
A System of Interstate Energy Cooperation Goals Based on Data Insight

Summary

To address the above-mentioned challenge, we propose.... Then a XX model is proposed as We propose ... based on Then we employ an XX model as The results show Lastly, we analyze ... to find

According to our empirical results, we propose some confident (sales) strategies and recommendations for We write a letter to (the marketing director of Sunshine Company) to summarize our analysis and results, together with our recommendations./ to assist with ..., we propose ...

Our framework shows a strong accuracy, robustness. It can be easily implemented to other data with our source codes.

Keywords: ext-Based Measure, Informative Text Selection, Reputation Quantification, Sales Strategy Formation

Contents

1	Introduction	1
1.1	Problem Background	1
1.1.1	Asian Giant Hornet Invading	1
1.1.2	Asian Giant Hornet	1
1.2	Our work	2
2	Assumptions and Notations	3
2.1	Assumptions	3
2.2	Notations	3
3	Problem 1:The Spread of Hornet	4
3.1	Data Preprocessing	4
3.2	Hypothesis Testing	5
3.2.1	Model	5
3.2.2	Conclusion	6
4	Problem 2:Classification Model	6
4.1	Model Introduction	6
4.2	Date,Longitude and Latitude Coding	7
4.3	Video Processing	7
4.3.1	Scene Detection	7
4.3.2	Video Keyframes	8
4.4	Image Processing	8
4.4.1	Sample Box Annotation	8
4.4.2	Data Augmentation	9
4.5	Image Characteristics Extraction	10
4.5.1	Focal Loss	10
4.6	Language Processing	10
5	Targets and Action	11
6	Sensitivity Analysis	11
7	Strengths and Weaknesses	11
7.1	Strengths	11
7.2	Weaknesses	11
8	conclusion	11
Memo		12
References		13
Appendices		14

1 Introduction

1.1 Problem Background

1.1.1 Asian Giant Hornet Invading

Vespa mandarinia is the largest species of hornet in the world, and the occurrence of the nest was alarming. Additionally, the giant hornet is a predator of European honeybees, invading and destroying their nests. A small number of the hornets are capable of destroying a whole colony of European honeybees in a short time. At the same time, they are voracious predators of other insects that are considered agricultural pests.

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.

From the first report of Asian giant hornets received, the public has played a critical role in the detection of Asian giant hornets in the Pacific Northwest. Washington State Department of Agriculture(WSDA) has set up platforms[1] to collect informations from the public and set traps to get hornets.

1.1.2 Asian Giant Hornet

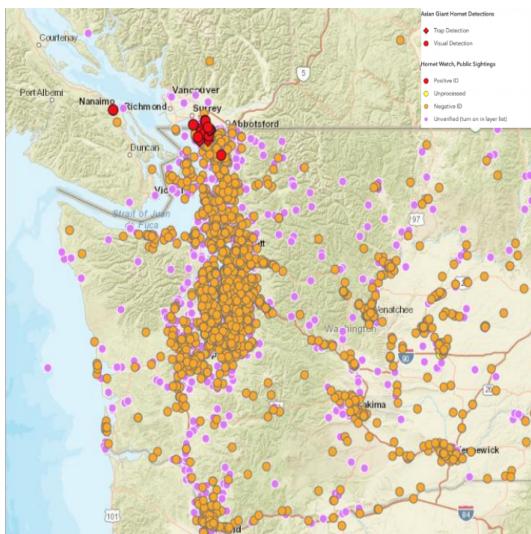


Figure 1: Asian Giant Hornet Detections[2]



Figure 2: *Vespa mandarinia*[3]

Name

- Common name: Asian giant hornet, sparrow wasp, murder hornet
- Scientific name: *Vespa mandarinia* Smith, 1852

Range of Activity

Usually to find food, the Asian giant hornet goes only 0.5 to 1.25 miles (1-2 kilometers) away from the nest (and never more than 5 miles (8 kilometers)).

