The Opioid Crisis

Xinmiao Wang

Department of Mathematics, Northwest University, Xi'an City, Shaanxi, China

wangxinmiao@126.com

Keywords: Cluster analysis, BP neural network

Abstract: On the basis of clustering analysis, the prediction of future drug reports. For the first part, we established a propagation feature model of the reported synthetic opioid and heroin cases. For each county, the clusters of each state are packaged by cluster analysis, and the clustering centers of the districts of the counties in each state are given. For the second part, we established the BP neural network model, and the feasibility test of the neural network trained with the 2010 data was conducted with the 2011 data, and it was found that the model had a high credibility. Finally, take 2017 as an example, the model gives the forecast of 2018 drug reports and the counties that need to worry.

1. Propagation model establishment and solution

1.1 Model Preparation

Information Collection

Since the propagation characteristics of opioid cases are closely related to geographical location, we collected 461 Longitude and latitude[1] of all the counties in the title, and used latitude and longitude as the factors affecting the reports of reports in the following analysis.

1.2 Establishment and solution of propagation models between states and counties

1.2.1 Cluster analysis for data stratification of counties and counties

First, cluster analysis[2] is carried out by using the two indicators of the specific performance of a certain drug in each county, the ratio of the number of drugs in the county to the number of drugs in the state, and the counties in each state are layered.

In order to exclude the influence of time and dimension on the drug reports, the average of the above two indicators after standardization in 2010-2017 is selected for cluster analysis;

In Ohio, for example, some of the conclusions from cluster analysis are shown below:
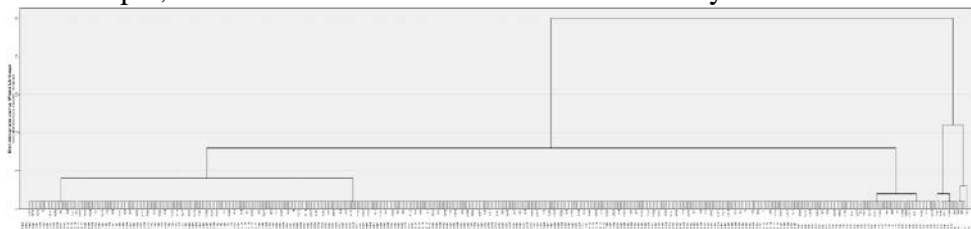

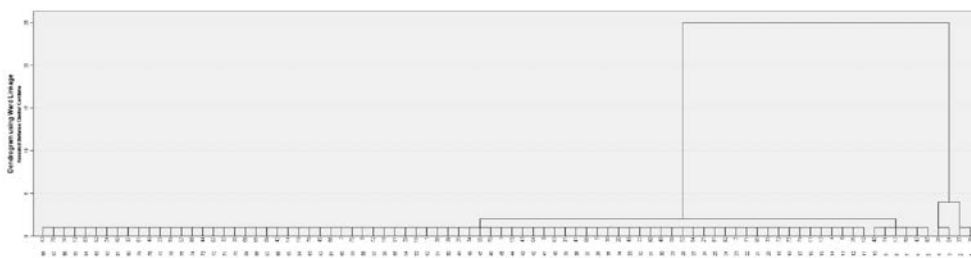Figure 1 Columnar ice crystals of synthetic opioids in Ohio

Figure 2 Columnar ice crystals of synthetic opioids in Ohio

According to the above columnar ice crystal map, we divide Ohio into three categories, and the codes of the counties corresponding to each grade are shown in the following Table:

Table 1. Ohio County Clustering Results

1	39001 39011 39013 39005 39003 39007 39009 39015 39023 39025 39043 39045 39051 39073 39077 39047 39017 39021 39031 39039 39053 39041 39069 39019 39027 39029 39035 39049 39055 39059 39063 39065 39057 39033 39037 39061 39067 39071 39081 39099 39103 39083 39101 39123 39127 39075 39085 39109 39079 39093 39091 39097 39107 39113 39147 39105 39115 39119 39121 39111 39137 39125 39139 3914 39159 39117 39131 39129 39135 39149 39153 39087 39089 39095 39157 39161 39133 39163
2	39165 39143 39167 39173 39175 39141
3	39155 39171 39151 39169

1.2.2 Traversing the search method to establish the cluster centers of each state

Using the traversal search method, for the zoning results of the counties in each state obtained by cluster analysis, traverse all the counties in each district, and combine the latitude and longitude coordinates of each county to find the sum of the shortest Euclidean distances from the county to all the other counties. The county corresponding to the minimum value is the cluster center of each packed area.

