

Drug trade analysis model based on location judgment

Xinyuan Liu *

School of Materials Science and Engineering, South China University of Technology, Guangzhou, China

* Corresponding author: 907875322@qq.com

Abstract. According to the current spread of new drug problem, this paper established the blocks the growth model of correction, to complete the determination of drug eruption time node, then based on the geographical position the drug eruption time model is established in this paper, using the fisher discriminant method, by geographical location, the estimated time of drug eruption and USES the multi-layer perceptron neural network to predict counties of drug eruption time, Finally, the robustness analysis is completed. This article establishes drugs.

Keywords: Fiser's criterion, neural network, Robustness analysis.

1. Introduction

Drug abuse in the United States has increased in recent years. There have been drug outbreaks in many places. Drug abuse leads to HIV epidemics, hepatitis transmission, mental disorders, and affects economics indirectly. In order to better control and supervise drugs such as opium and heroin, especially to deal with the spread of new drugs, it is necessary to do the statistics and analysis on the drug detection records of crime statistics bureau throughout the country. We need to establish mathematical models to identify the source of drug abuse, summarize the characteristics of drug transmission and take targeted measures. Now, we get a record of many types of drugs reported by counties in the five eastern states of the United States (Ohio, Pennsylvania, Virginia, West Virginia, and Kentucky) in 2010-2017 (may not be complete). In order to study the effects of other related factors, we also make use of the economic data from recent years. By studying the relationship between different variables and drug detection quantity, we can extract variables that are more valuable for studying drug detection quantity. Then we take action on these valuable variables. [1]

2. Data preprocessing

We first observe the drug detection situation in each state over the years. Trends in drug seizures in five eastern states from 2010 to 2017 is shown as figure 1. Drug seizures in five eastern states is shown as figure 2. [2]

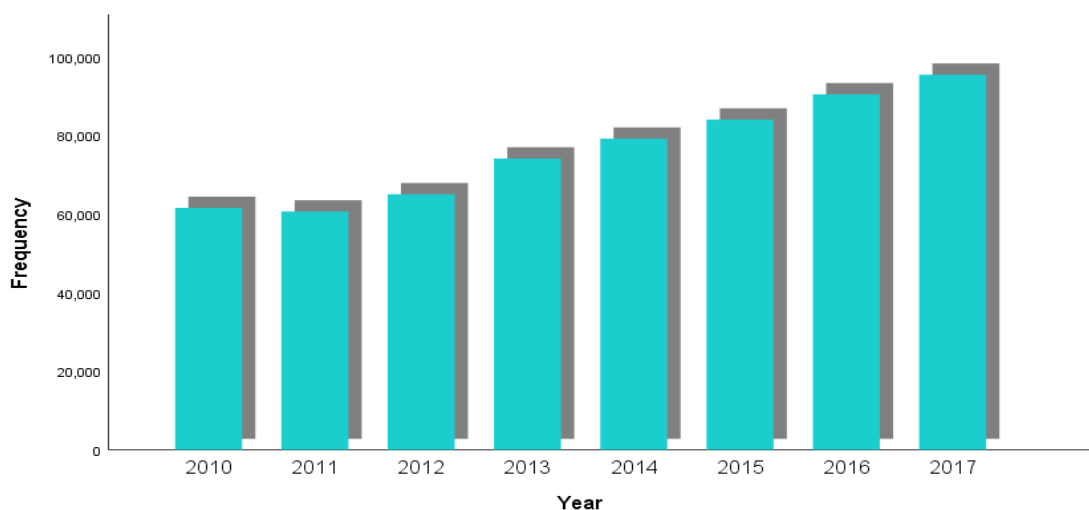


Figure 1. Trends in drug seizures in five eastern states from 2010 to 2017

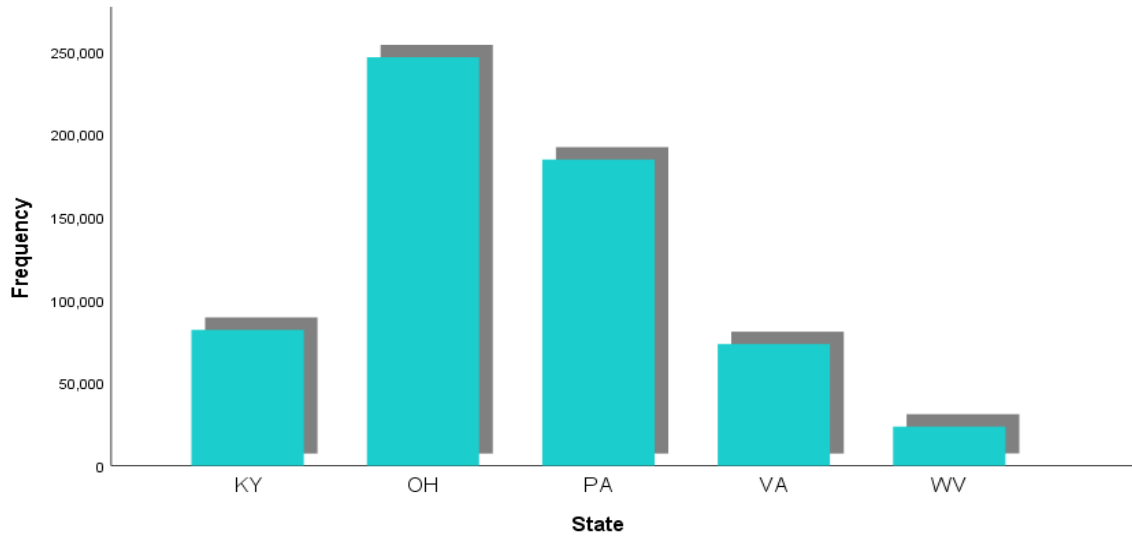


Figure 2. Drug seizures in five eastern states

3. Modified retarded growth model

3.1. Retarded growth model

Typical block growth model curve is shown as figure 3.

