

ARTICLE

Wolf is Coming—Dynamic Classification Prediction Model of Vespa Mandarinina

Yang Yue^{1,2*} Haomiao Niu^{1,2} Jiao Liu³

1. Mathematical Modeling Innovation Lab, North China University of Science and Technology, Tangshan, Hebei, 063210, China

2. School of Chemical Engineering, North China University of Science and Technology, Tangshan, Hebei, 063210, China

3. School of Clinical Medicine, North China University of Science and Technology, Tangshan, Hebei, 063210, China

ARTICLE INFO

Article history

Received: 1 March 2021

Accepted: 22 March 2021

Published Online: 21 May 2021

Keywords:

Neural network

Ecological equilibrium

Dynamic reproduction

ABSTRACT

Given the threat of Vespa mandarinina invasion to ecological balance, according to the data and information provided, the dynamic reproduction model of Vespa mandarinina is established by using natural domain interpolation, and the variation law of total bumblebee with time, latitude, and longitude is obtained. At the same time, we established the classification prediction model by using a neural network and established the mapping relationship between time and space to evaluation grade.

We meshed the area provided by the title, assigned values to the location of Vespa mandarinina (VM), and established a VM diffusion model with natural neighborhood interpolation. Its propagation process is simulated by cellular automata. It is determined that VM spreads in a circular shape centered at (122.93174°W, 48.93457°N) and (122.57376°W, 49.07848°N) in the Washington area, with the farthest distance being 1184.4 km and 985 km respectively.

We set up a classification prediction model for better classification. According to the image upload time and location, SVM and neural network are used for classification prediction, and the classification accuracy is 74.26% and 97.60%, respectively, and the neural network has higher classification accuracy. So we choose the neural network.

1. Introduction

The invasion of alien organisms has had a huge impact on local agriculture, food production, and species diversity. Recently, the VM was discovered in Vancouver and the neighboring Washington State area. The discovery of this invader has sounded a wake-up call for the locals. The frightening thing is that this insect is the largest known wasp in the world ^[1]. The accidental introduction of the

VM has brought serious consequences to European honeybees. While attacking the hive and preying on European honeybees, it also affected its foraging activities and living space. At the same time, they are voracious predators of other insects that are considered agricultural pests.

In addition to the threat to the beekeeping industry, the invasion of Vespa mandarinina in North America is also concerning for public health ^[2]. Their poisonous needles

*Corresponding Author:

Yang Yue,

Mathematical Modeling Innovation Lab & School of Chemical Engineering, North China University of Science and Technology, Tangshan, Hebei, 063210, China;

Email: 2090742207@qq.com

sting people and can cause severe allergic reactions and even death. Therefore, this caused greater anxiety among the locals, and Washington State has set up a help phone and web-site for this purpose to count the wasps. However, there are several other VM that may be confused for them. Because bumblebees spread quickly and become pests, it is necessary to reliably estimate the potential range of bumblebees in North America to assess the possible impact on humans and guide future efforts to eliminate bumblebees [3].

2. The Establishment of Dynamic Propagation Model

Based on problems and data analysis results, we summarize the requirements to be considered in the VM propagation prediction model:

- (1) The model can reflect the interaction between virtual machine propagation and diffusion in various regions.
- (2) This model can reflect the influence of the historical development of the various regions on their future.
- (3) The model must be able to simulate the change of virtual machines in all regions.

Based on the above conditions, we chose the cellular automata model. Cellular automata, proposed by Stanislaw Ulam and John Von Neumann of Alamos National Laboratory, New Mexico, USA, is a dynamic system based on time-space dispersion, which is widely used to analyze diffusion and propagation problems [4]. A cellular automaton is represented by a regular grid, each grid represents a unit, and the state of each unit is finite (open or closed). The cells around a cell are called neighbors. After the initial state is set at time $t = 0$, the next time $t+1$ will produce a new generation of cells. The status of the new unit is determined by the common status of itself and its neighbors at the previous moment, and so on, until the status of each unit is updated. The cells in the system are scattered in the grid, and the changes of neighbors are influenced by local evolutionary rules. Finally, under the accumulation of time and iteration of space, it spreads to all parts of the system to realize the change of the whole state [5].