The packaging areas for the five states are shown in the following figure:
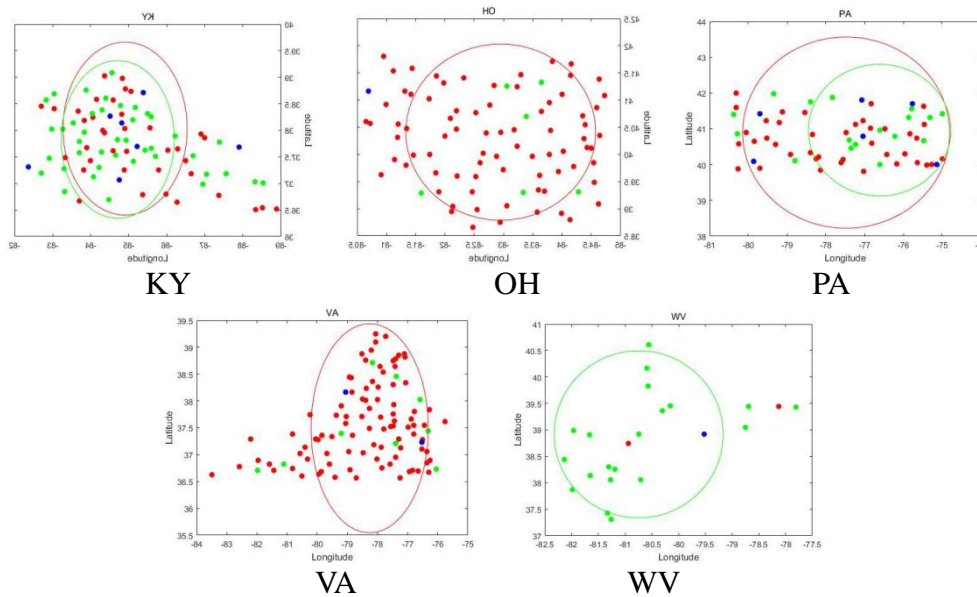

KY   OH   PA

VA   WV

Figure 3 State Cluster Zoning Map

The latitude and longitude are respectively horizontal and vertical coordinates, and red, green and blue respectively represent the counties of the states belonging to the first, second and third districts.

Since the degree of influence of the first to third zones obtained by the cluster analysis is decreasing, we use the cluster center of one area as the cluster center of each state. The cluster analysis results of each state are shown in the following Table:

Table 2. State Cluster Center Results Table

State	KY	OH	PA	VA	WV
Latitude	38.0331	40.4098	40.8936	37.4914	38.9154
longitude	-84.9031	-82.9516	-77.4765	-78.2577	-80.7476

2. Prediction model

2.1 Establishment of BP neural network model

2.1.1 Model selection

We decided to use data to train the BP neural network [3] and apply this model to predict when and where the number of future drug cases may exceed the threshold.

Since the latitude and longitude of a certain location is fixed, in order to avoid the occurrence of multi-valued functions, we select the longitude and latitude of the location of the drug case as a two-input of the neural network within a fixed year, but due to longitude, Latitude does not change over time, so we add the total drug reports in the year we want to predict as the third input to the neural network, and the drug reports for the year to be predicted as the output establish BP neural network mathematics model.

2.1.2 Processing of errors

In the first training process, because some counties did not have a drug case, there is no specific value in the given data, which leads to direct training error, and the results do not converge, so we extracted all the counties. The latitude and longitude data, first assign all the county's drug reports to zero, and then add the original data to get a new set of data (the new data gives zero value for cases that have not appeared).

2.1.3 Model feasibility verification

We used the total data of all the counties in the five new states in 2010 to retrain the network and forecast the data for 2011. After the forecast is over, we use the original data of 2011 to test and compare the results. Good predictive effect.

The self-test results of the state heroin and synthetic opioid cases are shown in Figures 4 and 5, respectively:
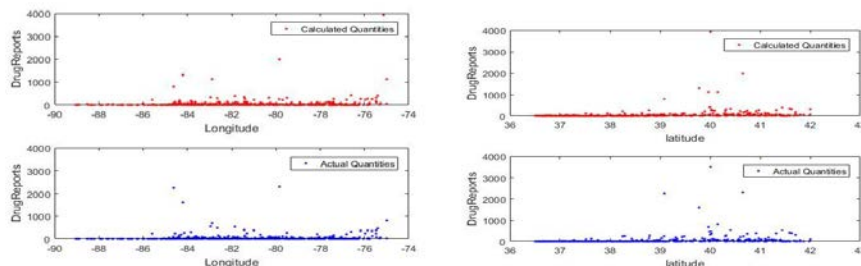

