally, the paper would expect our models to predict the future exactly as the paper would using the data used in the training process when making predictions in a production environment. The paper uses the SMOTE (synthetic minority oversampling technique) method to balance the class distribution by adding synthetic minority class samples to the data, which reduces the possibility of overfitting and improves the generalization performance of the classifier on the test set, resulting in more accurate results.

Combining the latitude and longitude results predicted by the gray model with the habits of Vespa mandarinia, which corroborate each other. 2. For the data set, multiple methods are used to extract valid data. And two methods are used to assist the classification. For the Logistic Regression, the parameters are adjusted to prove the robustness of the model.

References

[1] https://agr.wa.gov/departments/insects-pests-and-the papereds/ insects/hornets /data. Accessed 7 / 2/2021. Washington State Department of Agriculture.

[2] 2021MCM_ProblemC_V espamandarinia.pdf (from Pennsylvania State University Extension).

[3] Lu Yi. The Research and Application of Grey Forecast Model [D]. Zhejiang Sci-Tech University. 2014

[4] https://en.wikipedia.org/wiki/Geographical_distance. Accessed 7/2/2021.

[5] Bruno Trstenjak, Sasa Mikac, Dzenana Donko. KNN with TF-IDF based Framework for Text Categorization [J]. Procedia Engineering, 2014, 69.