Opioid Control Research Based on BP neural Network and Genetic Algorithm

Yu Zhou[1,*], Xuemei Liu[1], Haoyang Miao[2], Fanyu Zhao[3], Tianxin Liu[1]

[1]Southwest Petroleum University, Chengdu, China

[2]Chengdu University of Technology, Chengdu, China

[3]Southwestern University of Finance and Economics, Chengdu, China

*1355414236@qq.com

Keywords: control neural network; genetic algorithm; relative comparison optimization algorithm

Abstract: In this passage, we first analyze the data. And use the gray correlation degree to find the correlation degree between the counties. Then find out the source of specific opioids used--Philadelphia County. Next, we obtain the spread and characteristics of the incidents (cases) at the position level. Finally, based on the two factors of time and position, we establish a BP neural network model based on time and position. and then classified, and finally classified into five types of factors the neural network model based on control is established. finally, based on the conclusions obtained in the part one and two, we model the factor analysis method to analyze the main reason of affecting the abuse of opioids. According to the numerical size of the contribution rate can simplify the five main reasons to two.

1. Introduction

In recent years, the problem of drug abuse in the United States has become increasingly serious when the abuse of heroin has re-emerged, the quantities of synthetic drugs have remained high and legalization of cannabis has failed to reduce the use of other drugs. At the same time, the abuse of prescription drugs continues to spread. There are 100 million adults in the United States who suffer from chronic pain and they need to rely on opioid analgesics to relieve the pain. Drug abuse also has an impact on the important filed of the US economy. The proliferation of drug abuse has led to employment shortages in many places in the US. And the number of people who have died from excessively drugging has increased. By analyzing the characteristics of the spread of opioids in the United States and the abuse of opioids, there are many strategies for responding (respond)to drug crises proposed. Therefore, our task is to propose a possible strategy for dealing with drug crises by establishing mathematical models.

2. Related work

2.1 Problem Restatement

The problem that we need to solve in this paper are:

A mathematical model was developed to describe the spread and characteristics of synthetic opioids and heroin events between the five states and their counties, and to identify where opioids were first used in the five states.

Based on established mathematical models, the prevalence of opioids and heroin in five states was predicted and thresholds and corresponding times were determined.

According to the US Census social and economic data, analyze the causes and current situation of opioid abuse, and establish a mathematical model to clearly define the relationship between the trend of use and the socio-economic data of the US census.

A mathematical model is established to identify possible strategies for dealing with the opioid crisis and to test the validity of the model and the important parameter boundaries on which it depends.

  DOI: 10.25236/isrme.2019.113

2.2 Assumption

• By processing the data, we assume that the more drug reports in a state or a county, the more synthetic opioid and heroin incidents there are.

• When Principal Component Analysis is used, the factors that are dealt with can be seen as having no effect on the spread of opioids.

• When the spread of the reported synthetic opioid and heroin incidents (cases) in the position, the influence factors of each county such as the speed of the spread are the same and there is no individual difference.

• The values of each set of data in the census socioeconomic data table are in the range of error, so the estimates can be used directly.

• Counties with higher occurrences of opioids and heroin events in each state were extracted as cluster points each year, and a two-dimensional fit was established with their latitude and longitude.

• All the data in the attachment are considered to be authentic.

3. Our Model

3.1 Part one

3.1.1 Communication Center Model Based on gray correlation

Sometimes, it is difficult to identify any possible locations where specific opioid use might have started in each of the five states

The grey relational degree analysis method is a new system analysis method proposed by the grey system theory. The basic idea of the gray correlation is to judge the degree of association, which is the degree of association, according to the degree of similarity of the sequence curve geometry. The closer the curve is to the curve, the greater the correlation between the corresponding sequences and vice versa. There are assumptions that m evaluation targets and n evaluation indicators. Reference number is xx_0 ={ $xx_0(k)$|k=1,2,...,n }, Comparison number is xx_i ={ $xx_i(k)$|k=1,2,...,n }, i=1,2,...,m. Compare the gray correlation coefficient of the series i x to the reference number column x_0 on the kth indicator is

$$\xi_i(k) = \frac{\max_s \max_t |xx_0(t) - xx_s(t)| + \rho \max_s \max_t |xx_0(t) - xx_s(t)|}{|xx_0(t) - xx_i(t)| + \rho \max_s \max_t |xx_0(t) - xx_s(t)|} \quad (1)$$

