An Auxiliary Identification System for Destructive Pests

Weijian Mai¹, Fengjie Wu²*, Jiangtao Chen¹, Yisheng Tang¹, and Qingkun Liu¹

¹ School of Electronics and Communications Engineering, Guangzhou University, Guangzhou, Guangdong, 510006, China
² Laboratory Center of Guangzhou University, Guangzhou University, Guangzhou, Guangdong, 510006, China
*Corresponding author’s e-mail: wufengjie501@gzhu.edu.cn

Abstract. This work takes this Asian Hornet's invasion of Washington in 2020 as an example to build an auxiliary identification system for Destructive Pests to help the government's rescue work. To help the government to prioritize the resource to the most possible reports and improve the investigation efficiency, we built a binary classification model to classify the report that contains images and text. For image classification, we design a VGG16-based convolutional neural network with ImageNet pretraining and employ it to this specific task by transfer learning, which achieves superior performance with a high F1 score (95.08%) on our dataset. For text classification, we use the Latent Dirichlet Allocation (LDA) model to cluster words into different themes and adopt cosine similarity to calculate the similarity between the new report and the benchmark positive/negative sets. The results demonstrated that our approach can identify target pests efficiently and help the government to prioritize the resource to the most possible reports, improving the investigation efficiency.

1. Introduction

Currently, some governments have created helplines for the citizens to report sightings of destructive pests. However, these pests have similar appearance characteristics to other insects, which is confusing to the eyewitness. It is time-consuming and labor-intensive to manually distinguish whether it is the target pest from the pictures and text description information provided by the witnesses.

Specifically, due to the potentially severe damage of the Asian Giant Hornet, the presence of the Vespa mandarinia causes great anxiety to the local citizens. The government creates helplines for the citizens to report sightings of these hornets. Given the limited resources that the government have, we aim to help the government to prioritize the resource to the most possible reports and improve the investigation efficiency. Specifically, we designed a binary classification model to classify the target pest from reports that contain images and text descriptions. For image classification, we present a VGG16-based convolutional neural network using ImageNet for pre-training and employ it to this specific task by transfer learning, which achieves a superior performance of 95.08% in the F1 score. For text classification, an LDA model is proposed to extract discriminative information and cluster it into various themes. Based on this basis, this work builds the positive and negative text sets that contain different themes. When a new report is added, the sum of the cosine similarity of each theme is calculated between the report and each of the text sets. The greater the similarity, the higher the probability of belonging to this category. The classification results demonstrated that our approach can identify target pests efficiently and help the government to prioritize the resource to the most possible reports, which improves the investigation efficiency. Furthermore, we also discuss the update
frequency of our model so that the model could learn more information over time and achieve more robust performance.

2. Data Pre-processing
In this problem, we are given three datasets: a spreadsheet with 4440 reports of sightings, a rare file that contains the images submitted in the reports, and a spreadsheet mapping the image to the reports of sightings. The datasets include a unique label, the detection date, and submission date, the comments from both the reporter and the lab, the location, and the official classification for each sighting record.

Before creating our model, we pre-process our data by filtering, cleaning, and visualizing the data set, which allows us to obtain more reliable pieces of data and get a more intuitive interpretation of it.

1. Data split: we divide data into image and text sets according to the data type, and process them separately.
2. Statistics of data: In the dataset, there is a total of 4440 reports, in which there are 14 reports of Positive ID, 2069 reports of Negative ID, 2342 reports of Unverified, and 15 reports of Unprocessed.

3. Methodology

3.1. Overview
Figure 1 shows the overall flow chart of our model.

![Overall Flow Chart](image)

3.2. Report Image Classification based on CNN
According to the statistics of the data set, a majority of reports contain a photo of the insects, given the data type and the amount, we used the VGG16 CNN network to recognize the image in the report to classify the insect. If the classification result is not the Asian Giant Hornets, we will classify this report to Negative ID.

3.2.1. Image Data Set Construction
Due to the large difference in the amount of data between the positive and negative classes of this task, we need to expand the data for the positive class. Data augmentation skills including flipping, rotating, and adding different levels of Gaussian noise, were used to expand the dataset. Due to the lack of standard Asian giant hornet pictures in the data set, we also obtained some clear pictures from other materials and added them to it. Finally, we obtained a training dataset with 2970 negative and 150 positive images, a validation dataset with 470 negative and 30 positive images, and a test set with 470 negative and 30 positive images.
3.2.2. Model Implementation

We use the TensorFlow framework [2] to build our VGG16 [3] model and transfer learning to obtain the initial parameter from the VGG16 model trained on the Imagenet. Therefore, we can utilize the capability to extract features from the image in the recognition of the Asian Giant Hornets. In this model, we utilize Adam [4] and cross-entropy as our optimizer and loss function, respectively. The cross-entropy’s expression is as Equation (1).

\[
H(p, q) = -\sum_{i=1}^{n} p(x_i) \log(q(x_i))
\]  

In Equation (1), when \(x_i\) represents the input report, \(p(x_i)\) represents the real probability of \(x_i\), where we set \(x_i = 1\), and \(q(x_i)\) represent the prediction of the network of the input \(x_i\).