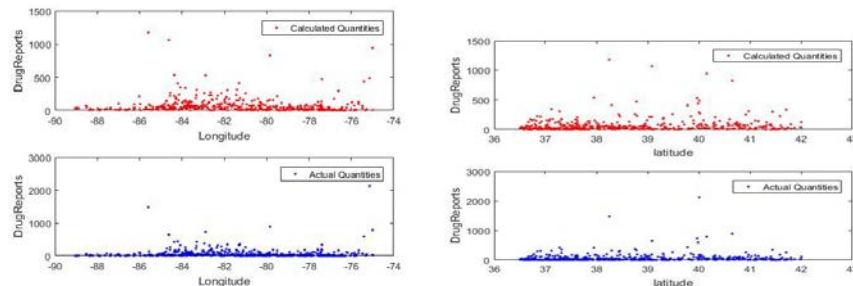Figure 4 Self-test results of heroin cases in all counties of five states

Figure 5 Self-test results of synthetic opioid cases in all counties of five states

Among them, the changes of the reports with the longitude are reflected in the left picture, and the changes of the reports with the longitude are reflected in the right picture; the red indicates the Calculated Quantities and the blue indicates the Actual Quantities. It can be seen from the figure that the actual value of the drug reports in 2011 is basically coincident with the predicted value, and the prediction effect is better.

2.2 Establishment of grey prediction model

First, the possibility of detecting the gray prediction model is tested. The existing total drug reports from 2010 to 2016 are used as input to predict the model output, which is the total drug reports of 2017. The 2017 forecast for the model is compared with the 2017 data given in the question, and a scatter plot is drawn:
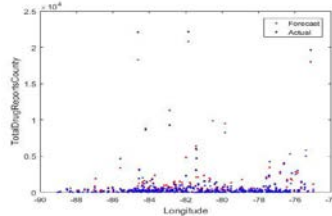

Figure 6. Comparison of the true and predicted values of the grey model in 2017

In the figure, the blue dot indicates the true value, and the red dot indicates the predicted value. The two are basically coincident, indicating that the prediction effect is more reliable.

Therefore, we will predict the model output by using the existing total drug reports from 2010 to 2017 as the input of predictive 2018, which is the total drug reports of 2018.

2.3 Establishment of evaluation model

Let us take the predicted 2018 reports as the example. We will use the gray prediction model[4] to find the 2018 total toll as an input. The longitude and latitude of the location will be used as the other two inputs to train the new one.

The neural network is used to predict the particles of the five states except 2018.

Next, we judge the county in the position corresponding to the portion of the 2018 year that the drug reports exceeded the threshold. We extract the portion of the data that exceeds the threshold to make a figure 7.

2.4 Solution of predictive model

We use matlab software to obtain the latitude and longitude of the counties in the five states that exceeded the threshold F_0 in 2018 as shown in the following figure:
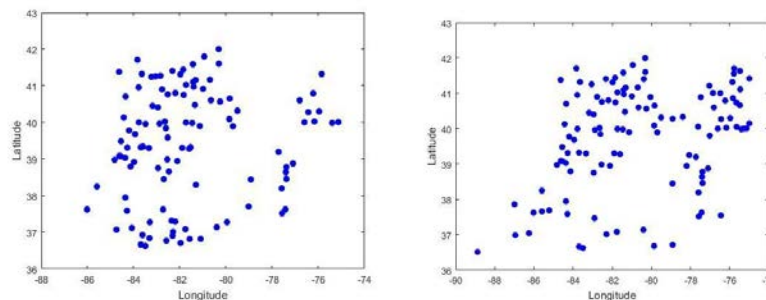

Figure 7.The latitude and longitude of counties where the number of drug cases exceeds the threshold in 2018

Among them, the scatter plot on the right indicates that the number of heroin cases exceeds the threshold, and the scatter plot on the left indicates that the number of synthetic opioid cases exceeds the threshold.

References

[1] https://simplemaps.com/data/us-cities

[2] Sarstedt, Marko, and Erik Mooi. "Cluster analysis." A concise guide to market research. Springer, Berlin, Heidelberg, 2014. 273-324.

[3]Sietsma, Jocelyn, and Robert JF Dow. "Creating artificial neural networks that generalize."

Neural networks, (1991): 67-79.

[4] Hsu, Li-Chang. "Applying the grey prediction model to the global integrated circuit industry." Technological Forecasting and Social Change, (2003): 563-574.