Table 1: Similar Species

Species	Similarity	Dinstinguish
Asian Giant Hornet	around 1.5 inches,yellow head,black belly,live on the ground or lower than 6 foots	
VespaCabro	size,shape,color are simlilar	small differences on face and chest,live higher than 6 foots
VespuiaSquamosa	only the queen may be confused in spring	much smaller in size
cicada killers	simliar size	no yellow in head and belly
Dolichovespula Maculata		small size,in the color of black and white,live on the tree or reef

Life Period

Like other social wasps, Asian giant hornets are an annual species that build new nests every year. When winter arrives, the current seasons' nests die out and the only individuals that survive are overwintering queens. When overwintering queens emerge in the spring, they seek out protected areas in the ground to begin building a nests.

Comparision with Other Species

While Asian giant hornets do not occur in eastern North America, there are a number of other large wasps that may be confused for them. Here introduce some species alike and there distinctions.

1.2 Our work

- Firstly,we try to predict the spread of hornet ,but just find that the spread is not regulated.
- Secondly,from all the information WSDA recievied ,several confirmed sightings of the pest have occurred in neighboring Washington State, as well as a multitude of mistaken sightings. So we construct a model to distinguish whether or not the observed object in one sighting record is Asian Giant Hornet.
- Then,we use this model into unverified records to figure out the identity of the object.
- And we find methods to update the model as new data comes in.
- Finally, we use the result to verify that the hornet in Washington State has been readicated.

2 Assumptions and Notations

2.1 Assumptions

To simplify our model and eliminate the complexity, we make the following main assumptions in this literature. All assumptions will be re-emphasized once they are used in the construction of our model:

1. e assume that the data given by official can fully solve the problem of the topic and do not need to obtain additional data as auxiliary information.
2. Assumption2.
3. Assumption3.

2.2 Notations

In this work, we use the nomenclature in **Table 2** in the model construction. Other nonfrequent-used symbols will be introduced once they are used.

1. work1.
2. work2.
3. work3.
4. work4.

Table 2: Notations

Symbol	Definition
I	The Intensity of the Heat Source
k	The Material's Conductivity
h	The Heat Transfer Coefficient
c	The Heat Capacity of Water ($=4200 \text{ J}/(\text{kg}\cdot\text{ }^\circ\text{C})$)
ρ	The Density of Water ($=10^3 \text{ kg}/\text{m}^3$)

3 Problem 1:The Spread of Hornet

As we want to figure out the regulation of the spread of hornets, we divide the process into two parts: the first is **Data preprocessing**. Firstly, we draw related pictures and tables to get more intuition. Then we process data for further analysis. The second is to test and try to find out the regularity of the spread of hornet by **Hypothesis testing theory**.

3.1 Data Preprocessing

Firstly we want to know about the basic information of these sighting records

Table 3: Basic Information of records

Positive number	14
Negative number	2096
Unverified number	2342
Unprocessed	15
Date range	2018-06-19 to 2020-08-01
Latitude range	45.4886 – 49.5480
Longitude range	-124.6650 – 116.8736

According to the table above, we decide to choose process data in these way:

- Select positive ID and unverified sample to analysis. That's because the sample of Positive ID is too small and random.
- Use the average position of each month's sighting site as the center of hornet, and search the trend of the species by the change of center point.
- Use the data after March, 2019. For it's when the first confirmed sighting appeared.