Where, $\rho \in [0,1]$ is Resolution factor, and $\max_s \max_t |xx_0(t) - xx_s(t)|$ are the two-level minimum difference and the two-level maximum difference. Some of the data obtained through correlation analysis are as follows:

TABLE I. Partial data obtained from correlation analysis

PHILADELPHIA	HAMILTON	ALLEGHENY	MONTGOMERY
0.970100793	0.970099825	0.970096152	0.970090253
BUCKS	JEFFERSON	DELAWARE	......
0.97007875	0.9700647	0.970058749	......

From the Table I, it can be seen that the county of Philadelphia is a possible location where specific opioid use might have started in each of the five states. At the same time, we also made a correlation analysis of 2011-2017 to prepare for the latter questions.

3.1.2 Time-based univariate analysis

According to the preliminary analysis of the data provided by NFLIS, we extract the drug report data of each state of each year and establish a model of Time-based univariate analysis between the year and the reported synthetic opioid and heroin incidents (cases). We assume that the year is the independent variable, and the data of the reported synthetic opioid and heroin incidents (cases)in every state as the dependent variable. The trend the data of the total drug reports in every states from

2010 to 2017 by MATLAB programming .The picture of the trend of total drug reports in every states is as follows:

From the trend of each state over time, we can roughly judge the spread of the reported synthetic opioid and heroin incidents (cases) between each state and county: the drug reports of Ohio has been rising, while the drug reports of Pennsylvania is declining year by year, and Virginia was fluctuating between 2012 and 2014. The numbers in West Virginia and Kentucky are basically stable.

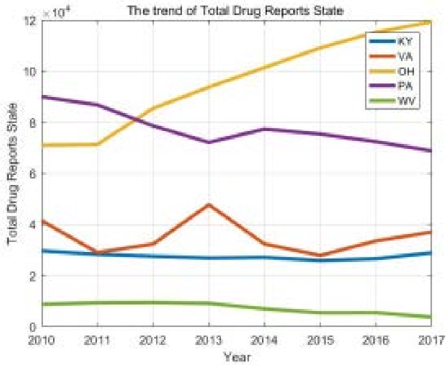

Figure1 The trend of Total Drug Reports States

Therefore, we take the Total Drug Reports data for each of the five states, and set the regression curve equation for the trend of the data over time and the regression is

$$y = a_1 x_1^m + a_2 x_1^{m-1} + \ldots + a_m x_1 + a_{m+1} \qquad (2)$$

Where a_i(i= 1,2,...,m+1) is constant, We fit the unary polynomial regression plot of all drug identification numbers over time in five states (see fig. 2)

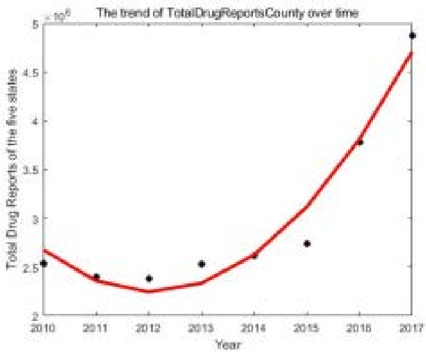

Figure 2 the trend of Total Drug Reports County over time

Conclusion: We use a model of Time-based univariate analysis , obtaining a linear change in the spread and characteristics of the reported synthetic opioid and heroin incidents (cases) in and between the five states and their counties over time as a quadratic change.

3.1.3 Space-based multivariate fitting model

We analysis and process the NFLIS data provided, then extract and filter the reported synthetic opioid and heroin incidents (cases) in and between the five states and their counties. We find that the reports of synthetic opioid and heroin incidents (cases) in and between the five states and their counties are associated to the positions of the counties. And we extract the counties whose reported synthetic opioid and heroin incidents (cases) are high, the quantities are over 1000. Then by Google Map, we find out the latitude and longitude data of the counties which are extracted by us.

