For office use only	Team Control Number	For office use only
T1 ________	**1901553**	F1 ________
T2 ________		F2 ________
T3 ________	Problem Chosen	F3 ________
T4 ________	**C**	F4 ________

2019
MCM/ICM
Summary Sheet

Analysis of complex drug diffusion phenomenon based on propagation dynamics

Summary

In response to the abuse and proliferation of synthetic opioids in the United States, we propose an innovative model and analyzed the data strategically.

In the first part, based on the characteristics of drug reports in the five states from 2010 to 2017, we establish a drug diffusion model based on the SIS propoation model. We use the entropy weight method to quantify the characteristics of direction of drug diffusion in and between the states. Further more, we use BP neural network to predict the origin of the drug and regions in which the number of drugs may exceed the threshold in the future.

Based on the simulation, we reach a conclusion which indicates the origins are*Lexington, Parkersburg, Cleveland, Laureldale, Bedford*, and the areas that may exceed the threshold in **2021** are *Cincinnati, Frankfort, Waston, Cleveland, Lynchburg*.

In the second part, we propose a model based on grey correlation analysis, Spearman correlation test and entropy weight method, which successfully excluded the not or less related influence factors to the conclusion. Based on the analysis, we modify our model in part 1 and significantly improve the stablity of SIS model and the accuracy of BP neural network model.

After the correction of the original model, the new experimental results show that there are eight factors having the greatest correlation with the number of drugs used including: *education attainment-population 25 years* and *over professional degree*, *relationship-Householder* and so on.

In the third part, we use the improved model and known data in Part 2 to simulate the year and region that exceed the threshold. According to the threshold, we propose corresponding prevention and control recommendations to the government. We strongly recommend government to make attention to specific county where the number of drug abuse incidents reaches the threshold, and use strong policies to restrain the drug abuse.

Keywords: **Drug Abuse, SIS Propagation Dynamics Model, Correlation Analysis, Entropy Weight Method, BP Neural Network**

Contents

1 Introduction

1.1 Problem Statement

In some parts of the United States, the abuse of synthetic and non-synthetic opioids has formed a new social crisis, and this abuse has shown a trend of spreading from local to periphery. Opioids have an impact on people of different levels and different age groups. If the number of opioid addicts continues to rise, the development of society will be greatly affected in the long run.

The NFLIS publishes an annual data report which covers every state from the crime lab. The report contains information on drug use in Ohio, Kentucky, West Virginia, Virginia and Tennessee counties, as well as data on the region's economy and population age distribution. On this basis, you can also collect the specific geographic coordinates of each county in the five states to increase the number of available data.

For the first part of the question, we need to create an appropriate model based on the data provided to find out the source of opioid diffusion , the spread trend of opioids , the threshold of opioids and predict the future changes in opioids.

For the problem in Part 2, we need to get additional information from the socioeconomic data of the given states and counties to make corresponding improvements to the model to make the results more accurate.

For part 3, we need to propose an effective strategy for opioid suppression and test the effectiveness of the strategy for U.S. Policy.

1.2 Analysis and Approach Overview

1.2.1 Global Analysis

From the analysis, it can be seen that:

Synthetic opioids within the state are spread out through counties with neighboring relationships, so it can be determined that the number of synthetic opioids between counties and counties within the state interacts with each other, therefore, when we construct the mathematical model, we can not exclude the interaction between independent individuals (counties and counties).

Synthetic opioids between states are spread through a more integrated state to a less integrated state, so it can be determined that changes in the number of synthetic opioids between states are affected by a centrosome.

In summary, we compare the propagation process of synthetic opioids with the propagation process of SARS virus[4], and choose to use SIS propagation dynamics model as the main line throughout the modeling process using BP neural network[3], entropy weight method, gray correlation analysis method. Finally , the traditional SIS model is improved to complete modeling and optimization.

1.2.2 First Part

Model Establishment: Establishing SIS dynamics propagation model

Set indicator H, which is defined as

$$H = \alpha_1 I + \alpha_2 F \tag{1}$$

which I denotes synthetic opium number of current year, and F denotes rate of change of synthetic opium this year. Set indicator h as the growth rate of H.

Calculate the diffusion trend: Calculate the trend of H indirectly to show the diffusion trend of synthetic opioids. At the same time, by using the location information (latitude and longitude) of a certain region [1], the calculation results are projected onto the map to reflect this trend, and the results are objectively displayed.