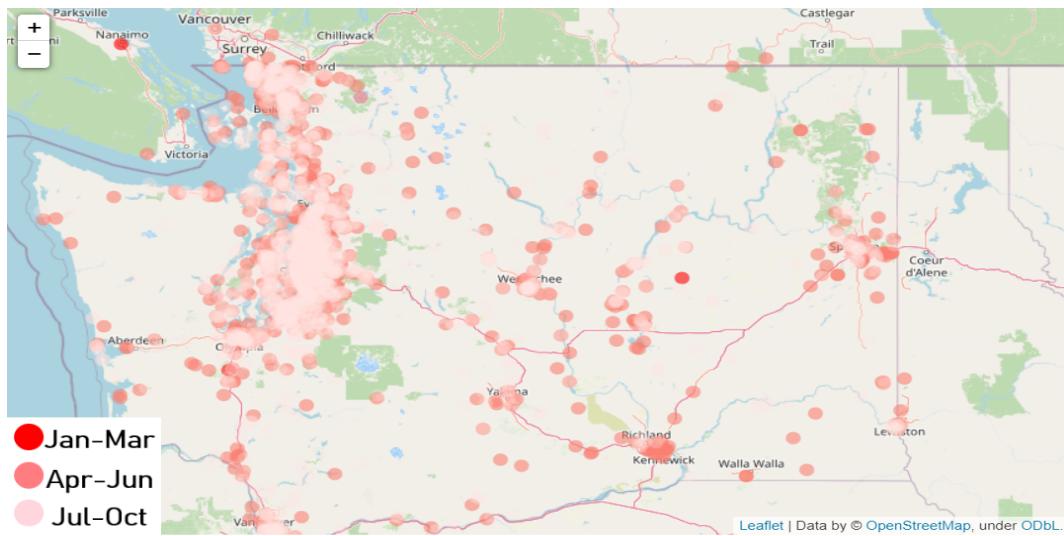


Figure 3: sighting sites

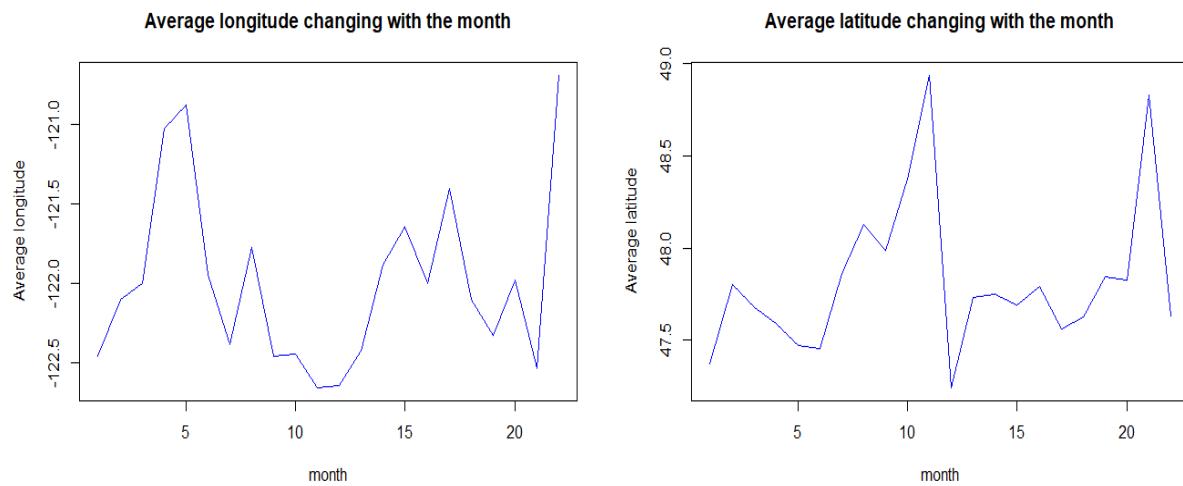


Figure 4: The change of center point

From the pictures above ,we find out that the movement of hornet is quite random,so we need to use some statistic methods to approve that.

3.2 Hypothesis Testing

3.2.1 Model

We want to test whether the center points of hornet are related by time or the route of the colony has some regulation. So we hypothesis that:

H_0 : These center points are independent $\leftrightarrow H_1$: These center points are not independent

And for it's two dimension data, we test the regulation in longitude and latitude apparently.

As we use ρ_i to represent the k-order lag correlation coefficient of the sample's longitude. The hypothesis can be written as:

$$H0 : \rho_1^2 = \rho_2^2 = \dots \rho_h^2 \leftrightarrow H1 : \exists k, s.t. \rho_k^2 \neq 0$$

And choose pivot quantity

$$Q = n(n+1) \sum_{k=1}^h \frac{\rho_k^2}{n-k}, \quad , h : \text{degree of randomness}$$

For $Q \sim \chi^2(h)$, as significance level is α , the rejection region is $Q > \chi_{\alpha-1}^2(h)$.