We establish a Position-based multivariate fitting model, using two-Dimensional fitting method to obtain the two-dimensional relationship between the number of the reported synthetic opioid and heroin incidents (cases) and latitude and longitude and its curve expression in each year. So as to analyze the spatial propagation characteristics of opioids and heroin events. A three-dimensional view of the location distribution and the number of the reported synthetic opioid and heroin incidents (cases) between the counties in 2010 and 2017 is obtained as follows:

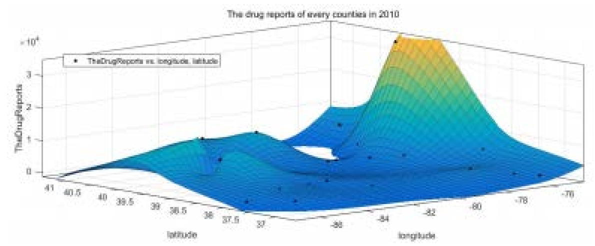

Figure 3 Distribution of drug identification numbers in 2010

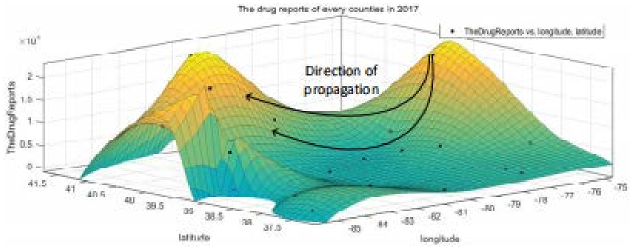

Figure 4 Distribution of drug identification numbers in 2017

Using the CRPTOOL toolbox in MATLAB, we input the datas of the latitude and longitude of every counties as the independent variable and input the data of Total Drug Reports County as the dependent variable, and then a relational expression is output as:

$$f(x', y') = 22 + 2.487\sin(0.2995\pi\, x'y') + 0.1266e^{-(0.8177y')^2}\ (95\%) \quad (3)$$

Conclusions: The spatial distribution of opioids and heroin events is characterized by: from Philadelphia County in Pennsylvania to the county near Akron, Ohio (its latitude and longitude coordinates (-81.625°W, 41.01°N)) and the county near JEFFERSON City, Kentucky (its latitude and longitude coordinates (-85.50°W, 39.55°N)

3.1.4 Time and space based BP neural network model

Opioid and heroin events can be obtained by univariate analysis. The number of opioids and heroin events is large, and the number of drug identification is also high. Therefore, the combination of opioids and heroin events is related to time and space. Considering time and space, we have established a BP neural network model based on time and space to study drug identification training.

Based on the previous time factors and geographical factors, a BP neural network twofactor analysis model was established. The BP neural network model consists of three layers of network structure, namely input layer, hidden layer and output layer.(see fig.5)

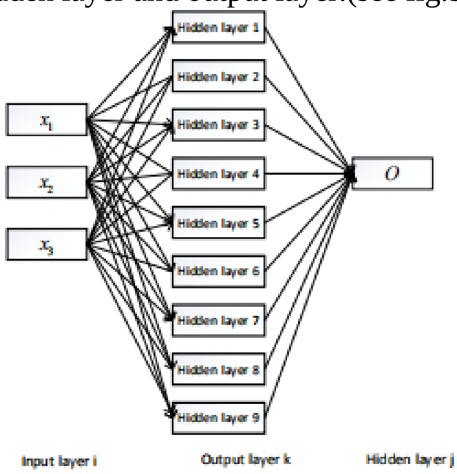

Figure 5 BP neural network structure

According to the input X= (x_1,x_{2i},x_3,)and expected output D=(d_1 ,d_2 ,...,d_l), Determine that the network input layer is 3, the hidden layer is 9, and the output layer is 1. Initialize the weight between each layer of neurons ,V_{ij} W_{jk} , Initialize the hidden layer threshold, the output layer threshold, the given learning rate is 0.01, and the neuron transfer function is a tangent sigmoid function tan sig . See Appendix 3 for the specific calculation process.