Find the source: Use the BP neural network to train the latitude and longitude coordinates of each county within the state, and the number of synthetic opiums in the current year, and push back to the source.

Finding where to happen in the future: Using BP neural network to train the latitude and longitude coordinates of each county in the state and the number of synthetic opioids in the current year, to find future data changes, and judge whether the occurrence is exceeded or not.

Find the threshold: the critical value of H, the critical value of h.

1.2.3 Second Part

Model Establishment: Establishing Grey Correlation Analysis Method Variable Correlation Analysis Model.

Model Solver: Regarding which factor of socioeconomic data affects the trend of diffusion, we use gray correlation analysis to screen out the factors that have a greater impact on synthetic and non-synthetic opioid diffusion from known political and economic data, and use entropy[2][5]. The power method finds the weight of each factor's influence. Furthermore, these factors with greater influence are put into the SIS model, and BP neural network is used to predict the spread of events in 2018 and beyond. Finally, it is concluded who is using drugs, which leads to the growth of drug use and addiction.

2 Assumptions and Notations

2.1 Assumptions

In the first part, we have the following assumptions to simplify the model:

- Does not consider the impact of population immigration between states and counties, and the impact of population births and deaths on the spread of drug abuse incidents.

- Assume that the governance capacity of each county is the same, regardless of the impact of the adjacent area between the county and the county on the ability to spread.
- The total number of incidents of drug abuse in the county is divided into drug abuse incidents and potential drug abuse incidents.
- The reduction in the number of drug abuse incidents only considers the reasons for government regulation and does not consider other factors.

In the second part, the following assumptions are made:

- In the gray correlation analysis method, the weight W_k is set to the commonly used 0.5.
- The factors that contain less information are negligible, otherwise it may lead to over-fitting in the model.
- In the process of improving the model, only those factors that are highly correlated with the growth of drug abuse events are added to the model, and the less relevant factors are ignored.

2.2 Notations

Table 1: All Notations

Notation	Meaning
I	The proportion of drug incidents that have occurred
S	Potential drug incidents accounted for
F	Growth rate of I
β	The probability of unit event propagation
	The probability of governance within unit time (year)
H	Indicator for measuring the severity of the spread of synthetic and non-synthetic opioid events in the region
a_1	H in the weight of I
a_2	H in the weight of I
h	Growth rate of H
T_j	Information Entropy of Entropy Weight Indicators
P_{ij}	The intermediate variable of information entropy calculated by entropy weight method
M	Entropy weight method index weights
X	BP neural network input

Y	BP neural network output
W_{ij}	BP neural network from $l-1$ layer j neurons and l layer number i neurons between the connection weights
$b_i^{(l)}$	BP neural network number l layer number i neuron offset
$net_i^{(l)}$	BP Neural Network l Layer i Input of Neurons
$E(i)$	Training error of a single sample of BP neural network
$X_0^{'}$	Grey Correlation Analysis Reference Data Column
$\lvert x_0(k)-x_i(k)\rvert$	Grey correlation analysis index sequence
$\min\limits_{i=1}^{n}\min\limits_{k=1}^{m}\lvert x_0(k)-x_i(k)\rvert$	Gray correlation analysis index sequence maximum value
$\max\limits_{i=1}^{n}\max\limits_{k=1}^{m}\lvert x_0(k)-x_i(k)\rvert$	Gray correlation analysis index sequence minimum value
$\zeta_i(k)$	Correlation coefficient of grey relational analysis method
r_{0i}	Correlation sequence of grey relational analysis
$r_{0i}^{'}$	The gray relational analysis weighted average correlation sequence accounted for a drug incident of I

3 Model Construction

3.1 Part I: Find out the source of opioid diffusion, the spread trend of opioids, the threshold of opioids and predict the future changes in opioids.

3.1.1 Entropy weight method to measure the severity of diffusion

The basic idea of the entropy weight method is to determine the objective weight according to the size of the index variability. Generally speaking, if the information entropy of an index is smaller, it indicates that the degree of variation of the index value is larger, and the amount of information provided. The more you play, the more you can play in the comprehensive evaluation, and the greater the weight. On the contrary, the larger the information entropy of an indicator is $E_j$which indicates that the smaller the index is worth to mutate, the less information is provided, and the smaller the role played in the comprehensive evaluation, the more weight it has small.