In the same way, we test the randomness of the latitude data by using latitude information to calculate $\bar{\rho}_i$ and \bar{Q} .

3.2.2 Conclusion

According to calculating result, neither Q and \bar{Q} doesn't fall in the the rejection region even in the significance level of 0.5 (P-value are above 0.5 in whatever the randomness is). **That means we can't say the spread of hornet can be predicted even the accuracy is 0.5**

4 Problem 2: Classification Model

4.1 Model Introduction

From all the records we get information in quite a lot of forms including **word, image, video and so on.**

- **Video:** video of the recorded object, including video after capture the insect, surveillance video contains the insect, video of outdoor observation and so on.
- **Image:** image with the recorded insect held on hand, captured by bottle, on a web, in flowers and so on.
- **Word:** notes recorded by the observer; comments replied by the officer
- **Other information:** longitude and latitude of the sighting site and the date.
- **Label Status:** whether or not the object is Asian Giant Hornet, including unverified.

So, we choose **Multimodal machine learning** to classify the records. That means we use all the given information to construct the classification model.

Another key problem of this issue is: the positive sample is too small (14 compared with the negative sample number: 2096). So the **sample imbalance** is great. We use **sample augmentation, focal loss, oversampling** to deal with this problem.

A more detailed explanation of the model can be seen in the picture below:

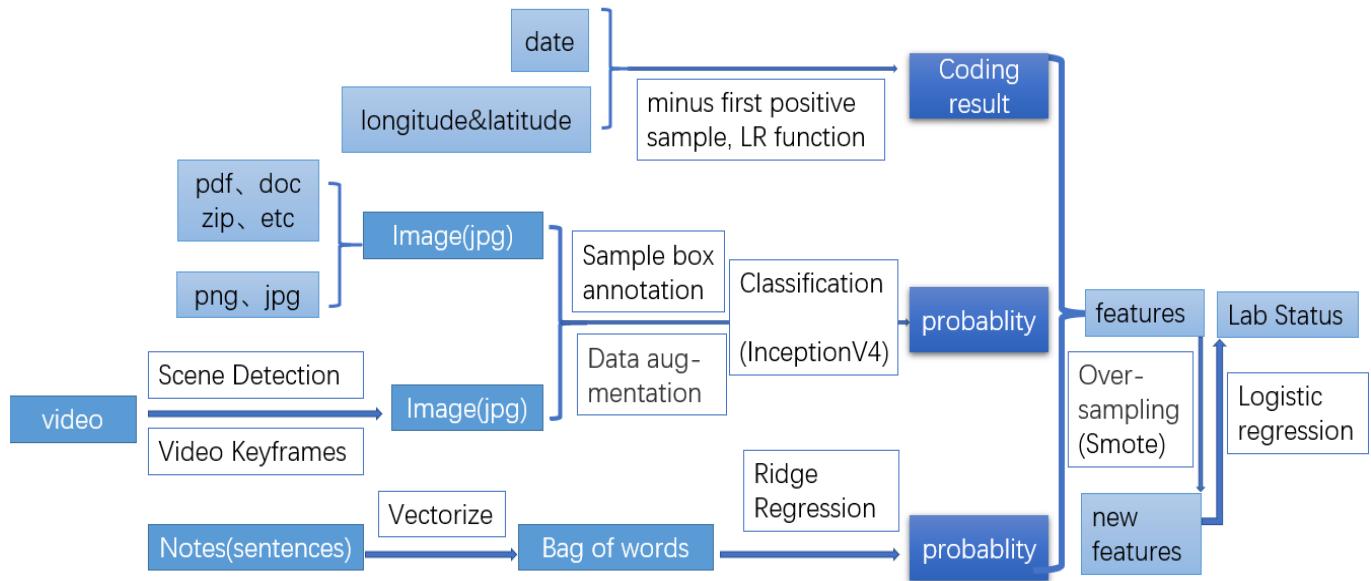


Figure 5: Multimodal Classification Model

4.2 Date,Longitude and Latitude Coding

To deal with these three index,we do two steps:

1. choose the first positive sample as reference(for date we counts the days between two samples) to eminate the influence of magnitude.
2. **LR function** for date: $f(x) = \begin{cases} \frac{1}{x} & |x| > 1 \\ 1 & |x| \leq 1 \end{cases}$ to limit the number.
3. **LR function** for longitude and latitude: $g(x) = x$,for the numbers are all less than 1.