2.1 Cellular Automata Model

The definition of two-dimensional cellular automata is as follows:

$$C = (D_2, S, N, f) \tag{1}$$

D_2 represents a two-dimensional Euclidean space, S is the cell state set, N is the cell neighborhood, F is the cell

state transformation rule.

For the unit located in the R unit, its state at that time is expressed as :

$$S(r, t) = \{S_1(r, t), S_2(r, t), \dots, S_k(r, t)\} \tag{2}$$

In which k States of the cell located in the r lattice at time t are shown. The corresponding States in the VM prediction model can be divided into the area where there is no VM nesting, that is, Clean Land (CL); There is an area where VM nests and no measures are taken to clean it up at time t , that is, Invaded Land (IL); There is an area where VM's nest and measures are being taken to clean it at t moment, that is, Land Being Cleared (BL); And the land of hibernation (HB), where wolves are hibernating.

N_q is the q 's neighbor of the cell located at r .

$$N = \{N_1, N_2, \dots, N_q\} \tag{3}$$

According to the propagation mode of virtual state, the distance from a new queen to build a nest is estimated to be 30 km. 30 km is equivalent to three or four cells, so the grid with 7×7 neighbors is defined as shown in Figure 6.

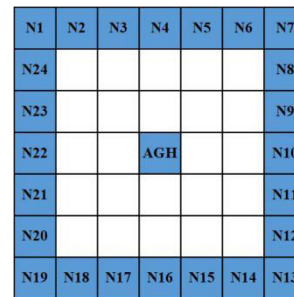


Figure 1. Cellular automata neighbor structure diagram

f is the transformation rule of the cellular state at position R from the time t to the time $t+1$. Where f_i represents the i th transformation rule.

$$f = \{f_1, f_2, \dots, f_m\} \tag{4}$$

If the current state of a cell is, then the j th transformation rule of its state at time $t+1$ is

$$S(r, t + 1) = f(S(r + N_1, t), S(r + N_2, t), \dots, S(r + N_q, t)) \quad j = 1, 2, \dots, m \tag{5}$$

First, randomly assign cells whose initial state is IL in a cell space of 530×1000 , and the number is less than 0.02% of the total cell space. Secondly, the cell space is traversed. According to the change probability determined by Robinet et al [6]. The evolution rules in the determined

topic are as follows:

Rule 1. If there are I cells in the states of IL and BL in the neighborhood of a cell in the state of CL, r will change to the state of IL with a certain probability ($i \times 0.14$).

Rule 2. At present, the cell in HB state will become IL state with a certain probability. (0.5).

Rule 3. At the present moment, the cell in the IL state will become a BL state with a certain probability. (0.3).

Rule 4. At the present moment, the cell in BL state will become CL state with a certain probability. (0.35).

Rule 5. All rules are executed at the same time.

2.2 CA Data Preparation

2.2.1 Spatial Data Interpolation

In the provided dataset attachment, there are 14 cases in which the wasp attribute is Positive ID in the report information, and each case report corresponds to different geographic information. It is very important to accurately interpolate the spatial attribute of its location (the probability of Asian giant bee VM invasion). From the spatial point of view, the more the locations tend to be the same, the more the probability of being invaded by VMs tend to be the same. The probability that two opposite points tends to separate is low^[7].

The purpose of using spatial interpolation is to supplement the spatial data that cannot be measured in space. In this topic, for various reasons, it is impossible to obtain the distribution of all wasps by measurement. At this time, interpolation technology is needed to simulate and generate these data, to understand the distribution of spatial regions as a whole.

The positive markers, untreated markers, negative markers, and unverified markers in this topic were assigned respectively, and the results are as follows:

$$\begin{aligned} Z_{PositiveID} &= 100 \\ Z_{Unprocessed} &= 25 \\ Z_{NegativeID} &= 1 \\ Z_{Unverified} &= 25 \end{aligned} \tag{6}$$

2.2.2 Grid Processing

The whole continuous space needs to be analyzed, so we grid the spatial data to make the analysis result better. To show the distribution of spatial data directly from the interface when analyzing spatial data, we generate isolines by interpolation, so that we can directly see the invasion probability of VM to different geographic locations from

the interface. Common meshing algorithms include rectangle meshing and triangle meshing. To make the algorithm simple and the program easy to implement, we use a rectangular grid and select natural neighbors to interpolate. The interpolation results are shown below.