Set indicator H, which is defined as

$$H = \alpha_1 I + \alpha_2 F \tag{2}$$

which I denotes synthetic opium number of current year, and F denotes rate of change of synthetic opium this year.

Data standardization
We first standardize the data of each indicator. Suppose that the K indicator is given by $X_1, X_2, ...X_K$, where $X_i = x_i, x_2, ...x_n$. Suppose the standardized values for each indicator data are $Y_1, Y_2, ...Y_k$, then

$$Y_{ij} = \frac{X_{ij} - min(X_i)}{max(X_i) - max(X_j)} \tag{3}$$

Find the information entropy of each indicator
The information entropy of the data is:

$$T_j = -ln(n)^{-1}\sum_{i=1}^{N} P_{ij}lnp_{ij} \tag{4}$$

Where $P_{ij} = \frac{Y_{ij}}{\sum\limits_{i=1}^{n} Y_{ij}}$

If $P_{ij} = 0$, define $\lim\limits_{P_{ij}\to+0} P_{ij}lnp_{ij} = 0$.

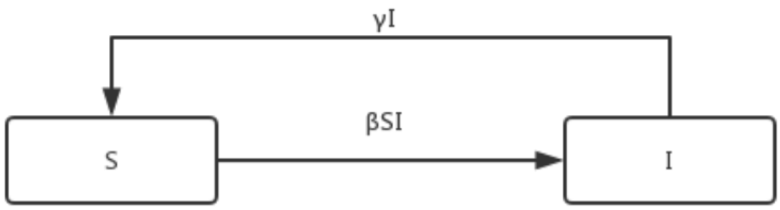
Weight of each indicator
According to the calculation formula of information entropy, the information entropy of each indicator is calculated as $T_1, T_2, ...T_K$. The weights of each indicator are calculated by information entropy as:

$$M = \frac{1-T_I}{k-\sum T_i}(i = 1, 2) \tag{5}$$

In summary, we can get the weight of both a_1, a_2.

3.1.2 Improved SIS model

The SIS propagation kinetic model can visually depict the virus diffusion process. We compare the diffusion of viruses and the diffusion process of synthetic opioids. The variables used in the model are divided into susceptible populations and infected populations, and the parameters of the model are improved to adapt to the proliferation of synthetic opioids between counties and states. The relevant assumptions have been stated, and the model is shown below:



SIS Model

According to the above model:

$$\frac{dI}{dt} = I(\beta S - Y) \tag{6}$$

$$\frac{dS}{dt} = -I(\beta S - Y) \tag{7}$$

Then when $\beta S - Y > 0$, the number of occurrences of the event begins to increase as part of the measure of the size of H.

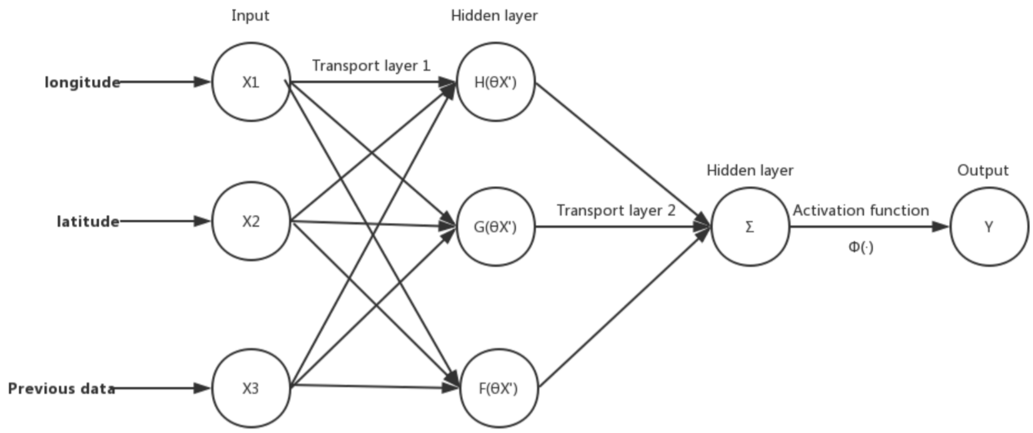
Based on the previous analysis, when we know that H is large enough, it indicates that the number of drug incidents in the region is large and the growth rate is large, and the spread is more serious. The government should implement relevant measures in the region to curb the occurrence of drug incidents.

