Problem Chosen	2021	Team Control Number
C	MCM/ICM Summary Sheet	**2106873**

Determine the buzz by discriminant

Using Information Evaluating Model to prioritize resource for investigation of Asian giant hornet

Summary

The Asian giant hornet, with its "invasion plan", crossed the bay and sneaked into Washington's territory. With plenty of food and comfortable temperature, Vespa mandarinia is eager to turn the land into its own "kingdom". However, the Washington government discovered its "ambition" and immediately took action to call on local people to report if they sight the "invaders", but due to the limited resources and time constraints, priority need to be given to looking into the most likely reports of Vespa mandarinia.

To address this issue, we **first** investigate the correlation between pest habitat expansion and time, and analyze the correlation between pest population and time, using Spearman coefficient on an existing **Regression Model** with 90% certainty that pest spread is related to time, followed by the image of the fitting curve to make preliminary predictions. We use a **cellular automata model** to simulate and predict the spread of hornets in space as well.

Secondly, we designed our own **Information Evaluating(IE) Model** for calculating the probability of mistakes in reports. We divide the mistakes of reports into three types: Regional mistake, Seasonal mistake, and Textual mistake. For the regional mistake type, the **K-means++**[1] clustering algorithm is used to classify the positive 14 data into two clusters, and then the **TOPSIS(Technique for Order Preference by Similarity to an Ideal Solution)**[2] algorithm and the mean method are used to calculate the probability of mistake in each report in terms of region. For the seasonal mistake type, the probability of mistake of report in each season is found to be 0.979, 0.696, 0.566, and 0.761. For the Textual mistake type, inspired by the existing **CE-VADER**[3] method, we use the extraction of feature phrases from the text to give the probability of mistake of the text. Then, we used hierarchical analysis to give the weights of 0.31, 0.62, and 0.07 for each type of mistake, and calculate the probability of final mistakes in each report. Then it is examined and proved the robustness of the model.

Thirdly, we use our model to provide a priority for the professionals: the smaller the mistake probability of report is, the higher the priority for investigation is. **Furthermore**, we also give the frequency of updating the model and the theoretical basis for eradicating hornets. We also perform a sensitivity analysis on the IE model we use and find that the weight of each type if mistakes has a significant effect on the final probability of mistake in the report.

Finally, we write a memorandum that summarizes our work and some advice of how to prioritize the limited resources for the Washington State Department of Agriculture.

Keywords: Spread of Vespa Mandarinia; Information Evaluating Model; Pest Prevention; Report Priority

Contents

1 Introduction

Which is known to all, the invasion of alien species will bring serious damage to the local ecological environment, biological diversity and economy. In September 2019, a colony of Vespa mandarinia (also known as the Asian giant hornet) was discovered on Vancouver Island in British Columbia, Canada. The nest was quickly destroyed, but the news of the event spread rapidly throughout the area. Later, the pest was confirmed to be observed in neighboring Washington State. The reproduction of Asian giant hornet and its impact on agriculture caused serious anxiety among the local people. Washington State has set up a helpline and website for people to report sightings of these hornets. However, because of limited human and financial resources, the state must decide how to prioritize its limited resources for further follow-up investigations. Although some reports have been identified as hornets, many other sighted proved to be other types of insects.

In order to improve the efficiency of the Washington government, we need to design a system for preliminary judgment of the reliability of the report given by the local people, which can be used to quickly find effective information and initially judge the reliability of report, and prioritize the given report for the specialists to process its verification. To achieve this goal, we establish two models. The first is to add spatial differences to the existing regression model; the second is an evolvable model that predicts the likelihood of a mistaken classification, which classifies the reports by their characteristic (key info, quality of the images in the report, etc.).

In this article, we utilize these models to predict the spread of Asian giant hornet in time and space, interpret the data provided in the public reports, and prioritize these public reports for additional investigation. Robustness of our models is examined by processing the "unprocessed reports" and comparing the result with the existing one. This two-pronged approach helped us rationally analyze the hornet crisis, and provided a method of information processing and interpreting, which help determine the eradication strategy of Washington department.

1.1 Problem restatement

Problem 1 Set up a concrete method for prediction of the spread of Asian hornet, using correlation to prove its feasibility, and provide its level of precision.

Problem 2 Because people mostly give classification of the insects based on empirical evidence and lack of knowledge of the Asian hornet, we establish a model to provide the likelihood of the mistaken classification.

Problem 3 Utilize our model to prioritize the report for the department staffs to investigate.

Problem 4 Determine the method and frequency of update of our model.

Problem 5 Tell whether the hornets have been whipped out based on our model.

1.2 Background

1.2.1 Appearance of Asian hornet

Asian giant hornet is the largest hornet species in the world. Workers range from *25 to 40 millimeters in length*. The hornet's head is a light shade of orange and its antennae are brown with a yellow-orange base. Its eyes and ocelli are dark brown to black. V. mandarinia is distinguished from other hornets by its pronounced clypeus and large genae.[4]

1.2.2 Habits and life cycle of the hornet

Like many other wasps, fertilized queens emerge in the spring and begin a new colony. In the fall, new queens leave the nest and will spend the winter in the soil waiting for the spring. When winter arrives, the current seasons' nests die out and emphthe only individuals that survive are overwintering queens. *A new queen has a range estimated at 30km for establishing her nest. The maximum foraging range of workers is 8 kilometers. Asian giant hornets always build their nest under the ground, while other local bees build aerial nest.*[4]

1.2.3 Predation of the hornet

Asian hornet hunts primarily arthropod prey, such as bee, scarab and longhorn beetles, large caterpillars, and spiders. After catching its prey, a hornet will decapitate it, chew the more muscular portions of the thorax into a ball, and then fly back to the nest. Asian hornet, however, does not eat bee itself. Once the local bee colony's defenses are overwhelmed, hornet will occupy the nest and harvest the protein-rich brood, leaving the dead bodies of bees alone.[4]

1.2.4 Other usual predators of local bees

Bears eat the honey in the hive, without killing the bee itself.

Skunks, bee-eater birds, crab spiders eat the bee itself, swallowing and digesting.[4]

1.2.5 Condition of Washington

Warm and hard-planted zones are ideal habitat for Asian giant hornet. Most parts of Washington are suitable for Asian hornets to live in. Figure 1 given below is based on the record from Global Biodiversity Information Facility (GBIF).[4]

2 Foundation of models

2.1 Assumptions

1. The whole Washington is suitable for Asian giant hornet to proliferate and every spot of Washington can be the suitable place for nest build up.
2. Every dead body of local bees located in the radius of 8 KM from the hornet nest indicate the presentation of a hornet.
3. Local people provide as much details of what he or she sees as possible based on his or her knowledge of hornets while reporting.
4. Not consider the impact of natural or manual effect on proliferation of hornets, except the changes of season.