3.2.3. Model Training

In the optimizer, we set the parameter \(l_p = 0.001\), \(batachsize = 20\), and train our network with the labeled data. In the training process, we iterate the training process 30 times and evaluate the accuracy using the cross-validation set. Dropout [5] with a probability of 0.2 is employed to the fully connected layer when training our model. The training loss and validation loss over training epochs are shown in Figure 2.

![Figure 2. Training loss and Validation loss over training epochs.](image)

3.2.4. Model Evaluation

After we train the program, we use the test set as the input of our model and calculate the F1 score of the prediction to evaluate the prediction accuracy of our binary classification model. Figure 3 and TABLE 1 show the evaluation result of our model.

![Figure 3. Precision and Recall.](image)

|        | RESULTS       |
|--------|---------------|
| Precision | 1.0000       |
| Recall   | 0.9080        |
| F1 score | 0.9508        |

TABLE 1.
3.3. LDA Model
The received report contains notes from both the reporter and the laboratory, which provides useful information to aid the prediction. In this section, we use the Latent Dirichlet Allocation (LDA) model [6] to process the notes and find the keywords related to Positive and Negative ID, and we also use this model to try to divide these words into five related topics such as the institution, location, season, appearance, and quantity.

3.3.1. Text Pre-processing
Before the text recognition, we need to correct the spelling mistakes, unify the forms of each word, and convert them to the lower case in the comments.

3.3.2. Steps of LDA Model
The LDA model mainly includes three layers, the words, the topics, and the documents. Based on the LDA model, we excavate the topic in the notes and analyze the most frequent words.

| TABLE 2. LDA symbol explanation |
|----------------------------------|
| $\alpha, \beta$ | Prior parameters of the Dirichlet function |
| $\theta, \phi$ | The parameter of the topic’s multiple distribution in the document |

The LDA model is based on the theory that every document chose a topic in a certain probability, then chose a word in a certain topic, and after repetition of the process, the document is generated. Both from the document to the topic and from the topic to the words obey the multinomial distribution shown as Equation (2) and Equation (3), respectively:

$$T|\theta = \text{Multinomial}(\theta)$$  \hspace{1cm} (2)
$$T|W = \text{Multinomial}(\theta)$$  \hspace{1cm} (3)

Given certain document $d_j$, the probability of the appearance of the word $w_i$ is denoted as Equation (4):

$$P(W_i|D_j) = \sum_{i=1}^{K} P(W_i|T = s) \times P(T = s|D_j)$$  \hspace{1cm} (4)

In Equation (4), $P(W_i|T = s)$ represents the probability of the appearance of $i$ words $w_i$ given certain topic $s$, $P(T = s|D_j)$ represents the probability of the topic is $s$ given the appearance of the $j$th documents.

| TABLE 3. POSITIVE TOPICS |
|--------------------------|
| Theme1       | Theme2      | Theme3       |
| wsda         | nanaimo     | september    |
| government   | blaine      | october      |

| TABLE 4. NEGATIVE TOPICS |
|--------------------------|
| Theme4       | Theme5       |
| one          | look         |
| a            | huge         |
From the LDA analysis, we have extracted the keywords from the text information and group them into positive and negative topics (TABLE 3 and TABLE 4), from which we visualize the word clouds (Figure 4 and Figure 5). Furthermore, we categorize them into five related topics such as organization, region, quarter, appearance, and quantity (Figure 6). These topics reflect critical factors to evaluate the report, and we will discuss a text classification model in the following thesis.

3.3.3. Text Classification Model

(1) Create Text Classification Data Set

First, we obtain five keywords from the Positive ID and Negative ID and add them to the text set in LDA. We convert the data into words and obtain the city’s name from its latitude and longitude so that we can process them more precisely, for example, we convert 2019/9/29 to “September” and obtain “Nanaimo” from the longitude and latitude [49.149394, -123.943134].

The result shows dramatic differences between the positive set and negative set: From the aspect of month of the report, the word “September” should be classified into the positive set and “December” to negative set, which is also in accordance with the insect’s habit that its population reaches its peak in autumn, and demise or hibernation in winter. From the aspect of the location of the report, the location that has Positive ID should be classified into the positive set, because there may be a nest near the discovery spot.

(2) Classify the New Report

1) Word segmentation: By implementing the jieba library, we segment the input notes into several parts, obtain the month’s word and the city’s name, and put these words into a text set.