4.3 Video Processing

In the given reports there are 92 vedio. Some videos almost keep quiet and can get quite clear image of the object, while others may contains great movement and can not recognize things clearly. In order to use all these information completely, we use **Scene Detection** to recognize **Scene Boundaries** and get **Video Keyframes** from video and porcess them with images together.

4.3.1 Scene Detection

Video scene detection algorithms generally use the degree of similarity difference between frames to measure, if the video frame is greater than a certain threshold, it is considered a new scene, otherwise it is not a new scene.

To do this there are mainly two ways:

- **scikit-video**[4]:it uses the similarity between color and fringe to detect.

- **ffmpeg**[5]:it chooses the frames considering the compression degree

Compared these two methods we use ffmpeg for it's quicker and more efficient.

4.3.2 Video Keyframes

Video Keyframes are frames used for Video compression and Video codec and contain complete information. Other non-key frames will be compressed by the difference between them. Video frames can be divided into three types:

- **I** :frame represents the key frame and is the most complete frame. Generally, I frame is selected for the video cover.
- **P** :frame single prediction frame, using the previous I frame or P frame, using motion prediction method for interframe prediction coding.
- **B** :frame bidirectional prediction frame, using bidirectional frame prediction coding.

And we firstly use **ffprobe** to get IBP frames, and then combine them to construct jpg image.

4.4 Image Processing

After combine the image provided and the image produced from the video, we cope with them in two ways:

1. **Sample box annotation**
2. **Data augmentation**

to make the information for efficient thus improve classification's accuracy.

4.4.1 Sample Box Annotation

In many pictures, the insect is hidden in the environment such as flowers and woods and the size of the insect is small.

So it's hard to distinguish the insect through these pictures. To deal with these we try to annotate the sample box. For the amount of pictures is large, we do this by **YoloV5**[6], an auto-annotate algorithm:

- First we annotate the sample box by hand to get 750 picture as train data.
- Then we use YoloV5 to deal with more than 2000 pictures.

The annotation result is pretty good as we can see in the picture below.

And as we can see in 8, the sizes of sample boxes are not uniform, so it's necessary to do **Sample Box Annotation** to extract information correctly.