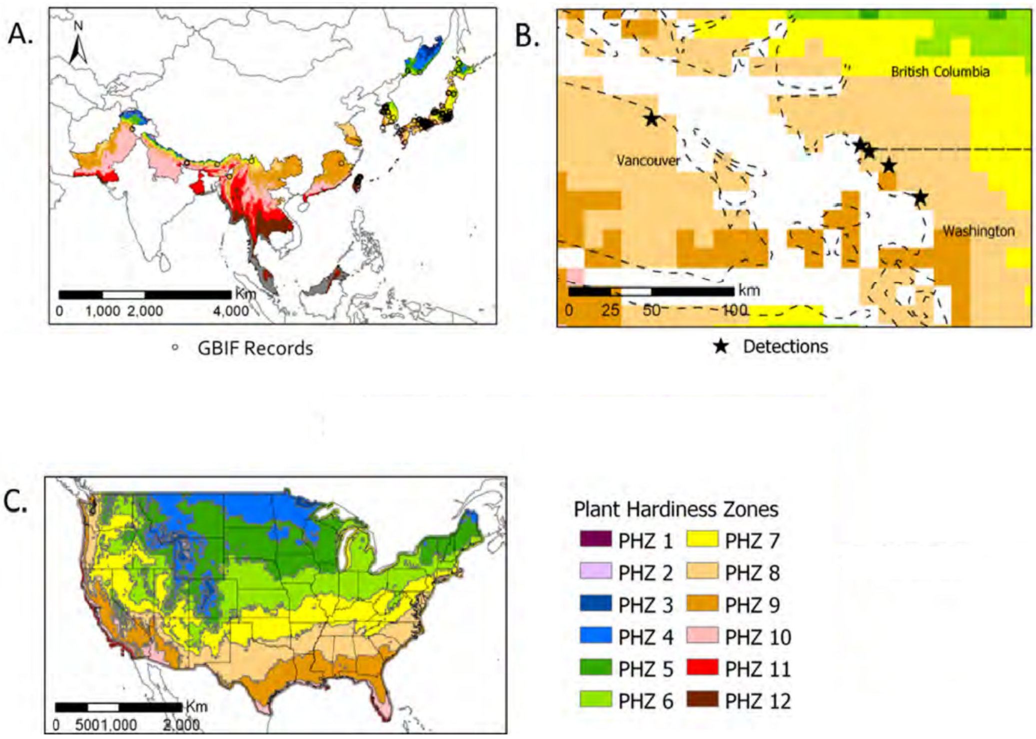


Figure 1: Plant Hardiness Zone maps; A. distribution of giant hornet in Asia; B. recent detections of giant hornets in America.

2.2 Data set processing

1. Discard the data which is before 2019 and the data reported in Canada.
2. Merge [Dataset.xlsx] and [Images by GlobalID.xlsx] into one file.

2.3 Nomenclature

For convenience, we use the following symbols in our article, cf. Table 1.

Table 1: Symbols and their Descriptions

Symbols	Definitions
X	Month
Y	Occurrences of hornets
H_1	Null hypothesis
H_0	Hypothesis of correlation of X and Y
N	Total sample
P_1	Probability of regional mistake
P_2	Probability of seasonal mistake
P_3	Probability of textual mistake
P	Probability of mistaken classification
P_e	Normalized P
r_s	Spearman correalation coefficient
ω	Weight vector

3 Our models

In order to discuss whether the spread of the pests can be predicted, we decide to use regression model and study the correlation between occurrences and time. Considering the geometry expansion of the pest, regional differences are added to the model.

3.1 Regression model

3.1.1 Geometry expansion

A new queen hornet builds a new nest within a radius of 30 kilometers. See Section 1.2.2. Using the positive ID data in the given data set, we plot a expansion trend of hornets from the beginning of 2019 to the end of 2020 in Figure 2. As is shown in the Figure 2, locations of nests of the hornets varies as time rolls on.

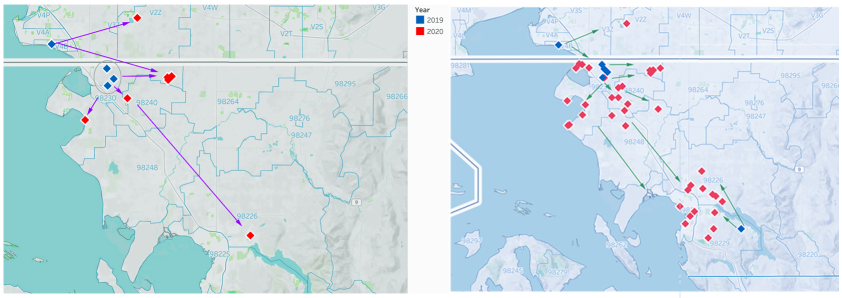


Figure 2: Expansion of the hornet nests. Blue - reported in 2019; Red - reported in 2020.

3.1.2 Augmentation of the sample

In view that there are only 14 positive ID in the data set, there are too few “direct” samples to illustrate the correlation. In order to obtain more samples, we take the following measures:

1. According to Figure 2, it can be seen that the nest of the hornets might relocate over time, which can prove its regionality. Meanwhile, according to background 1.2.2, we know that the foraging range of the hornets is 8 kilometers. Based on the above factors, we believe that the 8km radius of the vespa nest should be a high-probability area for occurrence of hornets.
2. Utilizing Matlab to convert the latitude and longitude into distance between reported location and confirmed location, we screen out reports from 4440 reports by whether the distance is under 8 km, and finally get 255 reports that are in high-probability areas.
3. In these reports, we notice that some of the unverified and negative ID report might directly or indirectly indicate the occurrence of hornets, because they contain some of the unique characteristics of Asian giant hornet given in 1.2.3. So we manually screen the unverified and negative texts by those unique characteristics, and get 46 reports that prove the existence of hornets. The screened-out reports are shown in appendix B.

3.1.3 Analysis of correlation

We extract the date info from the 46 reports, take time variable X with the interval of one month and occurrence variable Y, and then take the month X from 2019 to 2020 as the abscissa, and the occurrences Y as the ordinate to set up a coordinate system. After that, we plot a scatter diagram on the system in Figure . From the figure 3, it can be indicated that X and Y are not linear, so we use Spearman correlation coefficient instead of Pearson correlation coefficient.

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

d_i is difference between rank oders of X_i and Y_i

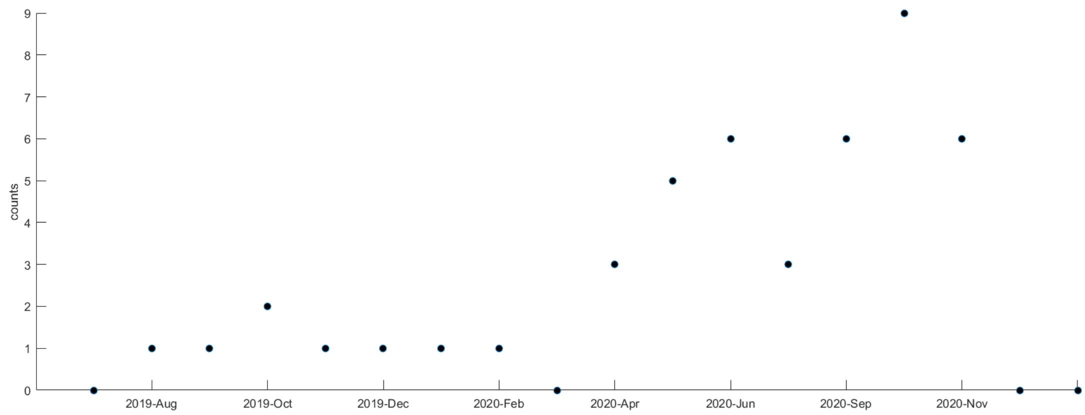


Figure 3: Occurrences of Asian hornetd over time

By calculating, r_s is 0.46. By checking the Spearman critical value table A, r_s>0.41, it is concluded that we can treat the null hypothesis H_0 as false and H_1 as true at the 90% confidence coefficient. It is believed that the number of occurrences of the hornets and time are significantly correlated.

