ISSN 2616-5872 Vol.3, Issue 1: 47-51, DOI: 10.25236/AJEE.2021.030110

A study of species distribution prediction based on kernel density estimation

Beixin Fang, Yanling Jiang, Yimeng Cao

School of Mathematics and Statistics, Lanzhou University, Lanzhou, Gansu 730000, China

Abstract: *The occurrence of Vespa mandarine in the State of Washington is terrible news for the local Western bees, as Vespa mandarine is a very effective predictor. The Washington State government is taking the Vespa mandarine invasion seriously and hopes that people will report it when pests are found. This article processes the sighting reports provided, establishes a model, predicts the potential distribution of coriander, and can effectively provide the possibility of a new report being positive so as to quickly determine the most likely positive report and help the government allocate resources rationally. First, a species distribution prediction model based on kernel density estimation (KDE) was established. We took the Gaussian function as the kernel function, performed KDE on the given data, and controlled the search radius of the kernel function to be 30km from the center of the nest to obtain a probability density heat map reflecting the possibility of coriander spreading to a certain area, which shows that the scope of the spread and the possibility of each location are predictable. It is also essential to consider the circumstances under which the hives are cleared. Therefore, when a hive is removed, we will update the KDE model. If the distance is less than 30km, the point closest to the coordinates of the cleared hive will be deleted.*

Keywords: *Kernel Density Estimation, Computer Vision, Vespa mandarinia*

1. Introduction

Vespa mandarine is the world's largest hornet and one of the most dangerous carnivorous insects [1]. It is native to temperate and tropical East Asia, South Asia, Mainland Southeast Asia, and parts of the Russian Far East. In late 2019, it was also found in the Pacific Northwest of North America, with a few more additional sightings in 2020, In Washington State in the United States and British Columbia in Canada, it has raised concerns that it may become an invasive species. Vespa mandarinia can devastate a colony of honey bees, especially dangerous for the introduced western honey bee because they lack the ability to protect their hives and fight back; a single Vespa mandarine can kill up to 40 bees in a minute due to its large mandibles, which can quickly decapitate its prey. As honey bees play an important role in the pollination process of crops, the spread of Vespa mandarine in the United States could severely impact agriculture and human health as well as the economy. Continuous monitoring to find and subsequently eliminate any discovered colonies is of prime importance to prevent Vespa mandarine establishment [2]. The Colony cycle of Vespa mandarine starts from mid-April after hibernation. Inseminat queens begin to search for nesting sites in late April [3]. Until early August, Queens mark a fully developed nest with about 100 workers. After mid September, queens will no longer lay eggs and die in late October. Then new queens will hibernate in moist underground habitats and repeat the cycle next year.To help to discover and locate the nest, the State of Washington has created helplines and websites for people to report sightings of Vespa mandarinia. However, many reports are mistaken Vespa mandarinia with other types of insects, which caused a waste of resources of government agencies. Thus, this is a model that can predict the spread of this pest over time.

2. Species distribution Prediction Model Based on Kernel Function

2.1 Principles of Kernel Density Estimation

Determine the center of the kernel function. On every data point, on which we will place a kernel function K later (Figure3). In our case, consider the data points as the locations of nests in One-dimensional space.

Place the kernel function on the data point. The kernel density estimate is:

ISSN 2616-5872 Vol.3, Issue 1: 47-51, DOI: 10.25236/AJEE.2021.030110

$$\hat{f}(x) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{x - x_i}{h}\right)$$