Through the mathematical model we established, we forecast the data from 2011 to 2017 and compare it with its actual value. The obtained line chart is shown in figure 6. After calculation, we get the relative error between the actual value and the predicted value is 4.9, so the model is desirable.

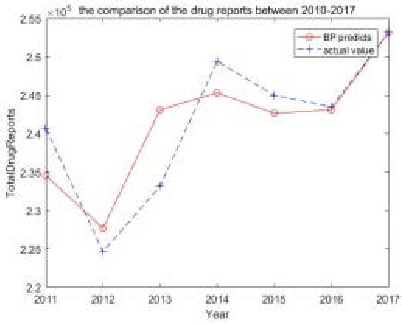

Figure6 The comparison of the drug reports between 2010-2017

The trend graph of the last 20 years is predicted by BP neural network (see fig. 7)

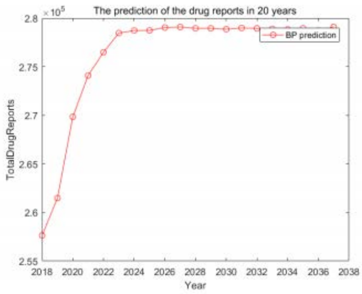

Figure 7 The prediction of the drug reports in 20 years

It can be seen from the figure that the threshold for drug identification is 2023, and has reached a peak of 27,884 in recent years and continues to maintain a high level. Through spatial analysis, it can be concluded that at 2023, the maximum position of the threshold is in Cleveland, Ohio (latitude and longitude coordinates (-81.66972°, 41.48222°))

3.2 Part two

3.2.1 Principal component analysis

According to the instructions, we pre-process the US Census Social and Economic Data According to the instructions, we pre-process the US Census Social and Economic Data Sheet and remove some inapplicable factors, such as data marked with (x). We assume that the value of each set of data is in the error range, so we remove the column containing the error range. Then, according to the data provided by NFLIS, the counties with more drug identifications and fewer drug identifications in each year are extracted, and the counties with faster increase of drug identification in different time periods. Compare the changes in the number of drug identifications in counties with more and less drug identifications in the same year and the counties growing over time, and then combine relevant information. Five major socioeconomic data affecting the number of drug identifications were screened out, and next we assume that other socioeconomic data have no effect on the number of drug dentifications.

Since there are still many different variables in the US Census economic data table, we hope to use less variables to explain most of the variation in the original data table, so we consider constructing a principal component analysis model.All the data in the NFLIS table are sorted according to the total number of drug identifications in each county, and then mapped to the census table. At the same time, the corresponding total number of drug identifications in each county is arranged in the corresponding order which is based on the the gray correlation method in the part one

Finally the Principal component analysis model is constructed.

Then use the sample data to perform multiple linear regression to get an interactive picture.(see fig. 9)

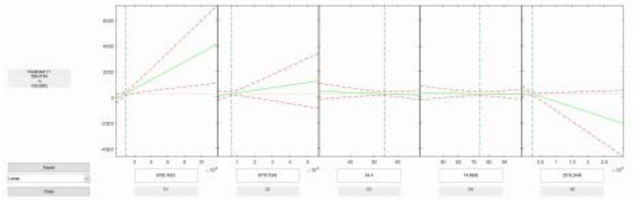

Figure 8 multiple linear regression interactive screen

The first one from left to right is the curve y(x1) confidence interval when x5,x6,x7,x8 is fixed, and the other graphs are analogous. The corresponding curve equation is:

$$Y=5.641*10^{2}+0.035x_4+0.020x_5-7.055x_6-1.799x_7-0.081x_8 \quad (4)$$

The standard deviation is 3.075, and the standard deviation is within the allowable range, so we draw five mean factors of affecting the abuse of opioids ,which are educational attainment (especially less than 9th grade)、residence 1 year ago (especially different house in the U.S.- different county)、marital status (especially divorced) 、grandparents (especially responsible for grandchildren – years responsible for grandchildren - 5 or more years)and ancestry (especially Scotch-Irish).