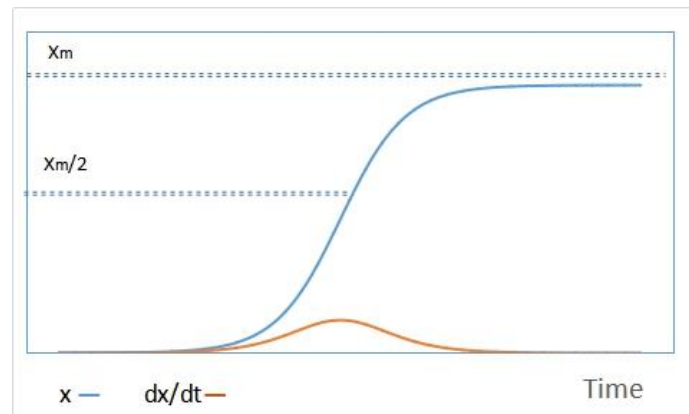


Figure 3. Typical block growth model curve

We believe that the number of specific drugs detected reflects the size of the local drug market. We assume that the trend of the drug market in a place is subject to the retarded growth model[1]. Typical retarded growth model hypothesis: The growth rate of the drug market x is a function that varies with x , And the ratio to the unrealized part of the market (for the market's maximum capacity X_M) is proportional to The proportional coefficient is the inherent growth rate r .

$$\frac{x_m - x}{x_m}, \text{ Expressed in differential: } \begin{cases} \frac{dx}{dt} = r \left(1 - \frac{x}{x_m} \right) x \\ x(t_0) = x_0 \end{cases}$$

3.2. Modified retarded growth model

By observing the actual trend of specific drug detections in a county, we found three types of changes, called A, B, and C. These curves are not very consistent with the block growth model.

Based on the observation of the three curves, we corrected the model. Three typical drug detection volume change curve is shown as figure 4.

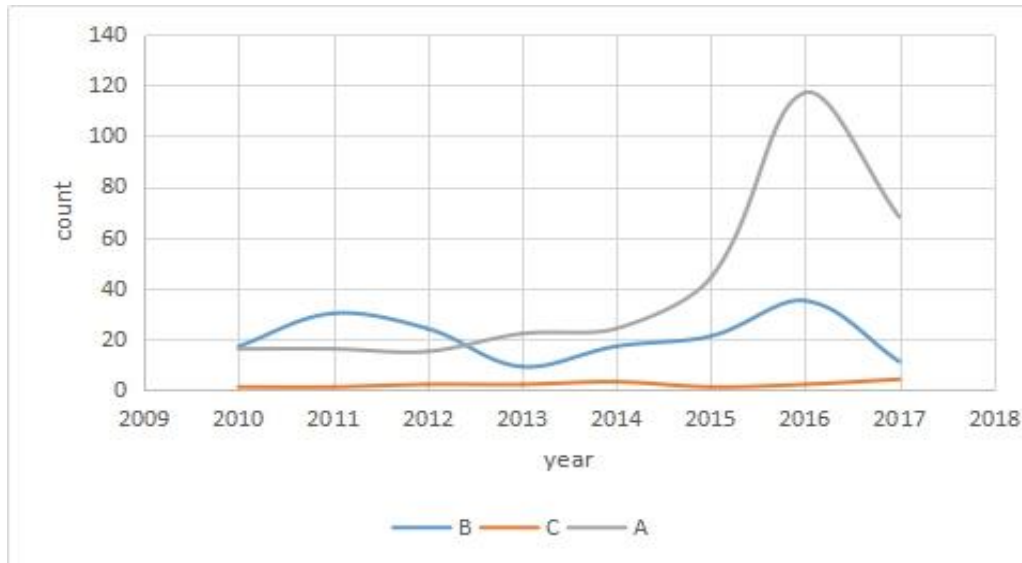


Figure 4. typical drug detection volume change curve

The modified model argues that the maximum capacity of the drug market is a variable over time, approximated as a piecewise function. The market was very large and the abuse of drugs developed rapidly at the beginning. When the market has grown to a certain extent, it attracts the attention of the government and police. After a crackdown by the government, the market capacity shrinks dramatically. [3]

The modified model can be described as:

$$\begin{cases} \frac{dx}{dt} = r \left(1 - \frac{x}{x_m(t)} \right) x \\ x(t_0) = x_0 \\ x_m(t) = \begin{cases} A \text{ huge amount, Before being found by the police} \\ A \text{ small amount, After being found by the police} \end{cases} \end{cases}$$

The modified model can explain the three types of curves very well.

For the A-type curve: A-type curve is the performance of the new drug that just entering into the local market. It has not attracted the attention of the police in the early stage, and the growth is rapid. At this time, the market has a large capacity. But when the police and government take measures to hit the market, its capacity is drastically reduced, which leads to the amount of drugs detection fell back.

For B-type and C-type curves: Both B-type and C-type curves fluctuate randomly within a certain range. B has a large range of fluctuation and the market share of drugs (the ratio of the number of a certain drug detected in that year of a county to the total drugs). B is considered to be a type of drugs that has existed for a long time in that area. C has a small fluctuation range, and the number of detections is only a single digit. Besides, the market share of C is also small. So it is considered that C is a type of drug that has not been erupted locally.