3.1.4 Data fitting and cellular automata

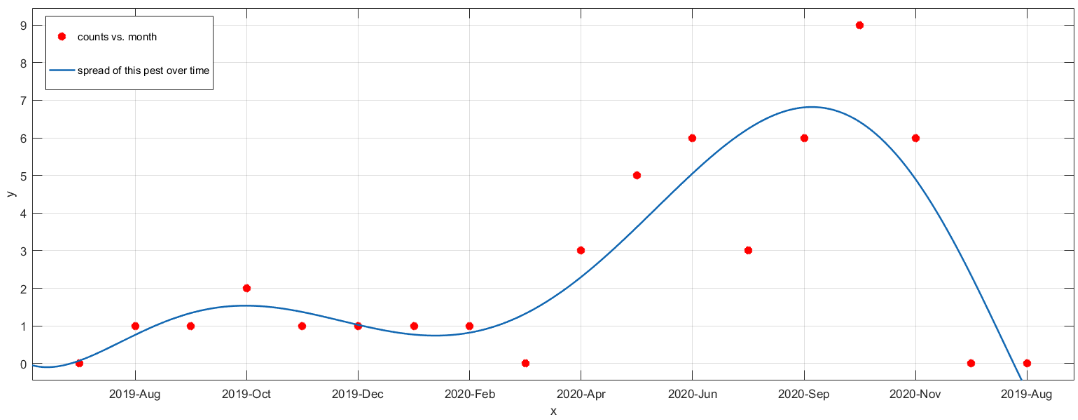


Figure 4: The fitting curve.

A polynomial regression curve fit of Figure 4 is performed using matlab, and the goodness of fit was 83%. The fitting curve is plotted in Figure 4. It can be seen from Figure 4 that the number of occurrences of the hornets varies with time. The number of occurrences of hornets reached a peak in summer and almost decreased to zero in winter, which indicated that the number of hornets meet with Section 1.2.2 and the fitting result is in line with the expected effect.

What stands out in the Figure 4 is that, with the change of time, the peak number of hornets will continue to increase while maintaining the existing conditions, but they will die in winter, leaving only the queen to hibernate. At the same time, we also perform a prediction of spatial distribution of the hornets, using cellular automata, as shown in Figure 5.

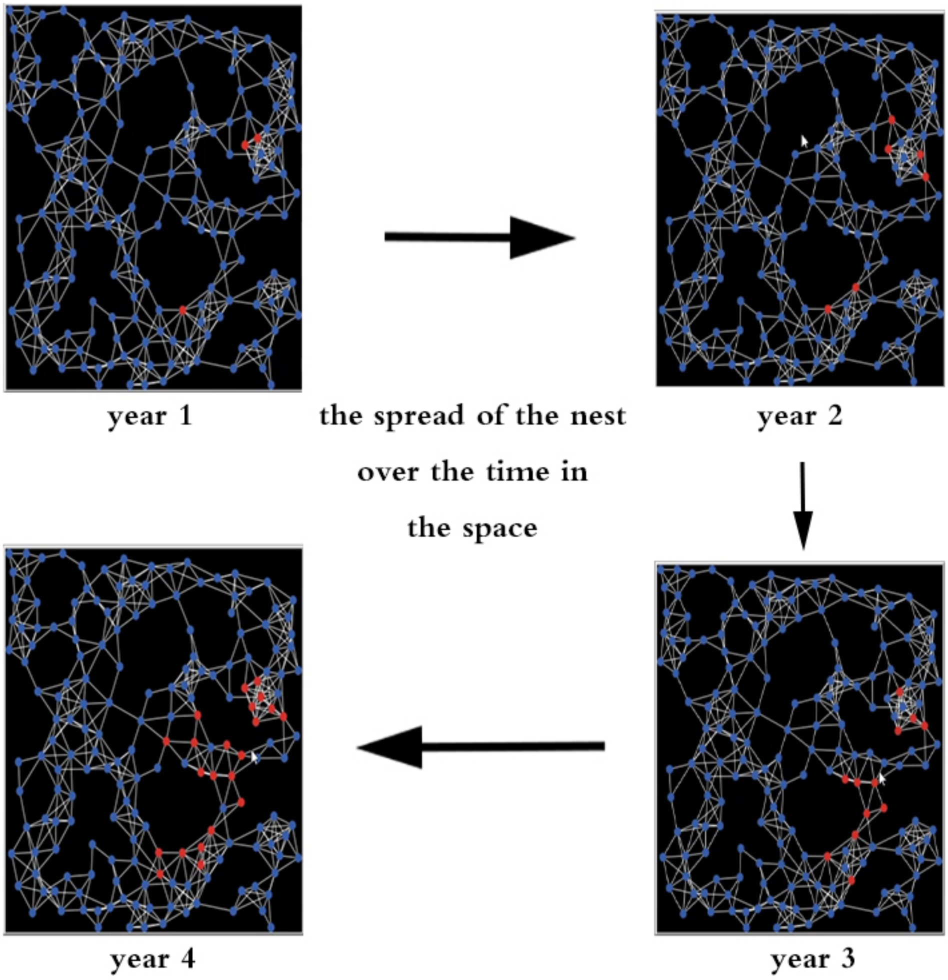


Figure 5: Spatial distribution of hornets

3.2 Information Evaluating model

Due to the lack of knowledge of Asian giant hornets, it is common that local people confuse Asian hornets with other local bees. Therefore, to save limited resources of the state department, we establish an Information Evaluating Model that can interpret the text in the sent reports and give the likelihood of a mistaken classification.

We conclude three sorts of mistaken classifications - regional mistake, seasonal mistake and textual mistake.

3.2.1 Data processing

In the data set, there are three valid types of reports – Positive ID, which are modified by the department staff to clearly be the presentation of giant hornets, Negative ID, which are modified by the staff to be definitely not the presentation of giant hornets, and the Unverified, which staff is not able to modify because of the low quality of the picture.

However, local people define whether the insect is a hornet at the very first stage based on his knowledge of hornets using their eyes, which outmatch any device or machine that can record an object. The reason why a report is unverified at last is that staff cannot modify, instead of classification itself being mistaken. So, the unverified data can not reflect on the validity of the report.

As a result, we screen out the Unverified data, use only Positive ID and Negative ID data.

3.2.2 Regional mistake

From background 1.2.2, we know the scope of activity and nesting of hornets, so the existence of hornets has a close relationship with the region. At the closer area of the nest, the more active hornets are, which leads to the greater the possibility of sighting the hornets, and lower possibility of a report to be mistaken. Conversely, the farther away from the nest, greater the likelihood of report mistaking is.

K – means ++ cluster algorithm If we calculate the distance from sighting spot to the 14 confirmed nests separately, and finally get an average distance, the amount of calculation is very large. In order to simplify the calculation, we first use the k-means++ clustering algorithm to simplify the calculation. The algorithm step is shown in Figure 6.

We separate the 14 data into two clusters using SPSS. cf. Table 2.