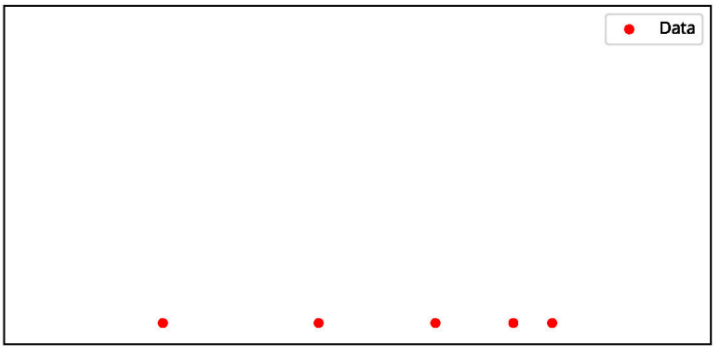

Figure 1: Data points

The KDE model established in this paper use Gaussian as kernel function placed on the data points, as shown in Figure4. A Gaussian Kernel is defined as,

$$K(x, x\prime) = e^{\left(-\frac{\|x - x\prime\|}{2\sigma^2}\right)}$$

The bandwidth, which is controlled by h, determines the search radius of the kernel function. When h gets larger, it will spread the kernel function, which leads to a search radius increase. The choice of bandwidth, in this case, is determined by the reproductive habits of Vespa mandarinia.

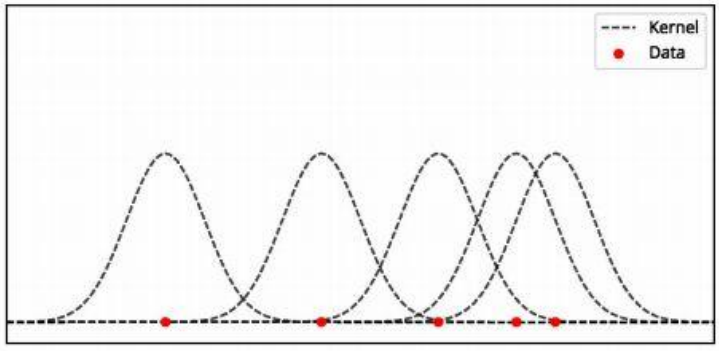

Figure 2: Kernel functions placed on data points

Calculate the Kernel Density Estimation. For a single kernel function, the density of the center of the kernel is the highest, and the edge can reach minimum to zero(depending on the kernel choice and if the kernel function has bounded support). To get the overall kernel density estimate, we sum the above kernel functions together, and then we normalize the estimate since every kernel function must have an integral evaluation to one. Now we have the kernel estimate curve, which is shown in Figure5.

Illustrate the model in 3D. The Vespa mandarinia built nests underground, which can be considered as a two-dimensional plane. Then consider two Vespa mandarinia nests with the coordinates (-2, -2) and (2, 2), and the range of building new nests for Vespa mandarinia is 2. Use the above steps to build a Kernel Density Estimate model using Gaussian kernel and use parameter $\hbar$ to control the kernel radius around 2. After we placed the kernel of the data points and calculated the Kernel Density Estimation (shown) in Figure 3, we could easily find out the possibility of a nest being established at certain coordinates.