3.3. Determination of drug outbreak time nodes

Through the modified block growth model, we have a deeper understanding of the trend of drugs in a certain place. We can classify the samples into three categories: A: The drug has been erupted in the recent years (2010-2017), B: This drug has been erupted before 2010, C: This drug has not yet erupted. Only samples of A can get specific time of drug outbreaks.

Combining the model and experience, we propose the following methods of discrimination:

If such kind of drug erupted in the recent years (2010-2017), the curve of A in the previous period is approximately exponentially increasing, and the first derivative and the second derivative are positive in the previous period.[4]

First-order derivative , Second-order derivative. $x' = \frac{(x_{t+1} - x_{t-1})}{2t}$ $x'' = \frac{(x_{t+1} + x_{t-1} - 2x_t)}{t}$

In order to eliminate the impact of the overall market changes in drugs, it is also necessary to observe whether there is a significant increase in the market share of such drug. After calculating the market share changes before and after the outbreaks for 20 groups. We think that if the ratio of market share in the next year to the market share the year before is larger than 8, it is reasonable to judge that there is an eruption of such kind of drug in that year[5].

A sample is considered an A-type sample if it satisfies both of these conditions. Otherwise, it should be classified as B or C. In the scope of observations, if the overall market share is small (less than 1%), or if the number of drugs detected is only a single digit, we believe that it is Class C. That is, the drug has not yet erupted in the region. Otherwise, it is Class B. That is, the area has already erupted before the observation time[6].Flow chart of discriminant method is shown as figure 5.

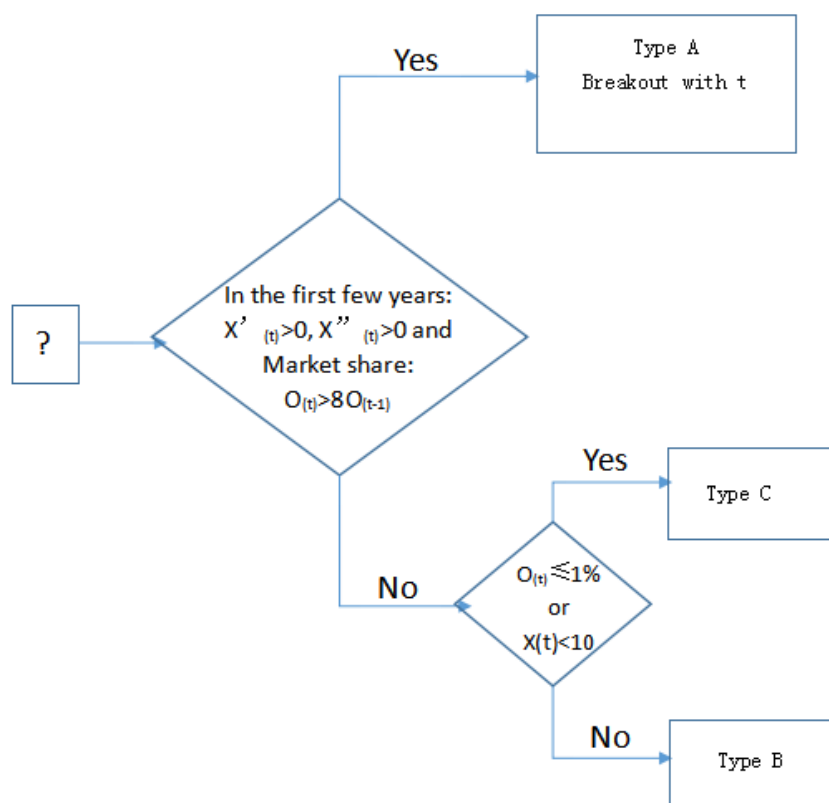


Figure 5. Flow chart of discriminant method

By using this discriminant method, we divide all the samples into three categories, A, B, and C, and obtain the specific outbreak time of the type A. [7]

4. Prediction model based on geographic location

4.1. Fisher's criterion

Fisher's discriminant method was proposed by Fisher in 1936. The method is to determine the function according to the criterion that the variance within the class is as small as possible and the variance between classes is as large as possible. The basic principle is to use projection technology to project k sets of p -dimensional data into a certain direction so that the projection group of data is separated from the group as much as possible. The idea of analysis of variance is used in the separation of different groups. [8]

Through Fisher discriminant method, the longitude and latitude are independent variables. We use the specific outbreak time of A and C as categorical variables, and fit the outbreak time of fentanyl

in different counties in Kentucky. There are 3 prediction errors in 18 counties, And the correct rate is 83.3%.(For the convenience of Prediction and statistics, we estimate C as the year 2020 after the observation range. Similarly, if there is B, it is estimated as the year 2009 before the observation range).

Using fisher discriminant method to predict the outbreak time of specific drugs in counties is shown as table1. Coefficient of classification function is shown as table2.

Table 1. Using fisher discriminant method to predict the outbreak time of specific drugs in counties

City	WHITLEY	PIKE	PERRY	KNOX	HARLAN	BELL	PULASKI	LAUREL	CARTER	BOYD	MONTGOM	JESSAMI	MADISON	JEFFERS	KENTON	FAYETTE	CAMPBELL	BOONE
Observed	C	C	C	C	C	C	16	16	16	16	15	15	14	14	13	13	13	13
Predicted	C	C	C	C	C	C	16	16	16	16	15	15	15	13	13	15	13	13

Table 2. Coefficient of classification function

	F				
	13	14	15	16	C
longitude	407.207	406.414	403.671	400.964	398.496
latitude	522.367	518.15	515.729	512.426	507.107
(constant)	27331.593	27102.645	26778.777	26426.658	26021.721
Fisher linear discriminant function					

If the longitude(lo) and latitude(la) of a county are known, the Fisher discriminant can be used to determine its category. The specific method is to input its latitude and longitude into the Fisher discriminant.

$$\begin{cases} F(13) = -407.207 * lo + 522.367 * la - 27331.593 \\ F(14) = -406.414 * lo + 518.15 * la - 27102.645 \\ F(15) = -403.671 * lo + 515.729 * la - 26778.777 \\ F(16) = -400.964 * lo + 512.426 * la - 26426.658 \\ F(C) = -398.496 * lo + 507.107 * la - 26021.721 \end{cases}$$

The F value of each categorical variable is compared, and larger the F value the larger the probability to be that kind of category.