Table 2: Two clusters of data

Cluster1	Cluster2
48.77753	−122.41861
48.98596	−122.65317

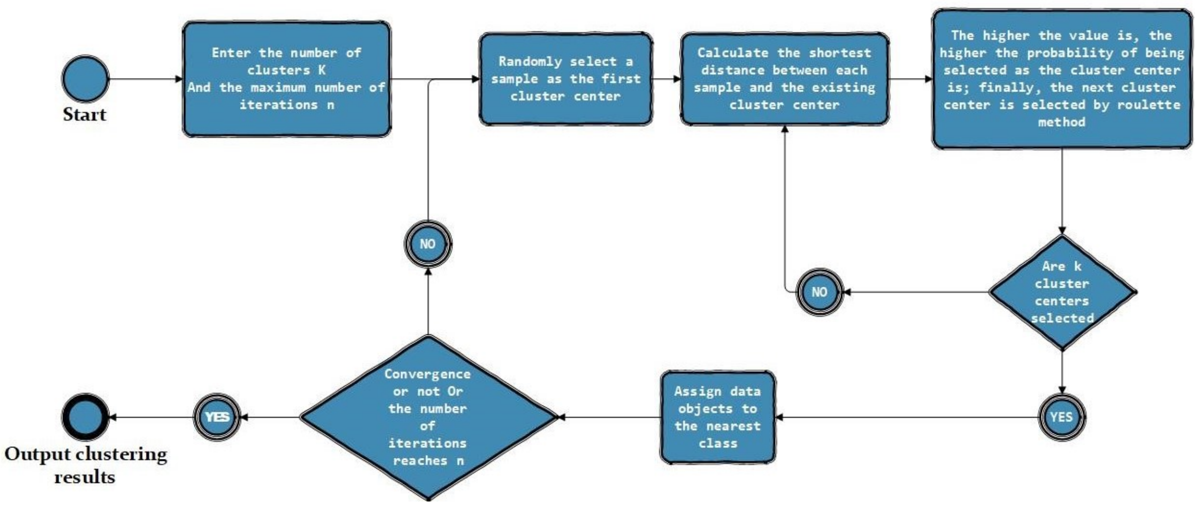


Figure 6: K-means flow chart.

TOPSIS algorithm In order to calculate the likelihood of a mistaken classification on regional mistake from the distance from each sighting spot to the nests, we use the TOPSIS algorithm, dividing the distance s from each point to the nests by the distance from the farthest sighting spot to the nests to get the probability P_0.

$$P_0 = \frac{s - min}{max - min}$$

Here, $min = 0$, max is the distance from farthest sighting spot to the nests.

Bigger the P_0 is, the longer the distance is. Smaller the value is, the closer the sighting spot is to the nest.

Method of mean value Calculate the P_0 from each sighting spot to two cluster points, and add them to get the average value to get the report's regional mistake probability P_1 . The closer P_1 is to 1, the more likely report to be mistaken, because the spot is very far from the nest and it is almost impossible to find a hornet. Obviously, the closer to 0, the lower the possibility of regional mistake. See Figure 7.

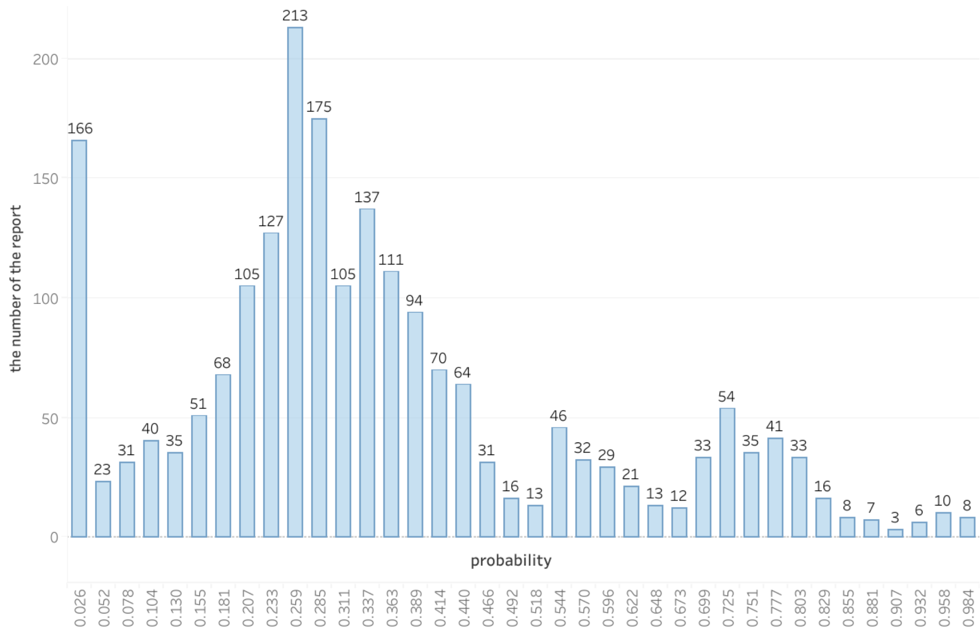


Figure 7: Regional mistake frequancy

3.2.3 Seasonal mistake

As shown in Figure 8, hornet almost cannot be found in winter.

We calculated the possibility of seasonal mistake in each season, and the result is shown in Table 2. The summer temperature is suitable and the hornets are proliferating. Likelihood of seasonal mistake is less, as the probability of a reporter sighting a hornet is greater. On the contrary, only the queen bee is left to hibernate in winter while other hornets die out.See Section 1.2.2. It is almost impossible for the reporter to come across a hornet, and the probability of mistake is extremely high.

For each report, we give a likelihood of seasonal mistake P_2 based on report's date.

3.2.4 Textual mistake

As Assumption 3, the reporter will describe what he sees in as much detail and accuracy as possible based on his knowledge of the wasp. In other words, if the report contains many details, the reporter has a better knowledge of hornets and the report is more likely to be the presentation of hornet.

Inspired by the existing CE-VADER model, we adopt the method of filtering by characteristics of the text, extract keywords in the data based on Section1.2, and find 10 most frequently occurring groups of words which reflect on the presentation of hornets. For each report, the more characteristics the report has, the more likely the reporter did sight a hornet. On the contrary, if the key words mentioned in the report are few or no key word, the less evidence reporter used to prove that it was a hornet, which means, reporter knew little about hornets, and the likelihood of mistaken classification is greater.

Size Large, big, huge, giant, largest, inches, feet, long, inch, size.

Appearance Orange, blank, yellow, red, brown, bright, singer, leg, stripes.

Behavior Hit, stung, kill, killer, murder, murderer.

Habitat ground, hive, under, holes.

For each report, we give a probability of textual mistake P_3 based on the occurrence of characteristics.

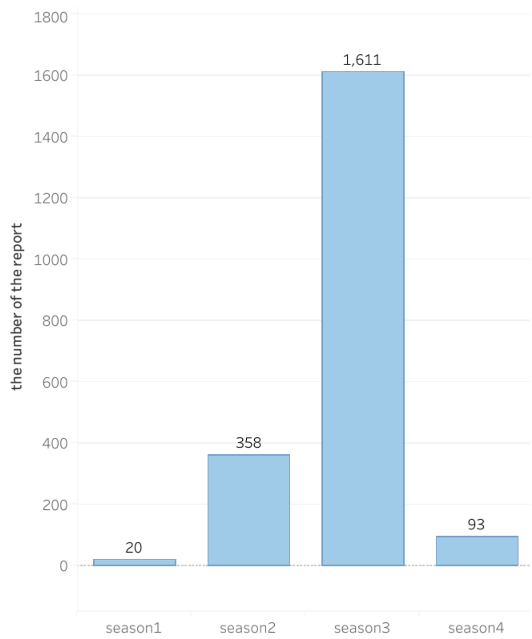


Figure 8: Season histogram

3.2.5 Likelihood of a mistaken classification of a report

After we get the report's mistake probability of season, region, and text, we decided to give the final likelihood of a mistaken classification using Analytic Hierarchy Process.