2) Key words extraction: Based on the LDA model, we extract the keywords from the notes and group them into five related topics and calculate the cosine similarity. Cosine similarity measures the angle between two vectors in the vector space, and the expression is as Equation (5):
\[
\text{Similarity} = \sum_{i=1}^{5} \cos (\theta_i) = \sum_{i=1}^{5} \frac{A_i + B_i}{\|A_i\| + \|B_i\|} = \sum_{i=1}^{5} \frac{\sum_{i=1}^{n} A_{ij} \times B_{ij}}{\sqrt{\sum_{i=1}^{n} (A_{ij})^2} \times \sqrt{\sum_{i=1}^{n} (B_{ij})^2}}
\] (5)

In the Equation (5), \(A, B\) is the eigenvector of two texts, and the \(i^{th}\) topic’s eigenvector is \(A_i = (A_{i1}, A_{i2}, ... , A_{in})\), and \(B_i = (B_{i1}, B_{i2}, ... , B_{in})\), respectively. The \(\cos(\theta_i)\) is the cosine of the angle between vector \(A\) and \(B\), which represents the similarity between \(A\) and \(B\) in \(i^{th}\) topics. The range of \(\cos(\theta_i)\) is [-1, 1]. The greater value means the two vectors are closer to each other in direction. The cosine similarity between two texts is the sum of all topics’ cosine similarity.

We calculate the similarity to both positive set and negative set and obtain \(\text{Similarity}_{\text{pos}}\) and \(\text{Similarity}_{\text{neg}}\), respectively, and give the prediction based on Equation (6).

\[
\begin{align*}
\text{Similarity}_{\text{pos}} < \text{Similarity}_{\text{neg}}, \text{Positive} \\
\text{Similarity}_{\text{pos}} > \text{Similarity}_{\text{neg}}, \text{Negative}
\end{align*}
\] (6)

Where positive means the report location has the Asian giant hornet and negative means the opposite.

3.4 The Combination of Image Classification and LDA Model

In our model, we give the Image Classification a higher weight than the LDA model. When the report contains a standard jpeg and png image, we will use our Image Classification model first since the VGG16 Net is more powerful. We defined a \textit{True positive} ranging from [0, 1] to represent the possibility of the positive set. The output of VGG16 is a Softmax classifier, and given \(p\) denotes the probability image is positive, the output is as Equation (7), and the classification result is as Equation (8):

\[
[ \text{positive} = p, \text{negative} = 1-p ]
\] (7)

\[
\begin{align*}
\text{For all images classified as positive, we sort them in descending of } p: \\
\text{St. } p \rightarrow 1, \text{True positive} \rightarrow 1
\end{align*}
\] (8)

When the report does not contain the image, we use LDA to analyze the notes and calculate the cosine similarity to classify the report. It is essential to note that the output is obtained from the difference in similarity between the positive text set, negative text set, and the notes in the report. The definition of the difference is as Equation (10):

\[
\delta_{\text{pos-neg}} = \text{Similarity}_{\text{pos}} - \text{Similarity}_{\text{neg}}
\] (10)

When \(\delta_{\text{pos-neg}} > 0\), the output is positive, and the greater the value, the \textit{True positive} is clustered to 1. We will abandon the reports classified as negative.

4. Discussions About The Update Frequency of Our Model

Our Image Classification model based on VGG16 achieves 99% accuracy in the F1 score, and we are confident with the model. The model should be functional unless there is a significant appearance change of the Asian Giant Hornets. Given the validity of our Image Classification model over time, we recommend adding new images from the newly found Positive ID reports and retrain the model annually.

Our LDA model mainly relies on the positive and negative text set, and it calculates the cosine similarity between the new report and the text set to generate the output. For this reason, we need to regularly update the words in the text set. Since the quantity of the Negative ID is relatively big and the new report does not significantly affect the result, we recommend updating the negative text set quarterly. However, the positive test set is small, we hope to keep it up to date. Thus, we recommend updating the positive text set once there is a new Positive ID report confirmed. This update serves a significant purpose because if a Positive ID report arises in an area, there is more likely to have more
Positive ID report in the neighbor area. Frequently update the model will help increase the sensitivity of the model is particularly risky areas.

5. Conclusions
This work takes the Asian Hornet invasion as an example to demonstrate the effectiveness of our model for government rescue work when limited information is available. Since the reports received from witnesses contain only image and text information, we constructed a unified model that includes image and text classification to identify the probability of the occurrence of target pests in the reported information. By ranking the probability, the government can effectively allocate resources to rescue the areas with a high probability of pest occurrence as soon as possible.

Acknowledgments
This work was supported in part by the Climbing Program of Guangdong Province, China, under Grant pdjh2020b0462, and in part by the National College Student Innovation and Entrepreneurship Training Program, China, under Grant S202011078109.

References
[1] Washington State Department of Agriculture website: https://agr.wa.gov/departments/fests-pests-and-weeds/fests/hornets/data.
[2] Abadi, Martin, et al. "Tensorflow: Large-scale machine learning on heterogeneous distributed systems." arXiv preprint arXiv:1603.04467 (2016).
[3] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).
[4] Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 (2014).
[5] Hinton, Geoffrey E., et al. "Improving neural networks by preventing co-adaptation of feature detectors." arXiv preprint arXiv:1207.0580 (2012).
[6] Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent Dirichlet allocation." the Journal of Machine Learning Research 3 (2003): 993-1022.
[7] K. G. Kim, “Book Review: Deep Learning,” Healthc Inform Res, vol. 22, no. 4, p. 351, 2016, doi: 10.4258/hir.2016.22.4.351.