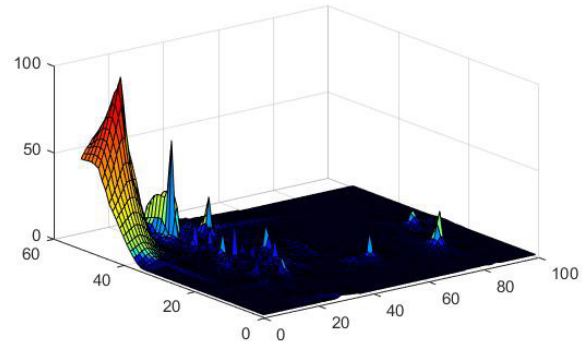


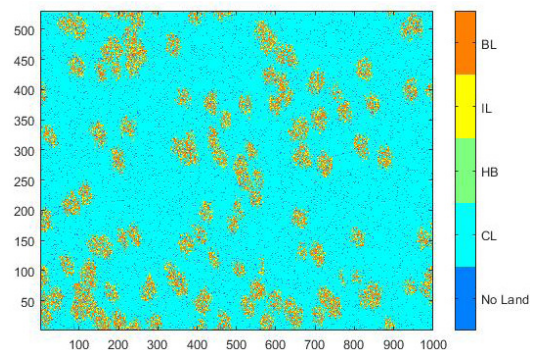
Figure 2. Probability spatial distribution of VM invasion

2.3 Analysis of Simulation Results

The results of the cellular automata simulation show that VM will spread around in an elliptical pattern with VM's current nest as the center, and the area of Clear Land will drop rapidly while the area of invaded land will increase in a blowout manner.

The predicted results showed that the growth rate of VM decreased after the 70th year, which might be caused by the following two factors: the population of VM increased sharply after the 70th year, the competition within the species of VM intensified, and the number of bees for the food of VM decreased; With increased government oversight, 60 percent of the VM nests found each year will be cleared. These factors resulted in a 17% decrease in VM transmission rate and a 29% decrease in population density^[6].

Because VM generally does not nest 3 to 6 feet above the ground, it usually nests underground. Therefore, it is difficult to find the nest of virtual machines and clean it up. Only 30% 40% of VM nests are found to be cleared every year; meanwhile, the cleaned nest may be rebuilt by other virtual machines, which worries residents^[6].



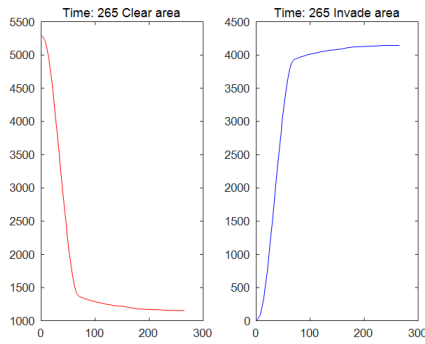


Figure 3. The transmission state of VM in 20 years and the number of CL cells and IL cells after 265 years

2.4 Washington Predicts Results

To avoid spatial autocorrelation caused by sampling deviation and over-fitting of specific areas, we take the reported geographical distance as the standard, refine the space based on distance sparsity, and exclude another event whose reported position is less than 5 km^[8].

The prediction result is as follows, taking the geographic location of the sparse positive ID as the initial value.

As shown in Figure 4, influenced by the topography of the Washington area, it is concluded that the nesting range of VM will be covered by ellipse centered on (122.93174°W,48.93457°N) after 20 years, and the farthest distance is 1184.4 km; The ellipse is centered at(122.57376°W,49.07848°N), covering 985 kilometers.