Let the judgment matrix $A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$

Then, using arithmetic average method, weight vector $\omega_i = \frac{1}{n}\sum_{j=1}^{n} \frac{a_{ij}}{\sum_{k=1}^{n} a_{kj}}$

After times of evaluation and observation with different value, we give the judgment matrix A 3.2.5 and perform a consistency check.

The result of the check CR is 0. If CR < 0.1, we consider the judgment matrix acceptable. At last, through the calculation, we get the Regional weight (ω_1) as 0.310, Seasonal weight (ω_2) as 0.620, Textual weight (ω_3) as 0.070.

Adding up P_1, P_2 and P_3, we get the final likelihood P. The greater the P is, the more likely the classification is mistaken.

3.2.6 Robustness of the IE model

In order to examine the robustness of our IE model, we use the method of proof by contradiction. If there are few cases identified as positive while P is large, then the model is robust.

Table 3: Judgement matrix A.

Judgement matrix	region	season	text
region	1.000	0.500	3.000
season	2.000	1.000	6.000
text	0.33	0.017	1.000

Table 4: Normalized judgement matrix A.

Judgement matrix	region	season	text
region	0.300	0.330	0.300
season	0.600	0.659	0.600
text	0.100	0.011	0.100

We use the TOPSIS method to normalize all probability of a mistaken classification as P_e. Then, we plot Positive ID data and its P_e in figure 9.

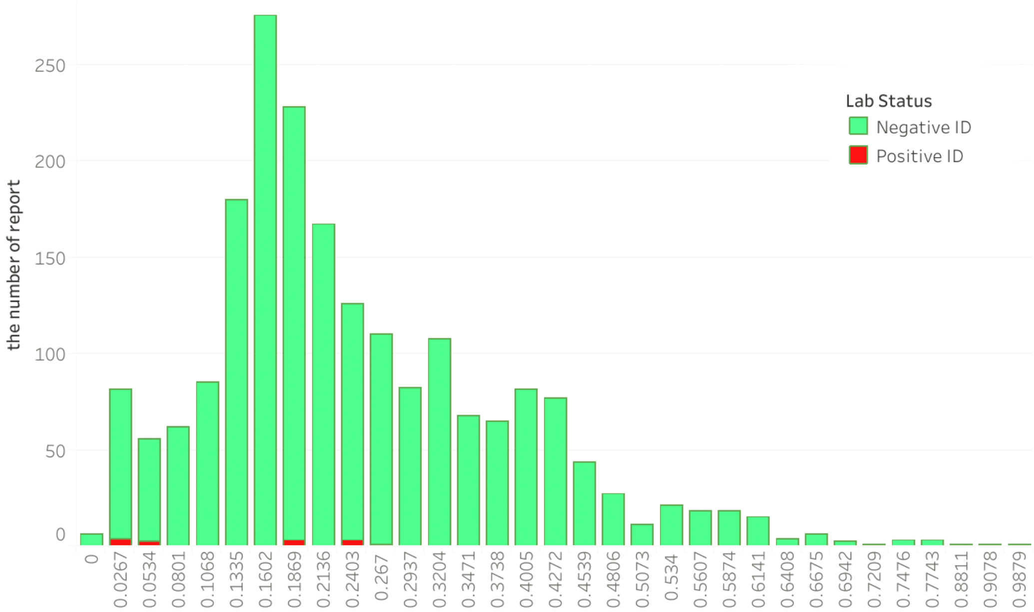


Figure 9: Positive and negative ID and their P_e

As shown in Figure 9, when P_e is close to 1, the positive ID occurrence is 0, and all positive ID have the P_e less than 25%, and 6 of them are less than 5%. That significantly show the robustness of the model.

3.3 Applicantion for prioritizing

We rank up the priority of every sent report based on theP_e from IE model. Report with higher P_e has a lower priority, for it is more likely to be a mistaken classification, while report with lower P_e has a higher priority.

In that way, the department staff can first identify the most probable presentation of hornets,which helps save a lot of resources. c.f. Figure 10

4 Further discussion

4.1 Update of the IE model

In order to help the state department to control the spread of hornets, our model needs to be updated according to the regularity of expansion and reproduction of hornets.

Regarding Regional mistake probability P_1, if the Washington government does not take any measures, over time, the hornet will spread in space as shown in Figure refÅs the number of nests increases, K -means++ algorithm needs more than previous two cluster. The cluster are increasing year by year. Then the distance from the sighting spot to nests will gradually decrease, and the probability of regional mistake will decrease a lot.

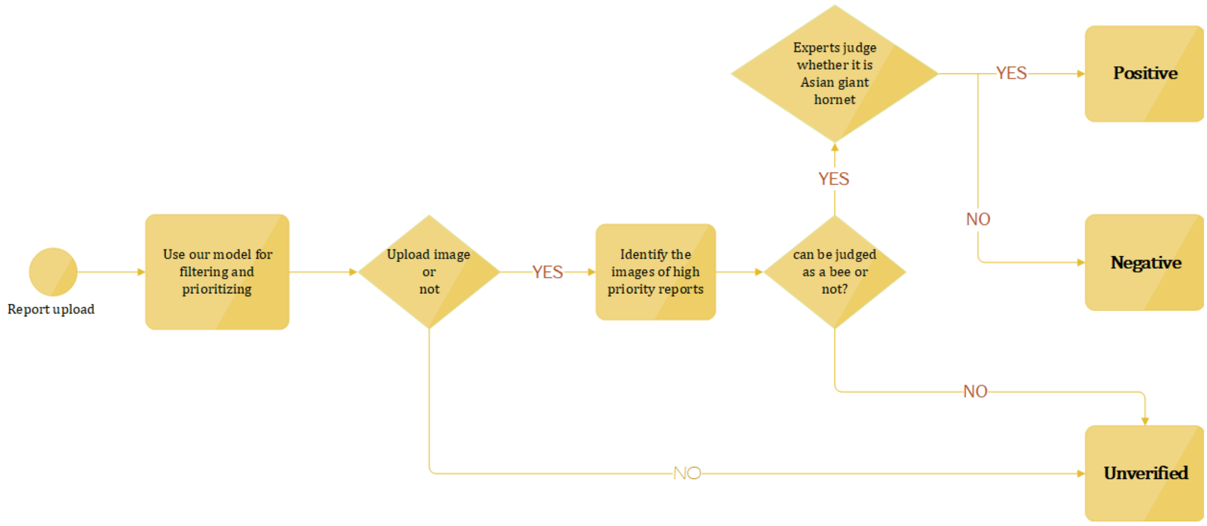


Figure 10: Flow chart for investigation of department staff

Regarding Seasonal mistake probability P_2, as the number of positive samples increases, we can more accurately predict the probability of hornet occurrence in each season. When the sample size is large enough, we can even predict the occurrence probability of occurrence by month or even by week.