If we only judge the county belongs to A or C, irrespectively of knowing the specific outbreak year of drugs in the type A, using the Fisher discriminant can get a greater accuracy. Using Fisher Discriminant to Predict the outbreak of Specific Drugs in Each is shown as table3.

Table 3. Using Fisher Discriminant to Predict the outbreak of Specific Drugs in Each County (A or C)

City	WHITLEY	PIKE	PERRY	KNOX	HARLAN	BELL	PULASKI	LAUREL	CARTER	BOYD	MONTGOM	JESSAMI	MADISON	JEFFERS	KENTON	FAYETTE	CAMPBELL	BOONE
Observed	C	C	C	C	C	C	A	A	A	A	A	A	A	A	A	A	A	A
Predicted	C	C	C	C	C	C	C	C	A	A	A	A	A	A	A	A	A	A

The accuracy rate was 88.9%.It can be seen that this method is effective for predicting the fentanyl outbreak time in Kentucky.

4.2. Use neural network multilayer perceptron

In addition to the Fisher discriminant method, the SPSS neural network multilayer perceptron can also achieve good results.

The neural network model originated from the study of the human brain thinking mode. The Fisher method is a linear method, and the neural network is a nonlinear data modeling tool. The input layer and the output layer, one or more hidden layers constitute the nerve. Meta-relationships between

neurons are assigned weights, and the training learning algorithm continually adjusts these weights during the iterative process, thereby minimizing prediction errors and giving prediction accuracy.[9]

It can be seen from Fig. 8 that the neural network multilayer perceptron is comparable to the Fisher method in the correct rate in predicting the specific time node. However, there are minor differences between the two in the specific sample forecast.

We use the longitude and latitude as covariates, and the outbreak time of drugs as a categorical variable. Through the distribution of 70% training set and 30% sample set, a neural network model with only one hidden layer and four elements in the hidden layer is proposed. As shown in Figure 9.