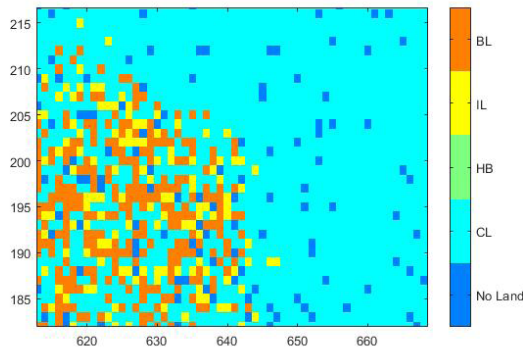


Figure 4. Propagation pattern of a single VM nest

Table 1. Positive ID Sparse Results

Latitude	Longitude
48.77753	-123.94313
48.92751	-122.81065
48.95558	-122.74501
48.97194	-122.70224
48.98099	-122.70094
48.99389	-122.68850
49.02583	-122.66103
49.06021	-122.64164
49.14939	-122.41861

2.5 Forecast the Resulting Hit Rate

According to the mode of transmission of VM, the nesting distance of new bees is estimated to be 30 kilometers. Taking 30 km as the standard, the spatial hit rate is defined as the ratio of the actual distance of each neighbor to 30 km in N_i ($i = 1, 2, \dots, 24$).

$$Hit_i = \frac{S_{N_i}}{30} \tag{7}$$

Determine the average hit rate:

$$Average Hit = \sum_i^{24} Hit_i / 24 \tag{8}$$

As shown in Table 2, Average Hit = 1.0012, and the closer the hit rate is to 1, the higher the accuracy of predicting VM spatial propagation by the model.

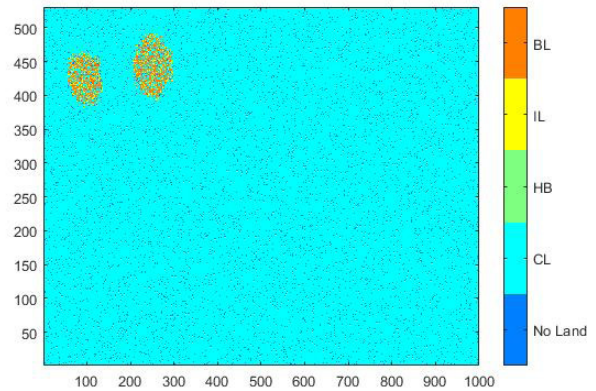


Figure 5. Forecast results of VM communication in Washington

Table 2. Space hit rate

Neighbour	N_1	N_2	N_3	N_4	N_5	N_6
Hit Rater	1.226575	1.042388	0.914235	0.867320	0.914235	1.042388
Neighbour	N_7	N_8	N_9	N_{10}	N_{11}	N_{12}
Hit Rater	1.226575	1.042388	0.914235	0.867320	0.914235	1.042388
Neighbour	N_{13}	N_{14}	N_{15}	N_{16}	N_{17}	N_{18}
Hit Rater	1.226575	1.042388	0.914235	0.867320	0.914235	1.042388
Neighbour	N_{19}	N_{20}	N_{21}	N_{22}	N_{23}	N_{24}
Hit Rater	1.226575	1.042388	0.914235	0.867320	0.914235	1.042388

3. The Establishment of Classification Prediction Model

Analyze the information provided on the topic based on time and space dimensions. To reduce the error rate of classification results, we try to establish a spatiotemporal comprehensive evaluation model with SVM and neural networks.

3.1 Support Vector Machine

Support Vector Machine was first proposed by Cortes and Vapnik in 1995, and it is widely used in linear and nonlinear classification. It shows advantages in solving the problems of a small sample, nonlinearity, and high-dimensional pattern recognition. Here, we set up an SVM data network [9]. The classification is based on the longitude and latitude of reporting points and discovery time of reporters, and the classification result is Lab Status.

3.1.1 Processing Flow

Different dimensions of data have different dimensions. To prevent the difference in the order of magnitude between input and output data, the original data is normalized first.

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (9)$$

Traditional SVM is only used as a linear classifier, while virtual machine classification is complicated, so it can't be classified visually by one line and one side. Here's a brief introduction to the kernel function.

General conditions are:

$$k(x, y) = \langle f(x), f(y) \rangle \quad (10)$$

$\langle x, y \rangle$ is the inner product of x and y . Using kernel function, the original space (Euclidean space R_n) is mapped to the new space (Hilbert space H), so that the hypersurface model of the original space corresponds to the hyperplane of feature space, thus completing the corresponding classification task [10]. Select linear kernel as kernel function and its inner product function is as follows.

$$k(x, y) = x^t y \quad (11)$$

After introducing the corresponding kernel function, the test set data can be used for classification training, and the trained SVM network can be used for classification testing of the original data.