3.2.2 BP neural network model based on controllable variables

Due to there are multivariate varieties when we take the five mean factors of affecting the abuse of opioids. Based on the neural network model established by the part one, we modify your model from Part 1 and construct a BP neural network model based on controllable variables which the input layer is extended from time, position (latitude and longitude), drug identification, educational attainment – less than 9th grade, residence 1 year ago , different house in the U.S.- different county, marital status – divorced ,grandparents – responsible for grandchildren – years responsible for grandchildren - 5 or more years and ancestry – Scotch-Irish, etc. The model includes the five new controllable variables. Under the trained neural network, adjusting the value of the variable of the input layer can obtain an output layer with relatively small relative error. The number of controlled cases is obtained through controllable variables.

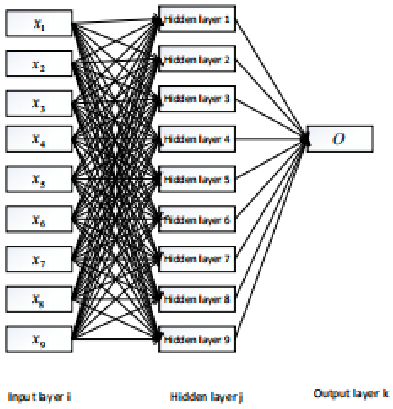

Figure 9 BP neural network structure

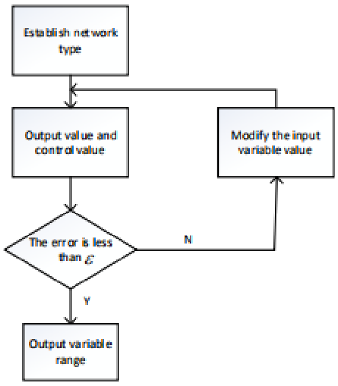

Figure 10 BP neural network with controllable variables

It is the output is

$$Q = \min \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2 \qquad (5)$$

According to the input X= ($x_1, x_{2i}, x_3, \ldots, x_9$) , and the expected output D=(d_1 ,d_2 ,...,d_l), the network input layer is determined to be 9, the hidden layer is 9, the output layer is 1, the weight V_{ij} W_{jk} between each layer of neurons is initialized, the hidden layer threshold d is initialized, and the output layer is Threshold e, given a learning rate of 0.01 and a neuron transfer function as a tangent sigmoid function tan sig . The number of controlled cases is obtained through controllable variables.

4. Sensitivity Analysis

4.1 Time-based grey prediction model

We use GM(1,1) model to check formal model in part one. The calculation process of the model is shown in Appendix 3. The curve is obtained by gray prediction (see fig.11)

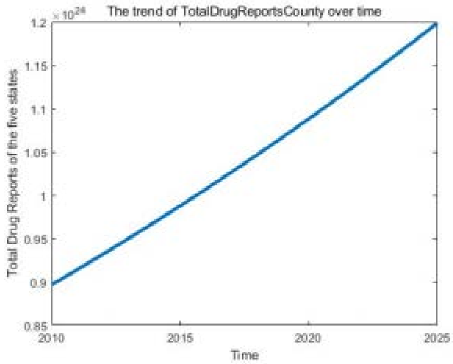

Figure 11 The trend of TotalDrugReportsCounty over time

From the figure, we can get the logistic function change of opioid and heroin events over time, in line with the final predicted output map of the neural network.

4.2 Stability Analysis of BP Neural Network with Controllable Variables

When the other independent variables are not changed, the original data is brought into the BP neural network training function, and the picture of the output through the BP neural network is as follows

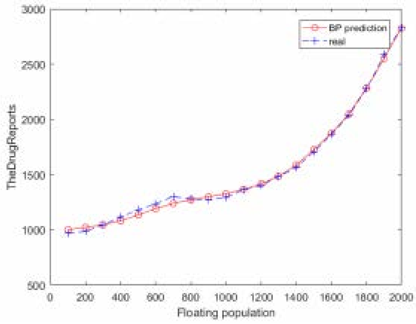

Figure 12 The number of cases varies with the floating population

By halving the data of the floating population variable of the model input layer (HC01_VC121) and regenerating the BP neural network training function, the relationship between the floating population and the case quantity can be regained, as shown in Figure 13.