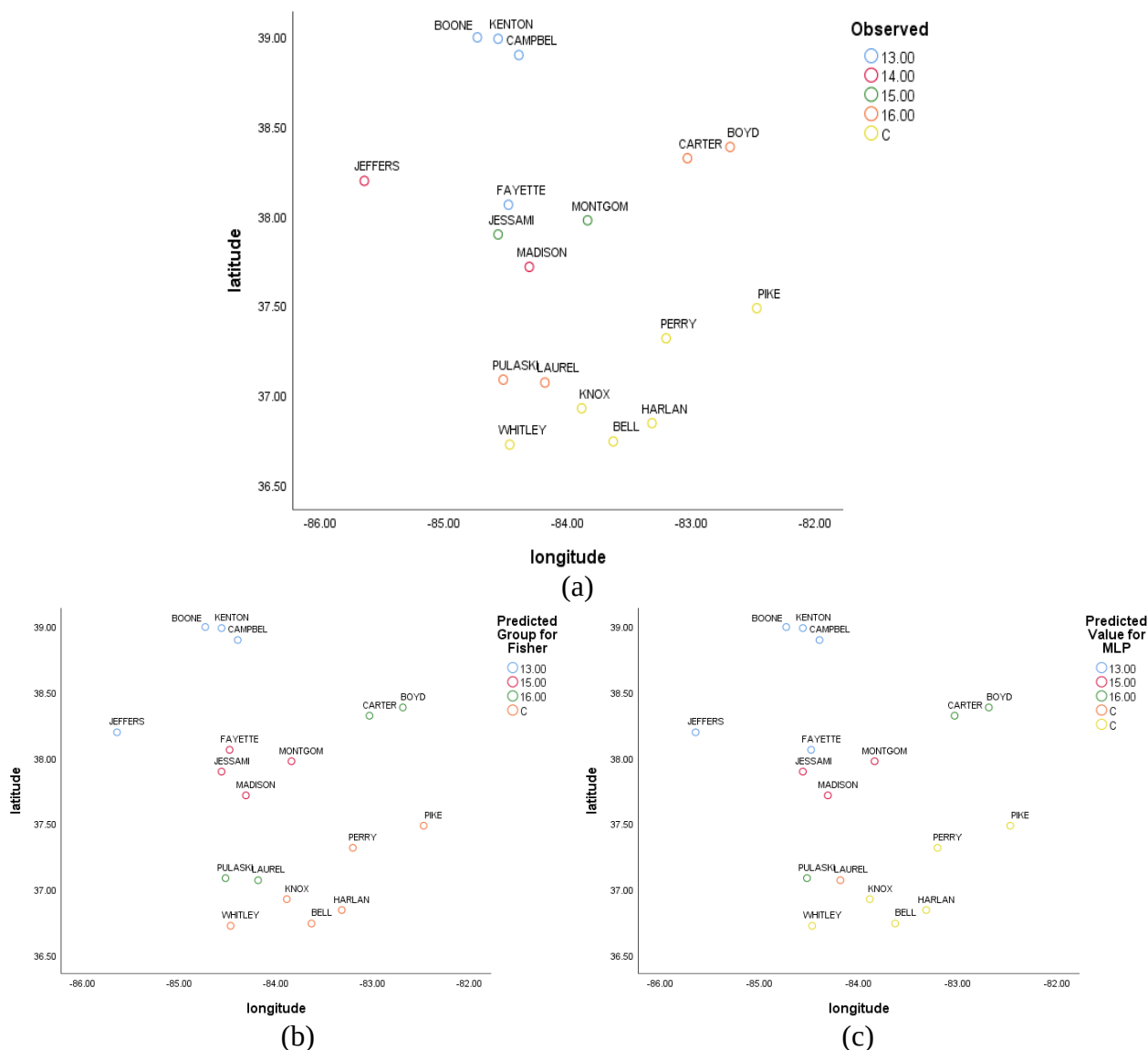


Figure 6. Prediction of the fentanyl outbreak time in Kentucky based on different prediction methods (a) observations; (b) Fisher Method, correct rate 83.3%; (c) Neural network multilayer perceptron method with a correct rate of 83.3%

Neural network multi-layer perceptron predicting the time of specific drug outbreaks is shown as figure 7.

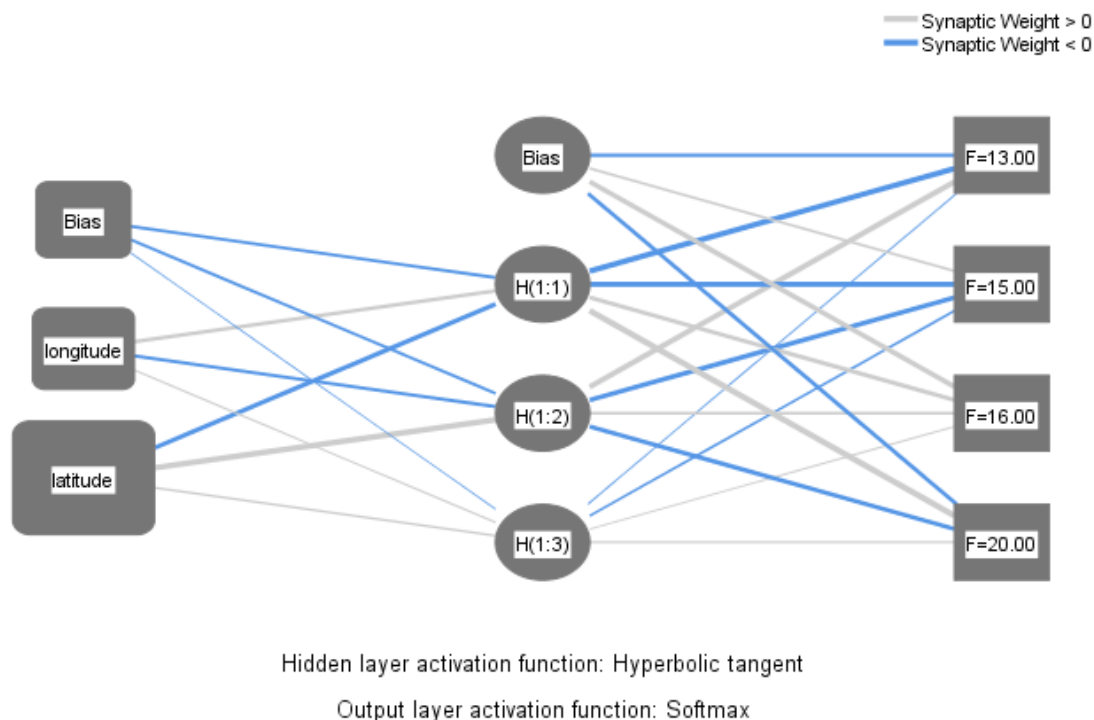


Figure 7. Neural network multi-layer perceptron predicting the time of specific drug outbreaks