3.1.2 The Classification Results

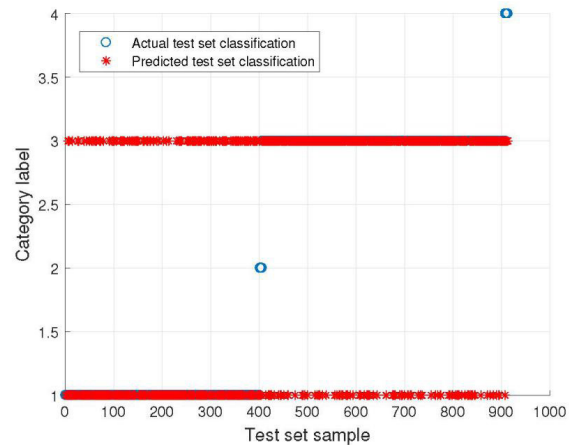


Figure 6. Actual & Predicted classification diagrams of the test set

As shown in Figure 6, witness reports are divided into four categories, represented by the numbers 4, 3, 2 and 1 respectively, corresponding to positive ID, negative ID, unprocessed and unverified respectively. Because the positive ID and unprocessed contain too little data, it is compressed to a single point in the figure, and most of the samples are negative and unverified.

To illustrate the classification accuracy, a new hit rate function, accuracy, is introduced. Accuracy is the proportion of correctly predicted classification result P_i to the actual category A_i .

$$Accuracy = \frac{P_i}{A_i} \quad (12)$$

The accuracy of classification is 74.2607%, which is low because there are too many differences in the number of samples.

3.2 Neural Pattern Recognition

The neural network is a mathematical model or computational model which simulates the mechanism and function [11] of the biological neural network and is used to estimate or approximate the function. It has been widely used. The neural network is applied to pattern recognition by using its adaptive characteristics [12]. The input variables are latitude, longitude and time reported, and the output variables are Lab Status.

3.2.1 Parameters Show

To make the model better predict the corresponding classification results, we selected 15 neurons in the hidden

layer and 4 nodes in the output layer and the training function used is (training by using scaled conjugate gradient backpropagation).70% of the data are selected for training, 10% for verification and 20% for testing. The model is structured as follows.

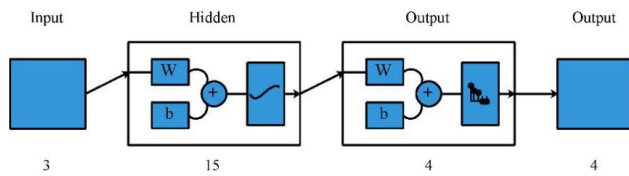


Figure 7. Schematic diagram of the model

After training, build the corresponding neural network relationship. See the following training rendering.

It can be seen from the below training curve that with the increase of training times, the error is decreasing. After 249 training times, the error has converged to the required precision and reached the global optimum.

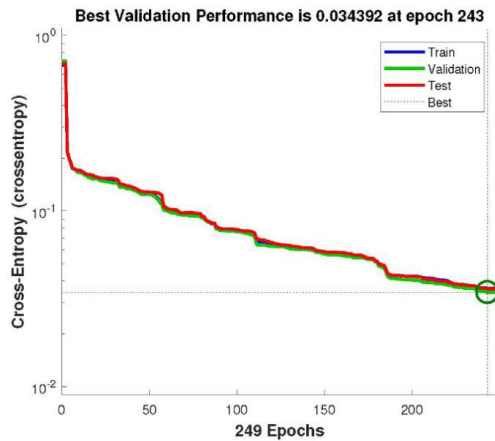


Figure 8. Training Curve

3.2.2 The Classification Results

After completing the training, the students enter the original data to obtain the original classification and prediction classification as shown in the following figure.