ISSN 2616-5872 Vol.3, Issue 1: 47-51, DOI: 10.25236/AJEE.2021.030110

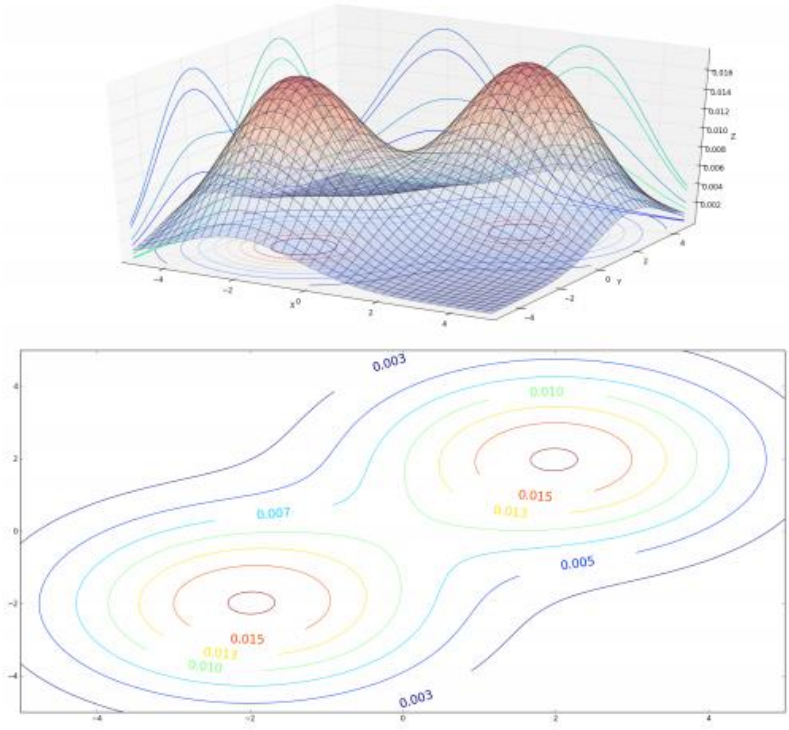

Figure 3: A Kernel Density Estimate based on two hypothetical nest locations

2.2 Fit Kernel Density Estimation

In this section, we will build the Kernel Density Estimation model using the given data and other reference information. First, we can take a rough look at the existing nest locations by scattering them on the map of the U.S. state of Washington and British Columbia and Canada. After summing up and normalizing the probability density of kernel functions placed on each point, we created a heat map of probability density as an additional layer under the scatter plot, which is shown in Figure 4. In order to get the possibility of a nest may be built in a certain location, we need to preform an integration over probability density functions that generate by the KDE method. Because our Kernel Density Estimation is preformed on a two-dimensional plane, we need to integrate over an area, which concerns the choice of R_I. To obtain an approximation of the probability at that point, we can take $R_I \to 0$ for integration. We could also use R_I= 30km to find a more robust result considering the biological meaning.

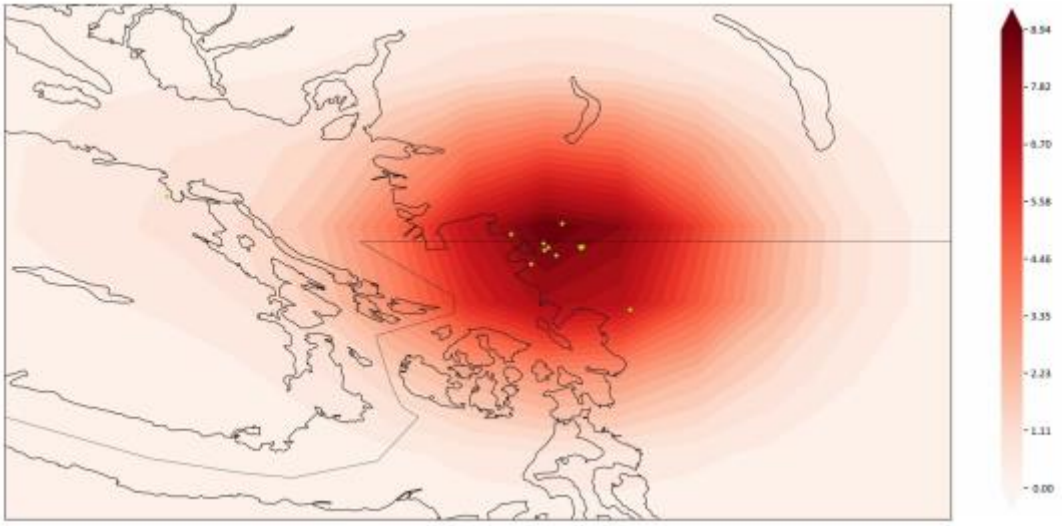

Figure 4: Probability density heat map

ISSN 2616-5872 Vol.3, Issue 1: 47-51, DOI: 10.25236/AJEE.2021.030110

The probability value of predicted spread based on each year of data is used as the sample data. Since the data are non-negative with annual measures and only one year of data is currently available and not strongly correlated with other data, we can use the gray prediction model [4] to analyze the accuracy of the prediction results.The sample data is entered as an original series, and then a new series is obtained by accumulating the original series, followed by the generation of the immediate mean of the new series. Finally, the development coefficient is calculated using the least-squares method. If the development coefficient is smaller, the accuracy of the prediction is higher.

