

Recognition of Bumblebee Behavior Based on AHP Model

Yingchao Hai*

Queen Mary University of London Engineering School, Northwestern Polytechnical University,
Xi'an, Shaanxi, 710129

*Corresponding author: yc.hai@mail.nwpu.edu.cn

Abstract. We propose a vespa mandarinia recognition model based on analytic hierarchy process to assess and classify reports of sightings. we construct the analytic hierarchy process (AHP) model to evaluate and classify reports. The four main indicators in the AHP model extracted from the data are image, position, note and detection time, and the corresponding weights of the four indicators are calculated. Then, we establish models based on the four indicators to determine the scoring criteria. For the notes part, we use Term Frequency-Inverse Document Frequency arithmetic to extract keywords and Long Short-Term Memory (LSTM) model to classify and score the notes. Based on the AHP model, scores of the report can be calculated. Next, we use the scores of the given reports to calculate the classification threshold and divide all the report into three status based on two thresholds: positive, negative and unverified. According to the corresponding formula, the two thresholds are calculated as 0.6276 and 0.3937 respectively.

Keywords: Probability selection model, Analytic Hierarchy Process, LSTM model, AHP

1. Introduction

Biological invasion has become one of the most serious problems faced by human beings, which has no doubt received great attention from the international community and scientists. Biological invasion refers to the process in which organisms from the original place of origin invade to another new environment through natural or man-made means, thus causing harm or even disaster to the biological diversity, agriculture, forestry, animal husbandry, fishery production and human health of the invaded place[1].

Image recognition technology based on deep learning, which can effectively improve the accuracy of image recognition and reduce generalization error, has become a research hotspot of crop pest identification technology. Image processing technology and convolutional neural network algorithm based on region detection are used to classify insects[2,3].

In the process of species identification, deep learning is often used to solve the problem. In this paper, AHP is used to evaluate and score the report of sightings, so as to classify it and improve the efficiency of the agricultural department in solving this problem.

2. Model assumptions and notation

2.1. Assumptions

- It is assumed that there must be beehives near the observation site. That is, there is one beehive for each observation site.
- Assume that colonies at all different locations are equally productive.
- Regardless of the fact that the vespa mandarinia travels in vehicles and over long distances.
- The influence of human factors is not considered in the construction of bee colony reproduction model.
- The effect of climate and temperature is not considered.

2.2. Notations

The key mathematical notations used in this paper are listed in Table 1.

Table 1. Notations used in this paper

Symbol	Description
$X^{(0)}$	Original data sequence
$X^{(1)}$	Accumulative sequence
Z	Adjacent value sequence
B	Data matrix
Y	Data vector
$\&(k)$	Residual
(X_0, Y_0)	The latitude and longitude of the center of a given location
(X_i, Y_i)	The latitude and longitude of the sightings
S	Evaluation score
O	The probability of wasps nesting
N_{before}	The number of the all the Asian hornet in the detection date
N_{all}	The number of the all the Asian hornet in the detection date
$\bar{S}_{pos}$	Average score of the positive report
$\bar{S}_{neg}$	Average score of the negative report
$\bar{S}_{unc}$	Average score of the uncertain report

3. Vespa mandarinia recognition model construction and solving

Due to hornets often build underground nests, it is not easy for non-professionals to find and clean them. In addition, hornets attack people and bring great threats to the ecological environment[4]. Therefore, this needs to be cleaned up by professionals. National agriculture department set up a hotline and a web site, people can through the website upload about the hornet's sightings, because the hornets and other insects are similar, does not have the ability to distinguish the people tend to upload the wrong sightings, the agriculture department need to screening of sightings, identify the most likely sightings, to maximize the use of resources.

In order to identify the wasps, the information which provided by witnesses should be used properly. First of all, we choose to carry on the preliminary screening to the data provided in the topic so as to realize the data preprocessing.

After the choice, the image, longitude and latitude, notes and detection date are selected as the four main index. The longitude and latitude determine the position of the hornet together. Based on this, we put forward the analytic hierarchy process to further study this problem.

3.1. Establish a scoring system to determine report priority

3.1.1 Position

According to the data in the question, we can see where positive reports are clustered together. Therefore, when the distance from the initial area is closer, the authenticity of the report can be explained to a certain extent. We can score the report based on the distance from the initial position provided by the report.