As for Textual mistake probability P_3, with the development of technology and the growing knowledge of local people, the new habits or characteristics of the hornets might also change the probability of the occurrence of the hornets, so as the characteristics used for filtering.

Ways to update model:

1. Add new cluster for K-means ++ manually every half year.
2. Renew data set every quarter. If the new Positive ID reach 100, make a new fitting curve, using the new data.
3. Add new characteristics for text filtering, once new characteristics of hornets are found.

4.2 Evidence of eradication of hornets

In terms of region, the clustering of Asian Giant Hornet gradually reduces. Finally, clustering is not even needed. In the end, there was only one nest left, not even one, which means that the pest has been eradicated. In terms of time, because of the life of the Asian Giant Hornet worker lasts only a year. So, after using our model to filtrate the report and judge, if in one year, the number of 'positive' is 0, we can say that the pest has been eradicated in Washington State.

5 Analysis of our model

5.1 Sensitivity analysis

We chose one of the “positive”, which is the 1296th of all specimens. Its probability of mistaken P_e is 18.9%(after normalization),we take its value as the dependent variable, change the value of Regional weight, meanwhile Seasonal weight also changes. Then, we observe the change of the result.

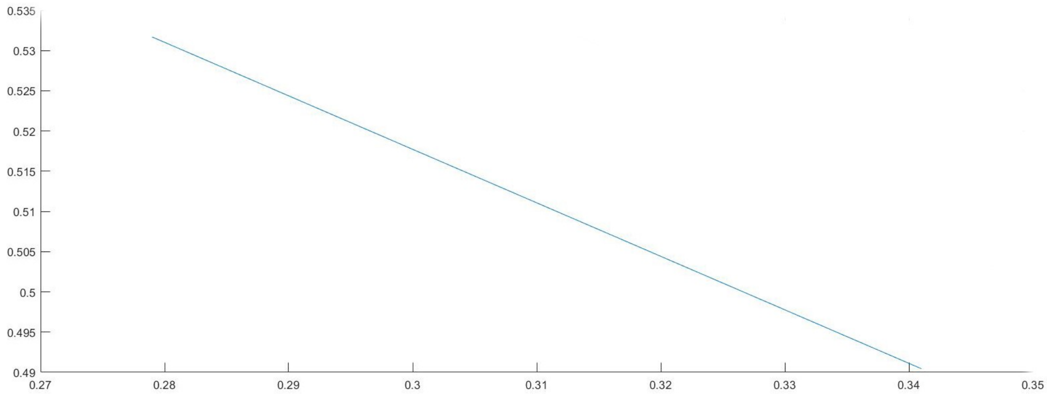


Figure 11: Sensitivity analysis

It can be seen from the image, the change of the weight has a huge impact on the results, and the influence is uniform. Therefore, in the actual evaluation, it need to be evaluated by many experts and perform the consistency test, take different values to find the most suitable weight.

5.2 Strengths and weaknesses

5.2.1 Strengths

1. A novel model founded on existing model and rigorous mathematical algorithm.

2. Capable and extraordinarily solve the practical problems. For example, Information Evaluating Model can lead to prioritizing investigation of the reports.

3. It is an evolvable model, whose accuracy can improve as the data set expand over time.

4. Models are established based on data and mathematical algorithms, quantitatively analyze problems, have mathematical foundations, and can be implemented by code.

5.2.2 Weaknesses

1. Weight coefficients of every types of mistake are hard to determine, for it requires professional biological knowledge and times of trials.

2. Actual situation of Washington and Asian giant hornet is much more complicated, for it can be influenced by many factors, such as human activity’ affect on climate and hornets’ reproduction.

6 Conclusion

After analysis, we know that the spread of Asian Giant Hornet can be predicted.In time, it is increasing. In space, it is expanding.

According to the given data, we create IE model and evaluated 2802 samples. Then, we can predict the likelihood of a mistaken classification.

Using our model, we can provide a priority to the professionals. The higher the priority, the more likely to be positive sightings. According to the habits and life cycle of the hornet, we need to add new cluster for K-means ++ manually every half year, renew data set every quarter. If the new Positive ID reach 100, make a new fitting curve, using the new data. And add new characteristics for text filtering, once new characteristics of hornets are found.

Using our model, if in one year, the number of positive is 0, we can say that the pest has been eradicated in Washington State.

At last, we report to the Washington State Department of Agriculture, summarize our model and explain How does our model work. How to prioritize these public reports with the limited resources of government agencies. And how to eradicate Asian Giant Hornet in Washington.

7 Memorandum

MEMO
TO:Washington State Department of Agriculture
FROM: Team#2106873
DATA: 8 February 2021
SUBJECT: Report Priority System

Dear Sir or Madame:

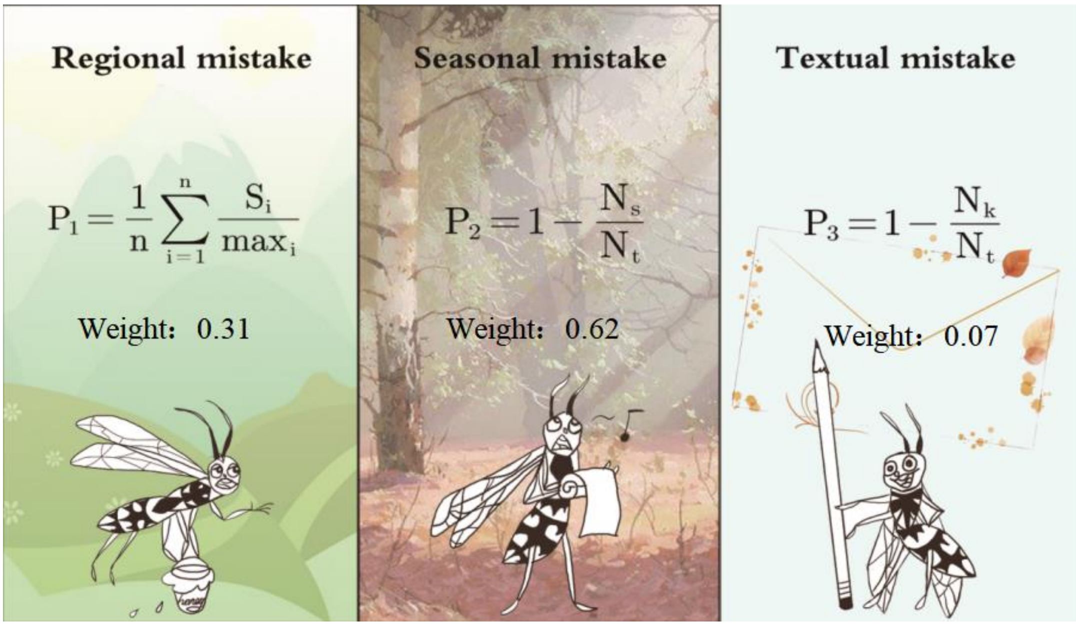
In order to prioritize limited resources to follow-up with additional investigation, we design one model, which is called IE model, that can be used for filtering the reports about Vespa mandarinia.

Now we are honored to write a memorandum to summarize our final investigative results in these 4 days and give you some suggestions.

Result

1. After data analysis and modeling, we find that the spread of this pest over time can be predicted. Over time, the number of pest is increasing. In space, its territory is expanding.

2. We create a model, which can predict the likelihood of a mistaken classification, the results demonstrate that the mistaken classification can be divided into three types: Regional mistake type, Seasonal mistake type and Textual mistake type.