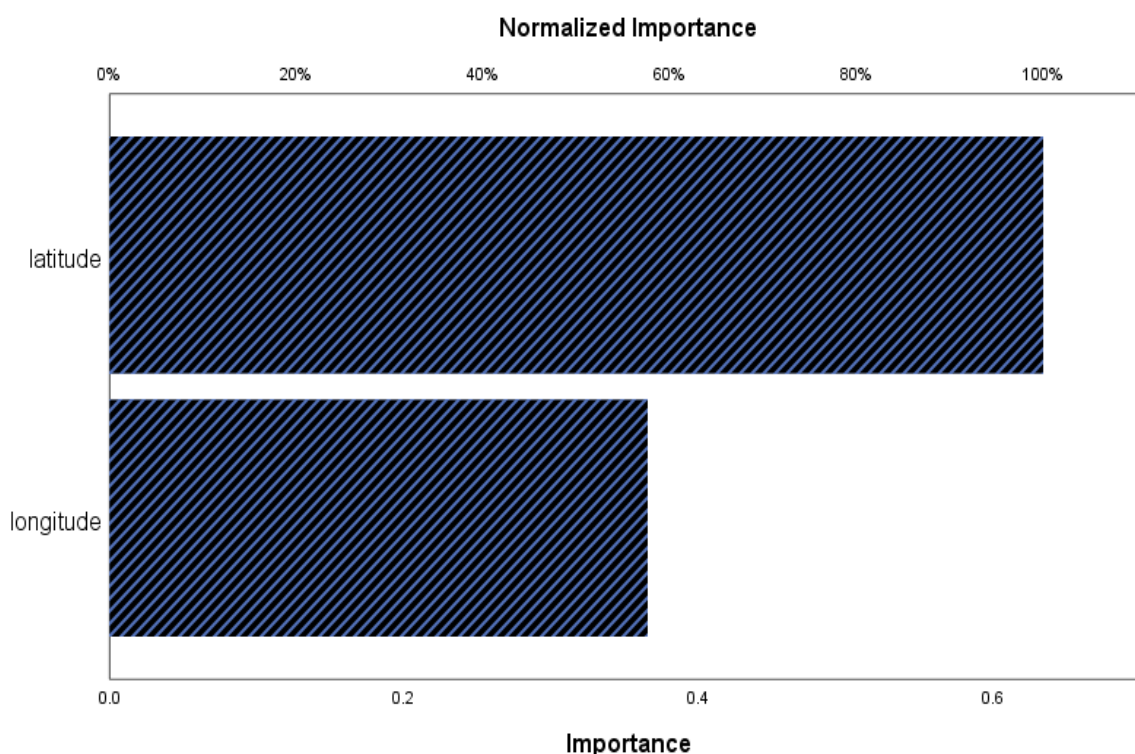


Figure 8. Independent variable weights in a neural network model.

Independent variable weights in a neural network model are shown as figure8. The neural network assigns more weight to the latitude. This suggests that latitude changes are the main reason for the uneven distribution of drug outbreaks in Kentucky. This is in line with our intuitive understanding of Kentucky's fentanyl propagation from north to south. when we divide the samples into A and C categories, regardless of the specific outbreak time but whether outbreak in the recent years, The neural network method performs even better. The correct rate is as high as 94.4% with only one of the 18 samples making an error. [10]

4.3. Robustness analysis

The trend of drug propagation in Kentucky is obvious and simple, which is spreading from northwest to southeast. Fisher method works well, and the neural network works better for it. But when we study other states, we notice that some states have more complicated drug propagation patterns. For example, the outbreak time of fentanyl in each county in Pennsylvania is shown as figure9.

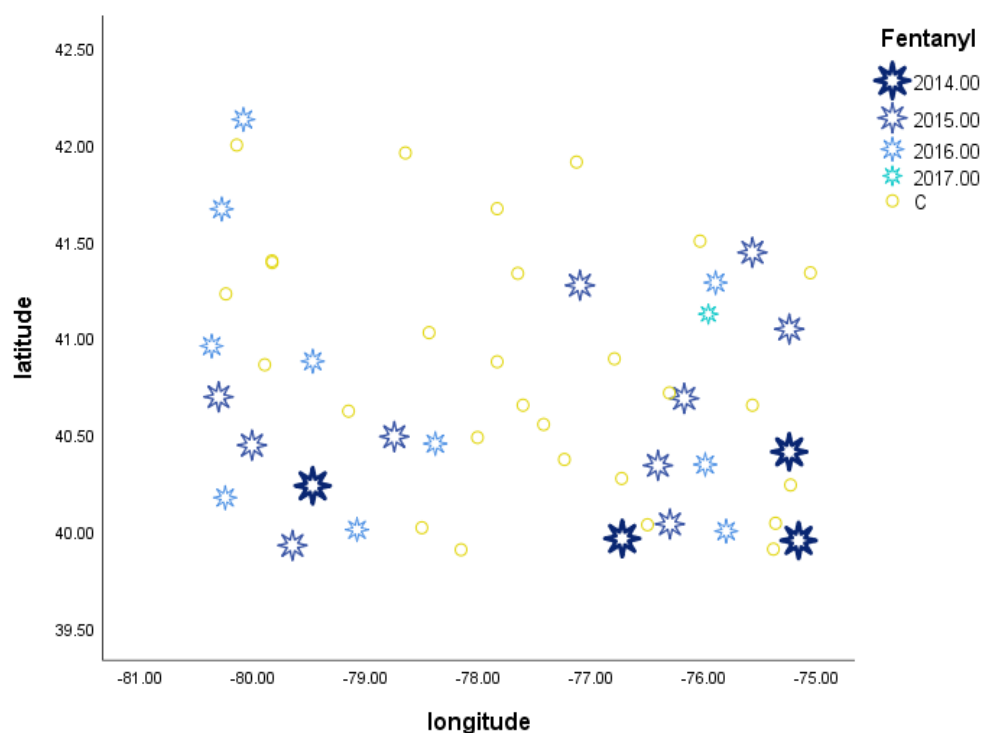


Figure 9. The outbreak time of fentanyl in each county in Pennsylvania

As can be clearly seen from the figure, the trend of outbreak time of fentanyl in Pennsylvania is not in a simple direction. Intuitively seen from this figure, Pennsylvania has two outbreak centers, one in the southwest and one in the southeast, which makes the propagation of fentanyl in Pennsylvania more complicated.

When we used the Fisher method to predict Fentanyl's outbreak time in Pennsylvania with only the longitude and latitude as the independent variables, the results are often not good. Fisher Classification Results is shown as table 4.

Table 4. Fisher Classification Results^a

		Observed	Predicted Group Membership					Total
			14	15	16	17	C	
Original	Count	14	3	0	1	0	0	4
		15	2	0	4	4	0	10
		16	2	0	7	1	0	10
		17	0	0	0	1	0	1
		C	7	4	7	5	5	28
	%	14	75	0	25	0	0	100
		15	20	0	40	40	0	100
		16	20	0	70	10	0	100
		17	0	0	0	100	0	100
		C	25	14.3	25	17.9	17.9	100
A 30.2% of original grouped cases classified correctly.								

The correct rate is only 30.2%, As can be seen from Figure 10, when we only use the geographical location as independent variables, the classification of Fisher's discriminant method makes the county in the same type geographically close to each other. For a state like Pennsylvania, which has multiple drug propagation sources, can not merely depend on geographic location to judge.