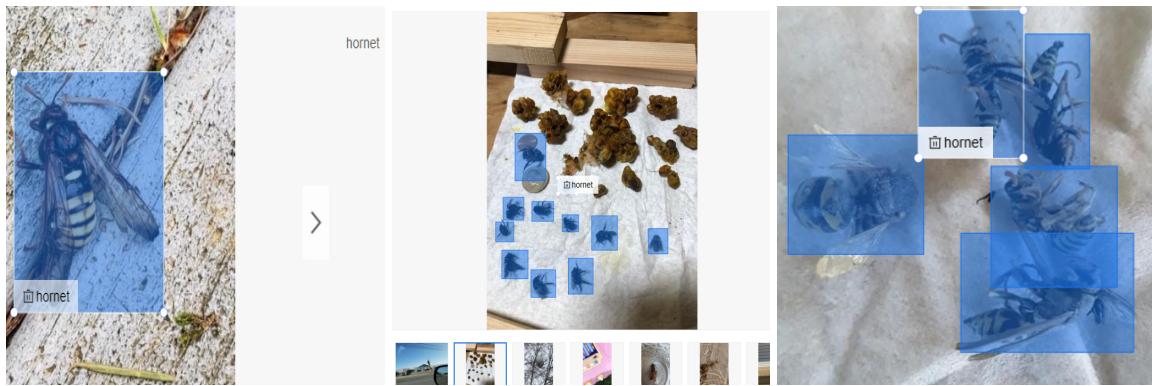


Figure 6: Sample box annotation by hand

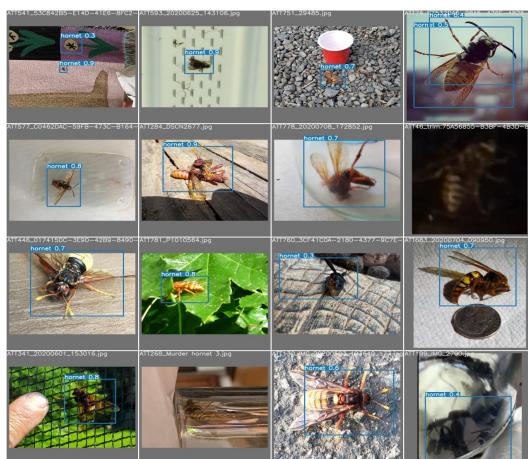
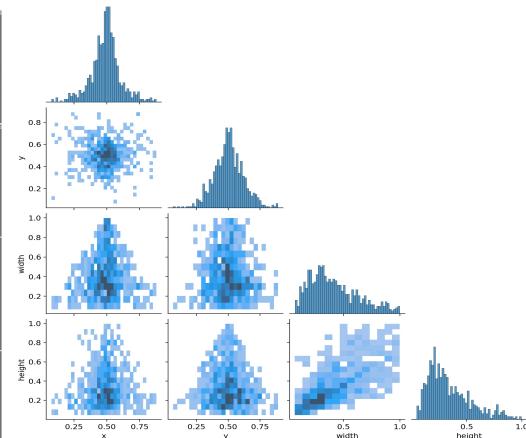


Figure 7: Sample box annotation by yolo

Figure 8: the size and position sample box¹

4.4.2 Data Augmentation

For the number of positive sample is too small, we use **data augmentation** to create more pictures from one positive sample through photo change. Thus we can get more information about how Asian Giant Hornet looks like.

As 8 shows, the size and shape of the sample box is not regular (The center of the sample frame is normally distributed and the height and width are skewed), so we choose to do the change randomly, that's to say the angle we rotate or the size we cut and other parameter are chosen randomly, so we can get more useful information.

More detailed operation can be seen in the table 4, and a result can be seen in the picture 9.

¹x,y represent middle point of the sample box, width and height describe the shape.

Table 4: Data augmentation operation

Operation	python code ²
Flip	RandomHorizontalFlip(), RandomVerticalFlip()
Rotation	RandomRotation(180, expand=True), RandomRotation(180, expand=False),
Affine	randomAffine(180, (0.3, 0.3), (0.4, 1), 3)
Crop	RandomResizedCrop(size), transforms.RandomCrop(size, pad_if_needed=True), transforms.CenterCrop(size), transforms.RandomPerspective(distortion_scale=0.3)
Color	transforms.RandomGrayscale(0.2), transforms.GaussianBlur(9), transforms.ColorJitter(0.8, 0.8, 0.8, 0.4)

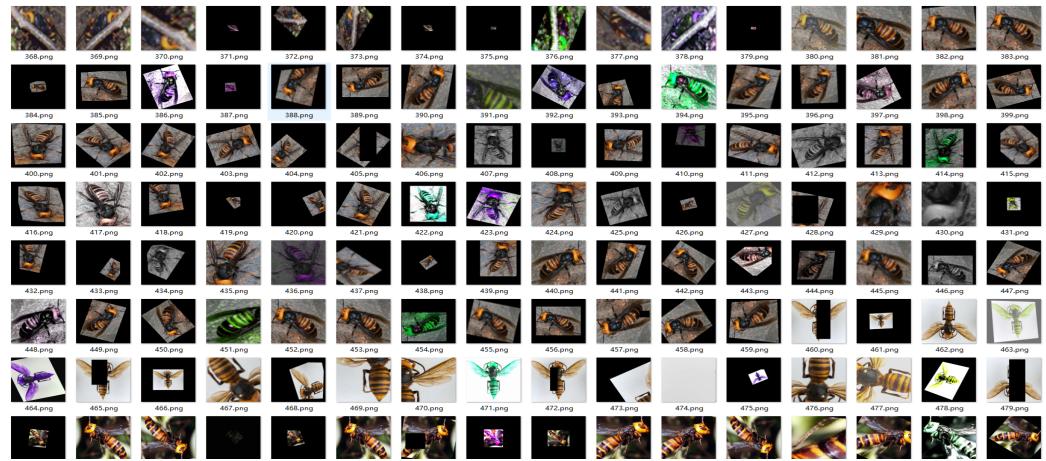


Figure 9: Data augmentation result

4.5 Image Characteristics Extraction

After we process all the image and video, we get dozens of pictures. Then we need to extract features from these samples for further classification.