3.According to the weight coefficients, we can find the probability of report to be mistaken. When the probability is close to 1, the positive cases are 0. All the positive cases are concentrated within the probability of less than 25%, among which 6 cases are concentrated within the probability of approximately 5%.

Conclusion

1. If we don't take any measures, the Asian Giant Hornet will sweep across the Washington state quickly. It will cause serious damage to the local ecosystem, especially to local bees. Their activities are seasonal. They are active in summer and autumn, inactive in spring and winter.

2. Our IE model takes full account of the characteristics of each specimen so the conclusions are very reliable. We can provide a priority to every new reports. In this way, the workload of the department staffs can be greatly reduced. Accuracy will also be greatly improved.

3. Because a new queen has a range of 30km for establishing her nest and the maximum foraging range of workers is 8 kilometers. Once we find a new nest or gathering point. The model needs to update. The method and frequency of updating have been given. Because of the life of the Asian Giant Hornet lasts only 1 year, we can judge, if in one year, the number of 'positive' is 0, we can think that the Asian Giant Hornet has been eradicated in Washington State. To sum up, by keeping an eye on the model update, our model can help you to find Asian Giant Hornet.

Suggestions

1. Before there are serious consequences, we need to find their nest as soon as possible and destroy it.

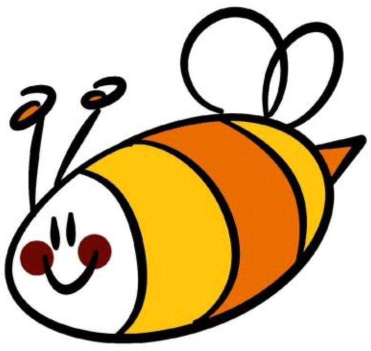
2. We suggest that the model should be updated regularly according to our article, so that the model could work with higher accuracy.

3. Asian giant hornets always build their nest under the ground, while other local bees build aerial nest. Their nest hard to find. If local people find their nest, you should reward them.

These are all results and suggestions our team provided to your department. Thank you for taking your time.

Hope that our models and these suggestions can help. Hope that Washington state can get rid of pest as soon as possible.

Your sincerely,
Team #2106873



References

[1] M. J, “Some methods for classification and analysis of multivariate observations.,” *Proceedings of the 5th Berkeley Symposium on Math. Statist. and Prob.*, vol. 1, 1967.

[2] R. A., “A-topsis – an approach based on topsis for ranking evolutionary algorithms,” *Procedia Computer Science*, vol. 55, pp. 308–317, 2015.

[3] C. J. Hutto and E. G. Vader, “A parsimonious rule-based model for sentiment analysis of social media text.,” *Eighth international AAAI conference on weblogs and social media*, 2014.

[4] U. S. D. o. A. Amber Tripodi, Trace Hardin, “Vespa mandarinia, asian giant hornet,” *New Pest Response Guidlines*, 2020.

Appendices

Appendix A

```
%this program is used to establish Information Evaluating Model
%write by matlab
Point1 = [48.77753,-122.41861];
Point2 = [48.98596,-122.65317];
pi=3.1415926;

for i=1:length(Latitude)
D1=distance('gc',Point1,[Latitude(i),Longitude(i)]);
dx1=D1*6371*2*pi/360;
D2=distance('gc',Point2,[Latitude(i),Longitude(i)]);
dx2=D2*6371*2*pi/360;
if(i==1)
dis1 = dx1;
dis2 = dx2;
else
dis1 = cat(1,dis1,dx1);
dis2 = cat(1,dis2,dx2);
end
end
max1= max(dis1);max2= max(dis2);
for i =1:length(Latitude)
```

```
t1 =  dis1(i)/max1;
t2 = dis2(i)/max2;
if(i==1)
p1 = t1;
p2 = t2;
else
p1 = cat(1,p1,t1);
p2 = cat(1,p2,t2);
end
end
p = (p1+p2)./2;
p_season = zeros(length(Latitude),1);
for i =1:length(Latitude)
if(DetectionDate(i)<=datetime(2019,3
,31))|| (DetectionDate(i)<=datetime(2020,3,31)
&&(DetectionDate(i)>=datetime(2020,1,1)))
p_season(i) = 0.021;
elseif (DetectionDate(i)<=datetime(2
019,6,30))||
(DetectionDate(i)<=datetime(2020,6,30)
&&(DetectionDate(i)>=datetime(2020,4,1)))
p_season(i) = 0.304;
elseif (DetectionDate(i)<=datetime(2019,
9,30))||
(DetectionDate(i)<=datetime(2020,9,30)
&&(DetectionDate(i)>=datetime(2020,7,1)))
p_season(i) = 0.434;
else
p_season(i) = 0.239;
end
end
shape = {'Large','big','huge', 'giant','largest'};
length= {'Inches' ,'feet','long','inch','size'};
color = {'Orange','blank', 'yellow','red',' brown',
    'bright'};
limbs = {'Singer','leg','stripes'};
behavior = {'Hit','stung','kill','killer'};
predation = {'Dead','decapitated','loss'};
nest ={'Nest','ground','holes'};
```

```
p_note = zeros(length(Latitude),1);
for i =1:length(Latitude)
words = strsplit(Note{i,1});
flag1 =0;flag2=0;flag3=0;flag4=0;flag5=0;flag6=0;flag7=0;
for j=1:length(words)
for h=1:length(shape)
if strcmpi(words{1,j},shape{1,h}) && flag1 == 0
p_note(i) = p_note(i)+(1/7);
flag1 = 1;
end
end
for h=1:length(longth)
if strcmpi(words{1,j},longth{1,h}) && flag2 == 0
p_note(i) = p_note(i)+(1/7);
flag2 = 1;
end
end
for h=1:length(color)
if strcmpi(words{1,j},color{1,h}) && flag3 == 0
p_note(i) = p_note(i)+(1/7);
flag3 = 1;
end
end
for h=1:length(limbs)
if strcmpi(words{1,j},limbs{1,h}) && flag4 == 0
p_note(i) = p_note(i)+(1/7);
flag4 = 1;
end
end
for h=1:length(behavior)
if strcmpi(words{1,j},behavior{1,h}) && flag5 == 0
p_note(i) = p_note(i)+(1/7);
flag5 = 1;
end
end
for h=1:length(predation)
if strcmpi(words{1,j},predation{1,h}) && flag6 == 0
p_note(i) = p_note(i)+(1/7);
flag6 = 1;
```

```
end
end
for h=1:length(nest)
if strcmpi(words{1,j},nest{1,h}) && flag7 == 0
p_note(i) = p_note(i)+(1/7);
flag7 = 1;
end
end
end
end
```

```
# This program is used to count the number of words in the sentence
#write by python
import io
import re

class Counter:
def __init__(self, path):
"""
:param path:
"""
self.mapping = dict()
with io.open(path, encoding="utf-8") as f:
data = f.read()
words = [s.lower() for s in re.findall("\w+", data)]
for word in words:
self.mapping[word] = self.mapping.get(word, 0) + 1

def most_common(self, n):
assert n > 0, "n should be large than 0"
return sorted(self.mapping.items(), key=lambda item:
item[1], reverse=True)[:n]

if __name__ == '__main__':
most_common_5 = Counter("test.txt").most_common(600)
for item in most_common_5:
print(item)
```