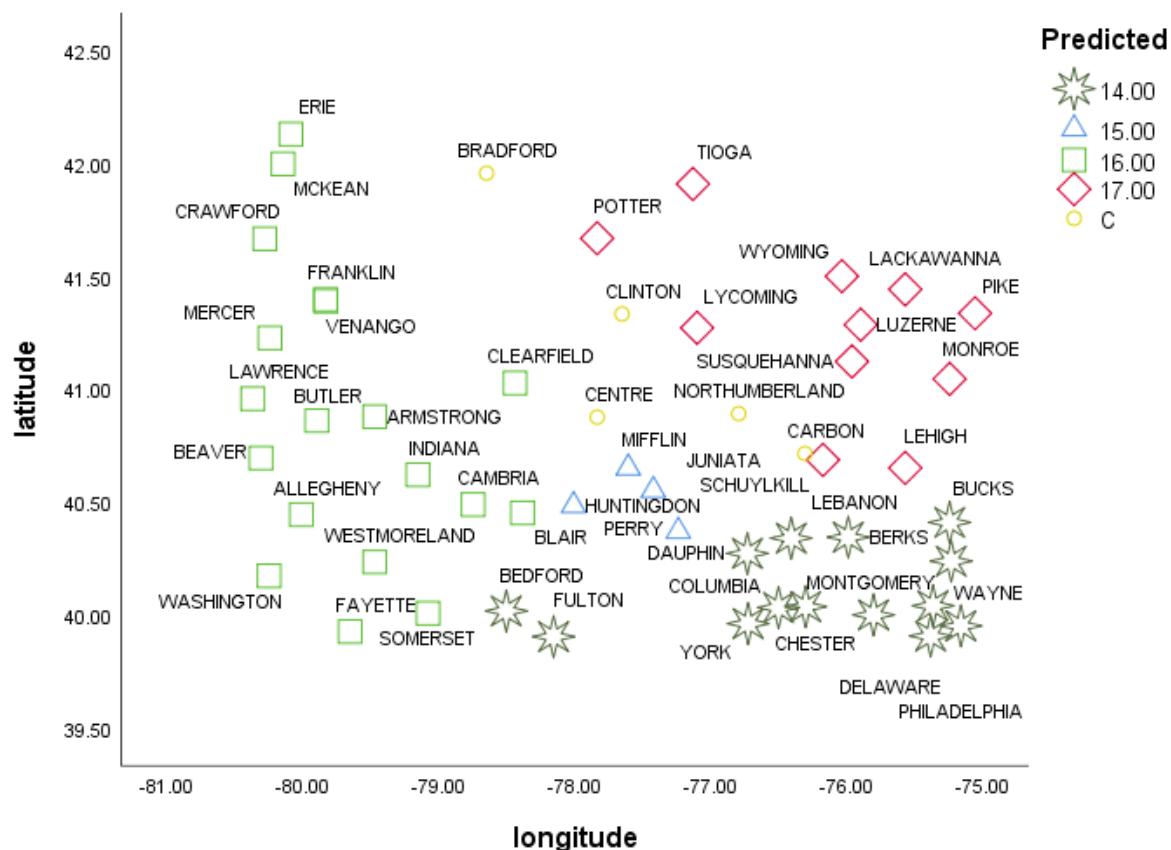


Figure 10. The prediction of Fisher method in the outbreak time of Fentanyl in Pennsylvania's counties

For Pennsylvania, the use of neural network multilayer perceptron, the results are all type C. It is because the type C in the sample accounts for more than 50%, causing huge interference to the neural network model. If we only use the neural network multilayer perceptron to predict the outbreak time, as shown in Figure 11 there are clear boundaries of different types, And the sources' information of the southwest is lost.