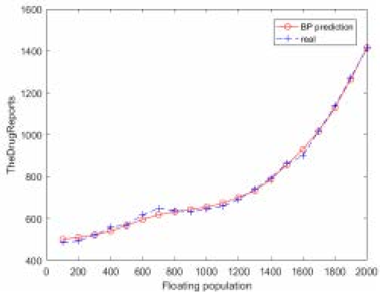

Figure 13 The relationship between the floating population and the case after halving

Average relative error:1.547

Through the analysis of Fig. 16 and Fig. 17, the floating population variable of the input layer is reduced, but the average relative error of the function after training by BP neural network is still small which indicating that the BP neural network with controllable variables is relative stable.

Under the condition that the variable population of the input layer is reduced, the quantity of drug identification in the output layer has changed significantly, which also indirectly indicates the influence of the floating population on the number of drug identification.

5. Strengths and Weakness

5.1 Strengths

We use three different methods of gray correlation, factor analysis and principal component analysis to analyze, process, and extract the data, avoiding the leakage caused by one method.

In the process of studying the propagation characteristics of the space, the threedimensional view can be used to observe the propagation trend very intuitively.

The BP neural network essentially implements a mapping function from input to output, which makes it particularly suitable for solving complex problems of internal mechanisms, and the global training results are not destroyed after the local or partial neurons of the BP neural network are destroyed. It will have a big impact.

When training, BP neural network can automatically extract the "reasonable rules" between output and output data through learning, adaptively memorize the learning content in the weight of the network, and apply it to new knowledge.

We skillfully determine the source of opioid and heroin events by screening the associations of the counties, and then establish a communication model to study the trend of dissemination.

The genetic algorithm we use does not have much mathematical requirements for the optimization problem solved. Due to its evolutionary characteristics, the intrinsic nature of the problem is not required in the search process.

5.2 Weakness

Under normal circumstances, when BP neural network training ability is poor, the prediction ability is also poor, and to some extent, as the training ability is improved, the prediction ability will be improved. However, this trend is not fixed. It has a limit. When this limit is reached, as the training ability increases, the predictive ability will decrease, that is, the so-called "over-fitting" phenomenon occurs.

When establishing a regression equation, which factor to use and which expression to use for the factor is only a speculation, which affects the diversity of the factors used and the immeasurability of certain factors.

Due to the excessive amount of data processing, only the most relevant data was selected when establishing the impact model of the problem-propagation model and the problem-based economic data on the number of drug identifications. In reality, the models we established maybe are not useful in the real living environment.

References

[1] Li De. Big Data Construction under the American Drug Abuse Crisis[J]. People's Rule of Law, 2018(Z1):159-160.

[2] Jiang Qiyuan& Xie Jinxing,&Ye Jun. Mathematical Models. 4th Edition [M]. Higher Education Press, 2011.

[3] Si Shoukui&Sun Yujing. Mathematical Modeling Algorithm and Application[M]. National Defence Industry Press, 2011.

[4] Sun Fangfang. Discussion on Grey Correlation Analysis Method and Its Application[J]. Science and Technology Information, 2010, 2(17): 886-888.

[5] Yang Yuhui& Xu Xiuli&Zhu Zhu. Overview of Opioid Abuse and Its Governance in the United States [J]. Chinese Medical Alert, 2017.

[6] Bao Yanping&Wang Tongyu& Wang Ziyun, et al. Analysis of heroin abuse characteristics and related factors in synthetic drug abusers in five regions of China [J]. Chinese Journal of Drug Dependence, 2015(6): 450-460.

[7] Zhao Zeyun,Wang Xuefeng.Research on Urban Traffic Flow Prediction Based on Neural Network[J].China New Telecommunication , 2019 , 21(04) : 59.