Appendix B

	A	B	C
1	GlobalID	Notes	Lab Status
2	{BBBA5BA	I was cleaning my gutters when I heard a snapping noise coming from the wall of my house. It was sitting on a dirt clump that looked to have been starting a nest. I backed away, and waited till evening to spray hornet killer on it. Next day knocked	Unverified
3	{7F3B6DB	Insects observed late Sept or early Oct	Positive ID
4	{EB4746A1	We found about a dozen of these bees in our backfield last October while we were storing bee boxes, Sorry no pictures	Unverified
5	{5EAD336	Hornet specimen sent to WSU	Positive ID
6	{F1864CC		Positive ID
7	{5AC8034	One dead wasp seen in Blaine, and suspect flying nearby	Positive ID
8	{13B67BC	Massive loss of bees, decapitated. No hornet specimens were located.	Unverified
9	{8F4C0AA	Saw a bee that looked like the picture in the Bellingham Herald. I did not notice if it had a skinny waist but it had pronounced striping on the bottom half and a vivid yellow at the head. It was at least as large as bumble bees, likely larger.r.	Unverified
10	{978317A4	Approx 2" long was yellow and black. Did not get a good look at its head. Was much larger than the yellowjacket queens that are out and about. It perched near a yellowjacket trap we have out then bounced off the living room window a couple of	Unverified
11	{055198DE	I heard a very loud buzzing outside my front door. I finally spotted it about 7 -9 feet away from me investigating my lawn. I watched it for about 5 minutes. It had a bright yellow head and tail (with stripes). It as at least 2 inches long (huge) and rea	Unverified
12	{49DC3BF	It was on the wet, muddy ground where the street sweeper was cleaned out at Blaine Public Works Dept, observed by me and two other employees. We covered it with a box but couldn't capture it before it flew away. It was had the orange face, pl	Unverified
13	{893D73D	the second time I saw one here -- we live next to Whatcom Falls Park. It was moving fast and making noise but since boyhood I have been very interested in insects. I even considered a career in entomology. It was a hymnoptera 2" long, and I d	Unverified
14	{C61FBDE	I was leaving a friends house and my drivers side window was open. I stopped at stop sign and noticed a large insect buzzing against my windshield. It was very large >1" hornet with distinctive yellow collared face of Asian Hornet. I quickly rolled up	Unverified
15	{1C6D0EA	Specimen collected by citizen scientist and reported to provincial government.	Positive ID
16	{0D607BE	I'll try to get a picture. It was about 2" long and it's body was round (fat). It was orange and black. Below the thorax, it was orange and black, alternating stripes. It was flying alone, looking for a spot to nest, I suspect.	Unverified
17	{AD56E8D		Positive ID
18	{FA174F57	I was watering my garden and heard the loudest buzzing. It wasn't like a hummingbird's wings, but like an insect, but just as loud as a hummingbird. I heard it coming toward me and I ducked, when I looked up a HUGE (2+ in) insect flew above me.	Unverified
19	{0D87B50	Did not see the hornet. I found several decapitated bees on my deck	Unverified
20	{D0270C3	The hornet was flying in a northerly direction toward our raspberry patch where there are a large number of bees. The hornet was flying much faster than other insects normally fly. There was a distinct loud buzzing sound.	Unverified
21	{FC6E894B		Positive ID
22	{13EBA05	Flew in an open window at our farm house. Dakota Creek runs through our property. There is a honey bee farmer in the field across the street. The obvious black and orange stripes were unmistakable. It was definable over an inch long. Its buzz was	Unverified
23	{A90D48A	While kayaking in Drayton Harbor, Blaine, I saw what I thought to be an Asian hornet struggling in the water. It was at least 2 inches long with yellow/orange & dark brown (black?) stripes and an orangey face. Unfortunately, I was unable to get a pi	Unverified
24	{4C97BC8	Large size. Looked like mini hummingbird. Had yellow and black bottom. Large tan or light brown head.	Unverified
25	{21870CE5	Large hot yet shape 2" long orange and black.	Unverified
26	{94EC5C0	1.5-2" in length. Landed on the rear tire of a fellow biker in Galbraith Bike Area. Significantly larger than any hornet I've seen before. Long dark body, orange head. Stinger looked like it was protruding. Didn't make to sting/no body curling/didn't see	Unverified
27	{882EC093	Next time in Blaine, there is a small hornet nest down by the public boat launch site in Blaine Harbor, I'll try and attach pics from my phone I took this am, the nest is a foot off the ground, just south and under the trees across from the bathrooms a	Unverified
28	{80FA4F9F	Was unable to catch it alive. It was healthy, almost two inches. Landed on a log in front of me, by the entrance to the park. Have lived in the Far East so it was a familiar sight to me, and I wasn't afraid.	Unverified
29	{A717D86	Spotted at outdoor dining in Birch Bay, posted to WSDA Facebook site	Positive ID
30	{FBD8524F	I seen the wasp on pavement went closer to get a look confirmed it wasn't a yellow jacket or other species of wasp went to grab my phone to get picture and a passing motorcycle made it fly off. It had black orange face and same color on lower bo	Unverified
31	{D423B68	I saw a large flying insect that was at least 3 times bigger than a wasp. Mainly black with orange stripes. I did not have my phone to take a picture. But it flew back and forth across our lower deck and then did the same thing on our upper deck. Nev	Unverified
32	{23000D4	It came and went too fast for a photo, but it was large with an orangey head, I'm almost positive it was the Asian giant hornet.	Unverified
33	{20B24FE1	It landed on our outside umbrella, about a foot away from where we were sitting. It was enormous...at least 2 inches long. It flew away before I could capture a photo	Unverified
34	{17BCC54	Was not able to take a picture of it because I ran away, but it was exactly as this picture. It was huge like 2 inches long, and orange looking with long stinger	Unverified
35	{2138197A	We caught and killed 2, 2nd on September 25th	Positive ID
36	{AA461F47	WSDA submitted for citizen report	Positive ID
37	{DEF5D82	Live hornet captured by WSDA staff	Positive ID
38	{44337587	I was outside with my neighbor and one landed on her back. We both saw it. It was huge and human like eyes. Golden is wings. Very vibrant color stripe. We took the kids to safety and I went back out and saw another in our field. You are welcome	Unverified
39	{A433310	Large bee about 2-2.5in long. Large abdomen. Dark orange markings. Unable to take photo	Unverified
40	{0FAC3767	doorbell cam image	Positive ID
41	{BEAC832	Dead hornet in light	Positive ID
42	{2D515E6	very large hornet hovering by my home's gutter. Largest I have seen, clearly not a yellow jacket	Unverified
43	{0437C2C	I'm almost positive that it was based on the size and what I've seen images of. It was an orange/black color and very fat. It was outside my window at work on Bakerview rd. I was unable to get a pic but will try again.	Unverified
44	{32B69B76	Looked like a bee but way way bigger. Looked like the one shown to me by the research team. Appeared to be the male Asian hornet	Unverified
45	{4E587C5E	Found a pile of wasps with their heads missing outside Bellingham mall.	Unverified
46	{A0909E9E	Saw only a side view about 8 inches from ear. Loud buzzing, about 1.75 in. long, fat. Startles me and it flew away.	Unverified
47	{9797D62	It was about 2 inches long and appeared to be looking for shelter and lost sight of it when it went into a gap in my house siding	Unverified