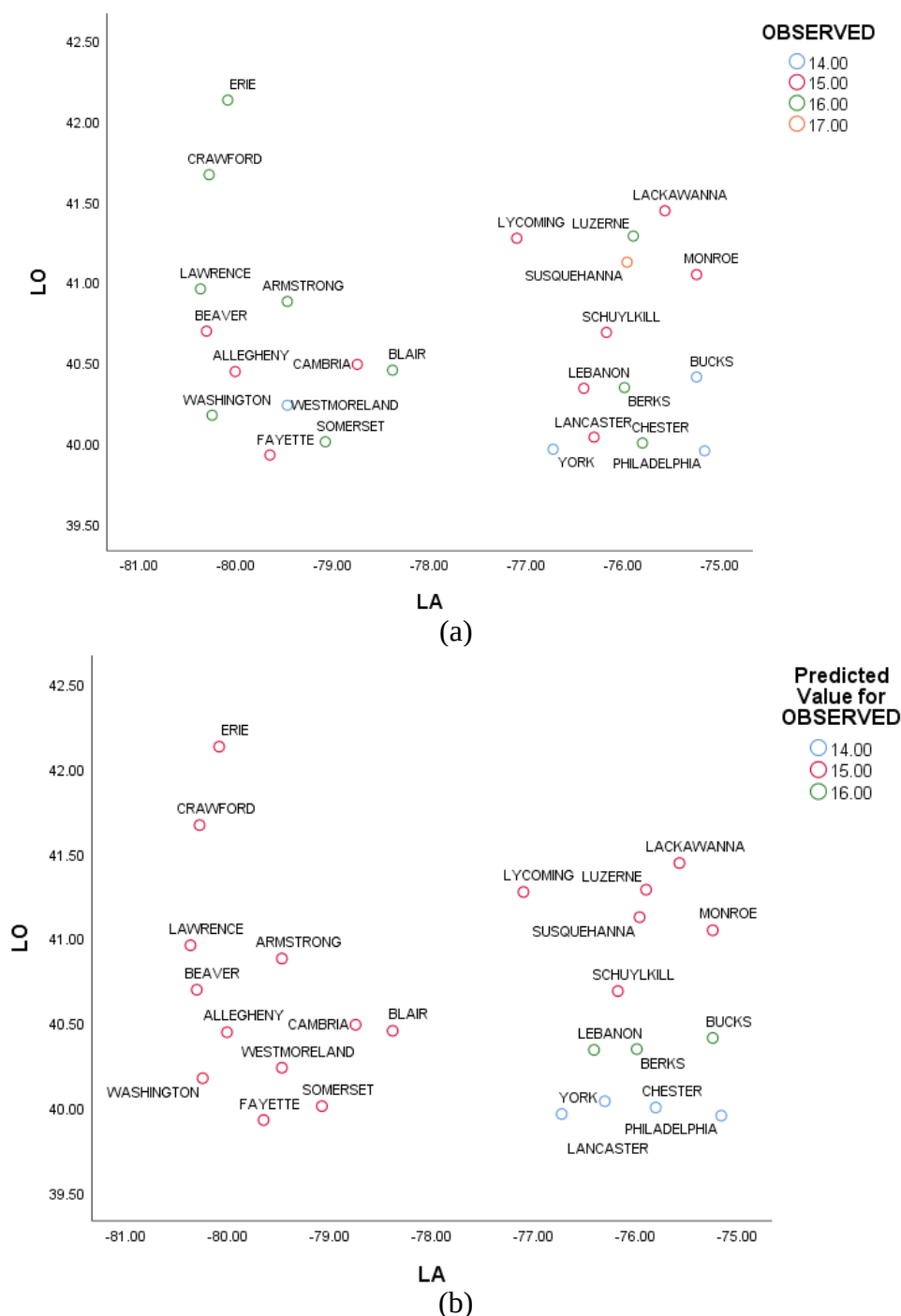


Figure 11. Mlp-predicted Fentanyl Outbreak time of type A in Pennsylvania's counties (A) Observed values; (b) Predicted values

5. Conclusion

We use a modified block growth model to characterize the outbreak of drug development in a place. This type of model serves as a standard to help us classify counties. Then, we use Fisher discriminant and neural network multi-layer perceptron to predict the time of drug outbreak in a certain place with certain independent variables. Essentially, this is a classification process. Finally, we use the weight analysis and principal component analysis in the neural network to find the independent variable that has the greatest influence on the amount of drug detection in a large Number of indicators. This is based on the assumption that "the factors that contribute to the total amount of drug This has also led to The discovery of a greater contribution to a particular drug amount of drug

detection perform well in judging the specific drug detection time, indicating that our assumption is realistic.

References

- [1] Liu Yang I, Xie Jin. Population prediction based on retarded growth model [J]. Shandong Industrial Technology, 2014, (11): 220 - 220. (In Chinese).
- [2] DING Dengwei, GAO Wensheng, LIU Weidong. Insulation Defects Discrimination in GIS by Fisher Discriminant Analysis of Partial Discharge [J]. High Voltage Engineering, 2013, 39 (4): 805 - 812. DOI: 10.3969/j.issn.1003 - 6520. 2013.04.006.
- [3] Speech Magnitude Spectrum Reconstruction from MFCCs Using Deep Neural Network [J]. Electronic Journals, 2018, 27 (2): 393 - 398. The DOI: 10.1049 / cje 2017.09.018.
- [4] Jiang Wenbin, Liu Peilin, Wen Fei. Speech Magnitude Spectrum Reconstruction from MFCCs Using Deep Neural Network [J]. Journal of Electronic Journals, 2018, 27 (2): 393-398. The DOI: 10.1049 / cje 2017.09.018.
- [5] Yuna Ro, Youngwood Ha. A Factor Analysis of Consumer Expectations for Autonomous Cars [J]. Journal of Computer Information Systems, 2019, 59 (1) : 52-60. DOI: 10.1080/08874417.2017.1295791.
- [6] Jun Wei, Xin Liu, Guoqing Jiang. Parameterizing sea surface temperature cooling induced by tropical cyclones using a Multivariate linear regression model with multivariate linear regression model [J]. Acta Oceanologica Sinica, 2018, 37 (1): 1 - 10. DOI:10.1007/s13131 - 018 - 1153 - 0.
- [7] Xu Changjin, Liu Zixin, Liao Maoxin, Li Peiluan, Xiao Qimei, Yuan Shuai. Fractional-order bidirectional associate memory (BAM) neural networks with multiple delays: The case of Hopf bifurcation[J]. Mathematics and Computers in Simulation, 2021, 182.
- [8] Thomas Paul, Rumjit Nelson Pynadathu, Lai Chin Wei, Bin Johan Mohd Rafie. EDTA functionalised cocoa pod carbon encapsulated SPIONs via green synthesis route to ameliorate textile dyes - Kinetics, isotherms, central composite design and artificial neural network [J]. Sustainable Chemistry and Pharmacy, 2021, 19.
- [9] Xiao Yong-Liang, Li Sikun, Situ Guohai, You Zhisheng. Unitary learning for diffractive deep neural network [J]. Optics and Lasers in Engineering, 2021, 139.
- [10] Kamran Iqra, Naz Saeeda, Razzak Imran, Imran Muhammad. Handwriting dynamics assessment using deep neural network for early identification of Parkinson's disease [J]. Future Generation Computer Systems, 2021, 117.