The longitude and latitude of the center of the known location (X_0, Y_0) can be get from the formula (1).

$$X_0 = \frac{\sum_{u=1}^n X_u}{n}, Y_0 = \frac{\sum_{u=1}^n Y_u}{n} \quad (1)$$

Due to the location data provided in the report is latitude and longitude, so we need to convert the latitude and longitude of two points into the distance between two points. The distance transformation formula is shown as formula (2).

$$d = R \arccos[\cos(Y_i) \cos(Y_0) \cos(X_i - X_0) + \sin(Y_i) \sin(Y_0)] \quad (2)$$

In the formula (2), R is the radius of the earth, (X_0, Y_0) is the longitude and latitude of the center of the known location, (X_i, Y_i) is longitude and latitude of the location of the reported sightings.

We can know that a new queen has a range estimated at 30km for establishing her nest, the scoring system can be constructed from this. When the distance is less than 30km, the score can be given as 1. When the distance is more than 30km and less than 60km, the score can be given as 0.8. When the distance is more than 60km and less than 90km, the score can be given as 0.5 and so on....

Since the accuracy would be reduced as a result of such scoring, we use the cubic equation to fit the curve, we can show the cubic equation in the formula (3).

$$S = -9.312 \times 10^{-8} d^3 + 6.459 \times 10^{-5} d^2 - 0.0154d + 1.4117 \quad (3)$$

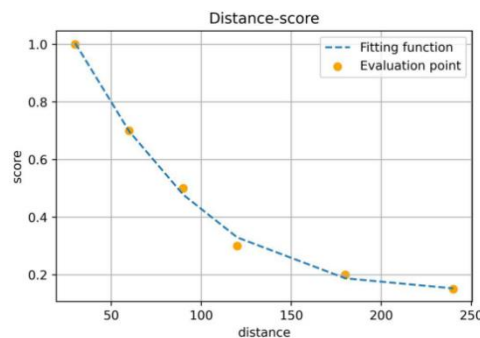


Figure 1. The relationship between score and distance

The relationship between score and distance is shown in Figure 1. Interpretation of result $y = \beta_3 x^3 + \beta_2 x^2 + \beta_1 x + \beta_0$.

Table 2. Parameter analysis

Parameter	Estimate of parameter	Confidence interval
β_0	1.4117	[1.1039, 1.7196]
β_1	-0.0154	[-0.0250, -0.0059]
β_2	6.459×10^{-5}	$[-1.7967 \times 10^{-5}, 1.4715 \times 10^{-4}]$
β_3	-9.312×10^{-8}	$[-2.9675 \times 10^{-7}, 1.1250 \times 10^{-7}]$
$R^2=0.9971$	$F=232.1991$	$P=0.0043$

According to the data in the Table 2, we can see that the equation as a whole is significant. The 99.71 percent of the value can be determined by this model, P is less than 0.05 and F is much larger than F tests the threshold. Therefore, we can say this equation is excellent.

3.1.2 Time

Using the grey prediction model established above, we can predict the number of Asian bumblebees at any time. The scoring criteria for time is proposed, the criteria are shown in the formula (4).

$$S = \frac{N_{before}}{N_{all}} \quad (4)$$

S represents the evaluation score, N_{before} represents the number of the all the Asian hornet in the detection date, N_{all} represents the number of the all the Asian hornet in the report date.

3.1.3 Image

Some eyewitness reports include images, image plays an important role in biometric identification. However, through data crunching, we found that less than 10 percent of the reports provided images, the number of images is small and not uniform, which makes image recognition become a difficult problem. The images in the report some are correct, but some are wrong. Therefore, we hope to reduce the weight brought by images in the subsequent model building.

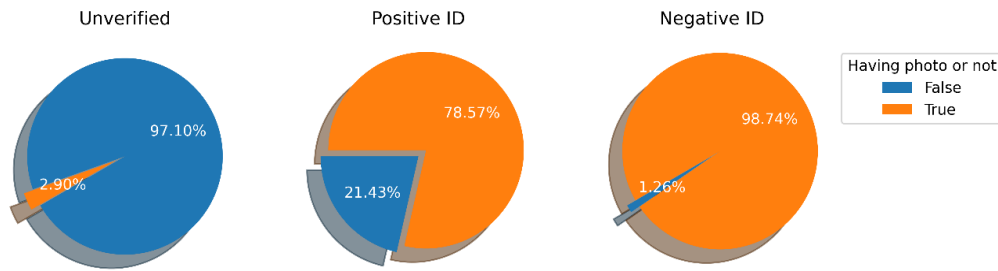


Figure 2. Proportion with or without attached pictures

Proportion with or without attached pictures is shown in Figure 2. From what has been discussed above, if the report adds image, it will be given 0.7 marks. And if the report does not add image, it will be given 0.3 marks.