3. Image Classification Model Based on Computer Vision

The primary objective of this classification is to identify the positive image, which means to recall as much positive sample as possible. So we could combine the label Negative and Unverified as Non-Positive to help just identify the positive sample. After classifying the image datasets to Positive/Non-positive, we use TensorFlow to train M_{cv}.

The primary objective of this classification is to identify the positive image, which means to recall as much positive sample as possible. So we could combine the label Negative and Unverified as Non-Positive to help identify the positive sample. The Confusion Matrix generated from the results of the test set is shown in Figure 5.

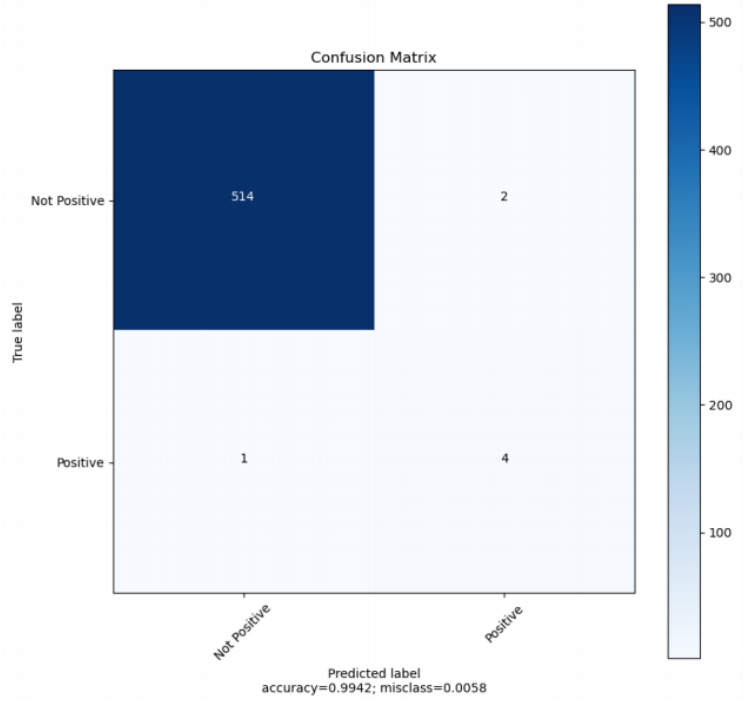

Figure 5: Confusion matrix

4. Conclusion

By establishing the prediction model of a species distribution based on Kernel Density Estimation and an image classification model based on deep learning, the prediction of the spread of the Vespa mandarinia and how to determine the optimal strategy through classification can be summarized as follow: using existing sample data to make a KDE species distribution map. From the map and KDE value, the spread of the Vespa mandarinia is predictable.

References

[1] Asian Giant Hornet | National Invasive Species Information Center. (n.d.). National Invasive Species Information Center. Retrieved February 8, 2021, from https://www.invasivespeciesinfo.gov/terrestrial/invertebrates/asian-giant-hornet

[2] Blonder, B., Lamanna, C., Violle, C., Enquist, B. J. (2017). Using n-dimensional hypervolumes for

ISSN 2616-5872 Vol.3, Issue 1: 47-51, DOI: 10.25236/AJEE.2021.030110

species distribution modelling: A response to Qiao et al.(). Global Ecology and Biogeography, 26(9), 1071-1075.

[3] Beazley, L., Kenchington, E., Lirette, C. (2017). Species distribution modelling and kernel density analysis of benthic ecologically and biologically significant areas (EBSAs) and other benthic fauna in the Maritimes Region. Ocean and Ecosystem Sciences Division, Maritimes Region, Fisheries and Oceans Canada, Bedford Institute of Oceanography.

[4] Xiangyun Liu, Hongqin Peng, Yun Bai Lueling Liao (2014). Tourism Flows Prediction based on an Improved Grey GM(1,1) Model. 138767-775.