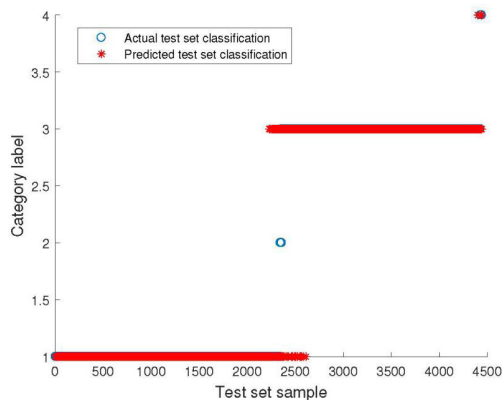


Figure 9. Neural network classification results

Eyewitness reports are also divided into four categories, and the classification is the same as the pictures. After calculation, the accuracy of the neural network is 97.6%. As can be seen from Figure 9, most of the eyewitness reports were correctly classified. The classification accuracy of the neural network is higher than that of SVM, and the normal distribution also has corresponding classification results, which shows a better prediction effect.

3.3 Results Analysis

After comparison of SVM classification and neural network classification, it is known that neural network has a better effect in classification prediction. Therefore, we choose to use a neural network to build the space-time comprehensive evaluation model. Input VM's location of the latitude and longitude, time, after the trained neural network, can get the accuracy of more than 97.6% of the classification results, if the positive ID, then need to send more staff to carry out field investigation.

4. Conclusions

According to the Vespa mandarinia invasion in Washington State, we established a diffusion model by using cellular automata. After the simulation and modeling, the results show that after 70 years, the growth speed of virtual machines slows down, resulting in a 17% drop in the propagation speed of virtual machines and a 29% drop in population density. In order to make better spatio-temporal prediction, we used support vector machine and neural network to establish mapping relationship between the location and time of the witness and the local wasp situation, and then establish network mapping for classification verification. The results show that neural network is superior to the support vector machine in classification prediction.

References

- [1] Requier, Fabrice, et al. "Predation of the invasive Asian hornet affects foraging activity and survival probability of honey bees in Western Europe." *Journal of pest science* 92.2 (2019): 567-578.
- [2] Beggs, Jacqueline R., et al. "Ecological effects and management of invasive alien Vespidae." *BioControl* 56.4 (2011): 505-526.
- [3] Tan, Ken, et al. "Honey bee inhibitory signaling is tuned to threat severity and can act as a colony alarm signal." *PLoS biology* 14.3 (2016): e1002423.
- [4] Cole, Vivienne, and Jochen Albrecht. "Modelling the spread of invasive species: parameter estimation using cellular automata." In *Proceedings Second In-*

ternational Workshop on Dynamic and Multi-Dimensional GIS (DMGIS'99. 1999.

- [5] Gao Jianhua. Several Improvement and Application of Evolutionary Strategy.2015. Wuhan University, Ph.D. dissertation.
- [6] Robinet, Christelle, Christelle Suppo, and Eric Darrouzet. "Rapid spread of the invasive yellow-legged hornet in France: the role of human-mediated dispersal and the effects of control measures." *Journal of Applied Ecology* 54.1 (2017): 205-215.
- [7] AbdelRahman, Mohamed AE, et al. "Deciphering Soil Spatial Variability through Geostatistics and Interpolation Techniques." *Sustainability* 13.1 (2021): 194.
- [8] Nuñez-Penichet, Claudia, et al. "Geographic potential of the world's largest hornet, *Vespa mandarinia* Smith (Hymenoptera: Vespidae), worldwide and particularly in North America." *PeerJ* 9 (2021): e10690.
- [9] Feng Guohe." Comparison of kernel function and parameter selection for SVM classification." *Computer Engineering and Applications*, 47.03(2011):123-124+128.
- [10] Sze Bi, Tinglei Huang." Design of Support Vector Machines Based on Linear Distance Kernels." *Journal of Beijing University of Electronic Science and Technology* (2013):478-481.
- [11] Li Song, Liu Lijun, and Zhai Man." Improved Particle Swarm Optimization Algorithm for Short-Term Traffic Flow Prediction Based on BP Neural Network." *Journal of Systems Engineering* 32.09(2012):2045-2049.
- [12] Li Xiaoyan. Research on the Optimization of Grey Neural Network Prediction Model.2009. Wuhan University of Technology, MA Thesis.