3.2. Analytic hierarchy process

3.2.1 Build an analytic hierarchy model

In this problem, the AHP method should be used to evaluate all the eyewitness reports and set priorities[5-7].

The factors contained in the problem are stratified. The factors are divided into three levels, which are the highest level (to screen out the sightings with high credibility), the middle level (to realize the indicators of screening), and the bottom level (all sightings).

3.2.2 Construct a pairwise comparison matrix

Based on four indicators selected from eyewitness reports: time, image, position and notes. Construct the judgment matrix of the first-grade evaluation index.

$$A = \begin{bmatrix} \frac{w_1}{w_1} & \frac{w_1}{w_2} & \frac{w_1}{w_3} & \frac{w_1}{w_4} \\ \frac{w_2}{w_1} & \frac{w_2}{w_2} & \frac{w_2}{w_3} & \frac{w_2}{w_4} \\ \frac{w_3}{w_1} & \frac{w_3}{w_2} & \frac{w_3}{w_3} & \frac{w_3}{w_4} \\ \frac{w_4}{w_1} & \frac{w_4}{w_2} & \frac{w_4}{w_3} & \frac{w_4}{w_4} \end{bmatrix} \quad (5)$$

Calculate the maximum eigenvalue.

$$\lambda_{\max} = \frac{1}{n} \sum_{i=1}^n \frac{(AW)_i}{w_i} \quad (6)$$

Calculate the eigenvectors.

$$W_i = \frac{\bar{W}_i}{\sum_{j=1}^n \bar{W}_j} \quad (7)$$

Where $W = [W_1, W_2, W_3, W_4]$.

3.2.3 AHP Consistency Test

After the judgment matrix is constructed, the relative weight of each element in two levels is calculated by the judgment matrix, and the consistency test is carried out[8]. It is not allowed for the judgment to deviate too much from the consistency, so the consistency test of the judgment matrix is needed. The specific test steps are as follows:

Step 1: Calculated consistency index CI .

$$CI = \frac{\lambda_{\max} - n}{n - 1} \quad (8)$$

Step 2: Check the criteria for testing the consistency of the judgement matrix from the relevant data $RI(n)$. Relationship between RI and n is shown in Table 3.

Table 3. Relationship between RI and n

n	1	2	3	4	5	6	7	8	9
RI	0	0	0.52	0.89	1.12	1.26	1.36	1.41	1.46

Step 3: Calculate the random consistency ratio of the judgment matrix CR .

$$CR = \frac{CI}{RI} \quad (9)$$

If the consistency ratio is less than 0.1, the judgement matrix of AHP has satisfactory consistency, namely its consistency degree is acceptable.

3.2.4 Result of AHP

The judgment matrix of each index is given by the expert scoring method to reflect its relative critical degree[9]. According to the result of data processing, the judgment matrix of these indexes has passed the consistency test. Index judgement matrix is shown in Table 4.

Table 4. Index judgement matrix

	Position	Time	Note	Image
Position	1	2	3	3
Time	1/2	1	2	2
Note	1/3	1/2	1	1/3
Image	1/3	1/2	3	1

According to the index judgment matrix calculation, we can get the corresponding weight of each index, the index shown in the Table 5.

Table 5. Evaluation index and weight

	Position	Time	Note	Image
Weight	0.4448	0.2581	0.1076	0.1896

After consistency check calculation, the coincidence indicator $CI = 0.0550$; and the consistency ratio $CR = 0.0618$. $CR < 0.10$, so the consistency of the judgment matrix A is acceptable.

The scoring model established above can be used to calculate the scores of each indicator of each eyewitness report, and then the weight calculated by the analytic hierarchy process can be used to solve the corresponding scores for each eyewitness report. After that, we will use the calculated scores to classify each eyewitness report.

$$S = \sum_{i=1}^4 S_i w_i \quad (10)$$

Them into three kind according to the score.

$$\bar{S}_{pos} = \frac{\sum_{i=1}^a S_{pos}}{a} \quad (11)$$

$$\bar{S}_{neg} = \frac{\sum_{i=1}^b S_{neg}}{b} \quad (12)$$

$$\bar{S}_{unc} = \frac{\sum_{i=1}^c S_{neg}}{c} \quad (13)$$

In the formula (11) (12) (13), $\bar{S}_{pos}$, $\bar{S}_{neg}$, $\bar{S}_{unc}$ represents the average score of the three types. The result of the average score of the three types is shown in the Figure 3.