To do this, we choose a **CNN model** by GoogLeNet: **Inception V4**. For it's much quicker than Resnet and other picture classification networks.

4.5.1 Focal Loss

Through

4.6 Language Processing

In the given information, there are notes with each record, to deal with the words we need to code them first. We use **N-grams** to **vectorize** the sentence. This takes two steps:

²Python code using torchvision.transforms package.

1. take adjacent words to construct combinations.
2. represent the combinations with their frequency in whole sample.

Through then we change a sentence into a string of numbers.

5 Targets and Action

6 Senitivity Analysis

7 Strengths and Weaknesses

7.1 Strengths

- Strength1.
- Strength2.
- Strength3.

7.2 Weaknesses

- Weakness1.
- Weakness2.

8 conclusion

Different methods may show different ranks of effect according to the users' demand, the conditions outside the bathtub and of the bathtub, and so on. However, according to our research in Section 6, these recommended methods are still useful under various conditions, if the details (e.g. the temperature and the time of period) are properly adjusted. So the users can choose and adjust the bath strategy based on the rules below:

Rule I *High temperature and large flux can make the water attach the optimum temperature earlier, meanwhile it may cause uneven temperature and waste of water.*

Rule II *Evenness comes from continuous inflow with appropriate temperature, a little higher than the optimum temperature in usual.*

Rule III *The best strategy to have an excellent bath is to heat up the water before it cool down.*

Memo

Dear users:

Thank you very much for choosing our bathtubs. Each piece of your support is our valuable treasure which can encourage us to develop better bathtubs. We express our sincere thanks to you and your family! For this kind of bathtub, we have some suggestions for you:

1. You don't have to worry about the full-water condition, because when the bathtub is full of water, the overflow drain can drain out the excess water.
2. With the consideration that you will have your special demands on the bath water, we give you some suggestion through so that you can choose what you prefer (or just freestyle you bath).

References

- [1] Facebook. Asian Giant Hornet Watch. <https://www.facebook.com/groups/hornets>.
- [2] Washington State Department of Agriculture. 2020 Asian Giant Hornet Public Dashboard. <https://agr.wa.gov/departments/insects-pests-and-weeds/insects/hornets/data>.
- [3] Wikipedia. Asian Giant Hornet. https://en.wikipedia.org/wiki/Asian_giant_hornet.
- [4] Scikit-Video . <https://github.com/scikit-video/scikit-video/blob/master/skvideo/measure/scene.py>.
- [5] ffmpeg . https://ffmpeg.org/ffmpeg-filters.html#select_002c-aselect.
- [6] YoloV5 . <https://github.com/topics/yolov5>
- [7] Lu J.A., Shang A., Xie J., Gu P., MATLAB-Solution of Partial Differential Equations (in Chinese). *Wuhan University Press*, 2001.

Appendices

Appendix1

```
1 # Scene Detection
2 ffmpeg -i video name -filter:v "select='gt(scene,0.1)',showinfo" -f null - 2>&1
3 ffmpeg -i 666051400.mp4 -filter:v "select='gt(scene,0.1)',showinfo" -f null -
4 2>&1
5 #uses ffprobe get IPB time:
6 ffprobe -i 666051400.mp4 -v quiet -select_streams v -show_entries
7 frame=pkt_pts_time,pict_type
8 #transform IPB to jpg
9 ffmpeg -i 666051400.mp4 -vf "select=eq(pict_type\,I)" -vsync vfr -qscale:v 2 -f
10 image2 ./%08d.jpg
```