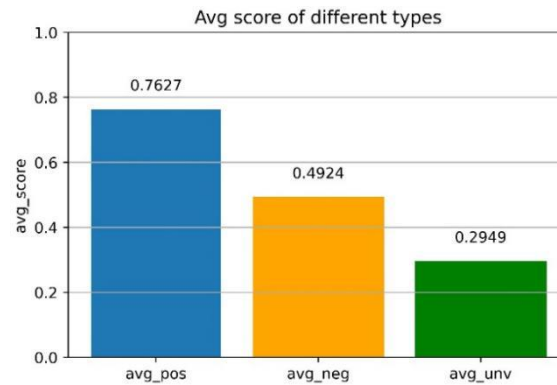


Figure 3. Result of the average score of the three types.

Then, we will introduce the threshold value as the boundary value.

$$A_{pos} = \frac{\bar{S}_{pos} + \bar{S}_{neg}}{2} \quad (14)$$

$$A_{neg} = \frac{\bar{S}_{neg} + \bar{S}_{unc}}{2} \quad (15)$$

After calculation, we can get the value of the A_{pos} , A_{neg} , the former equals 0.6276 and the latter equals 0.3937. This is to say, if the score is larger than 0.6276, we know this report is positive example, if the score is larger than 0.3937 and less than 0.6276, this report is negative example, and if the score is less than 0.3937, this report is uncertain example.

4. Conclusion

Biological invasion has become one of the most serious problems faced by human beings, which has no doubt received great attention from the international community and scientists. Biological

invasion refers to the process in which organisms from the original place of origin invade to another new environment through natural or man-made means, thus causing harm or even disaster to the biological diversity, agriculture, forestry, animal husbandry, fishery production and human health of the invaded place. Biological invasion will be an inevitable problem in the future, so we need to make preparations in advance to deal with it. We can control invasive organisms as soon as possible by establishing a good website in advance and completing the screening algorithm[10].

In our model, we use a grey prediction model to make a simple prediction of the population size and give a prediction of hornet location diffusion based on the literature. Then, four main factors are selected from eyewitness reports, which are detection time, location, note and image. Then we define the scoring standard of four main factors, and take the four main factors as the evaluation index to construct the analytic hierarchy structure and use the analytic hierarchy process to analyze the report. The report is classified by using known data to build a threshold for classification.

This model can be used to prevent other biological invasions, we only need to change the relevant parameters and weights, and substitute the living habits and data of other species into the model to realize the prediction.

References

- [1] Qinglong Wang, Bailian Li. Advances in biological invasion models [J]. Tech review, 2011, 29(10): 71-79.
- [2] Honsong Yan. Research on Deep Learning Based Natural Scene Text Recognition and Generation Algorithm [D]. Henan University, 2020.
- [3] Lemei Jiang. Insect recognition method based on deep learning and its application [D]. Agricultural University of Hunan, 2019.
- [4] A Susano, S Sutrisno, A Darmawan, et al. Determination system of designing the best beach object in bail by using AHP method. 2019, 1175(1)
- [5] Sang-Woon Kim, Joon-Min Gil. Research paper classification systems based on TF-IDF and LDA schemes. 2019, 9(1): 1-21.
- [6] Hongwei Pang. Research on insect image recognition technology based on deep learning [D]. University of Chinese Academy of Sciences (School of Artificial Intelligence, Chinese Academy of Sciences), 2020.
- [7] Ming Tan, Shouyum Shen. Plant selection in Zhurong Peak scenic spot of Hengshan Mountain based on analytic hierarchy process [J]. Journal of Hunan City University (Natural Science Edition), 2021, 30(01): 45-50.
- [8] Pengrong Ji, Weisong Huang, Xiangyong Hu. Study on the characteristics of grey prediction model [J]. Systems Engineering Theory and Practice, 2001(09): 105-108+139.
- [9] Gonggui Cai. The living habits of the wasp [J]. China's bee industry, 2019, 70(10): 35.
- [10] Pramita Sree Muhuri, Prosenjit Chatterjee, et al. Using a Long Short-Term Memory Recurrent Neural Network (LSTM-RNN) to Classify Network Attacks. 2020, 11(5)