3.1.3 BP Neural Network

BP neural network algorithm is one of the most effective multi-layer neural network learning methods. Its main features are signal forward transmission and error backward propagation. By continuously adjusting the network weight value, the final output of the network is as close as possible to the expected output. To achieve the purpose of training.

In the improved SIS model, we use BP neural network to train historical data (such as the latitude and longitude of a place, the proficiency of synthetic opioids this year), and then record the characteristics of all the data. To ensure that the correlation between counties is not excluded, but also to complete the prediction of future data and the backstepping of the birthplace.

The BP neural network diagram shows:



Operational process
Input vector:

$$X = [x_1, x_2, x_i x_m], i = 1, 2, , m \tag{8}$$

Output vector:

$$Y = [y_1, y_2, y_k y_n], k = 1, 2, , n \tag{9}$$

The output of each neuron in the i layer is:

$$h^{(l)} = [h_1^{(l)}, h_2^{(l)} h_j^{(l)} h_s^{(sl)}], j = 1, 2, , sl \tag{10}$$

sl is the number of l layer neurons

$$h_i^{(l)} = f(net_i^{(l)}) \tag{11}$$

$$net_i^{(l)} = \sum_{j=1}^{sl-1} W_{ij}^{(l)} h_j^{(l-1)} + b_i^{(l)} \tag{12}$$

Let W_{ij}^{i} be between the j neurons in the $l-1$ layer and the i neurons in the l layer Connection weight. $b_i^{(l)}$ is the offset of the i neurons in the l layer, then

$$H_i^{(l)} = f(net_i^{(l)}) \tag{13}$$

$$Net_i^{(l)} = \sum_{j=1}^{sl-1} W_{ij}^{(l)} h_j^{(l-1)} + b_i^{(l)} \tag{14}$$

Among them, $net_i^{(l)}$ is l layer i neuron input, $f()$ is the activation function of the neuron. The nonlinear activation function is usually used in multi-layer neural networks instead of the linear activation function, because the multi-layer neural network based on the linear activation function is essentially a superposition of multiple linear functions, and the result is still a linear function.

Activation function

Nonlinear activation function used by BP neural network:

$$f(x) = \frac{1}{1+e^{-x}} \tag{15}$$

$$f(x) = \frac{1-e^{-x}}{1+e^{-x}} \tag{16}$$

BP Algorith

Given a training sample of m, the error function is:

$$E = \frac{1}{m} \sum_{i=1}^{m} E(i) \tag{17}$$

Among them, $E(i)$ is the training error of a single sample:

$$E(i) = \frac{1}{2} \sum_{k=1}^{n} (d_k(i) - y_k(i))^2 \tag{18}$$

So,

$$E = \frac{1}{2m} \sum_{i=1}^{m} \sum_{k=1}^{n} (d_k(i) - y_k(i))^2 \tag{19}$$

The BP algorithm updates the weights and offsets in each iteration as follows:

$$W_{ij}^{(l)} = W_{ij}^{(l)} - \alpha \frac{\partial E}{\partial W_{ij}^{(l)}} \tag{20}$$

$$b_i^{(l)} = b_i^{(l)} - \alpha \frac{\partial E}{\partial b_i^{(l)}} \tag{21}$$

Among them, α is the learning rate, and its value range is (0, 1).

3.2 Part II: get additional information to make corresponding improvements to the model to make the results more accurate.

Grey relational analysis refers to the method of quantitative description and comparison of a system development and change situation. The basic idea is to determine whether the connection is tight by determining the geometric similarity of the reference data column and several comparison data columns. It reflects the curve. The degree of association between them.

This method can usually be used to analyze the degree of influence of each factor on the result. It can also be used to solve the comprehensive evaluation problem with time. The core of the problem is to establish the parent sequence with time according to certain rules, and to evaluate each object over time. The change is taken as a subsequence, and the degree of correlation between each subsequence and the parent sequence is obtained, and the conclusion is drawn according to the correlation size.

Determination of analytical index system
Let n data sequences form the following matrix:

$$(X_1', X_2', ..., X_n') = \begin{bmatrix} x_1'(1) & x_2'(1) & ..... & x_n'(1) \\ x_1'(2) & x_2'(2) & ..... & x_n'(2) \\ .... & ..... & ..... & ..... \\ x_1'(m) & x_2'(m) & ..... & x_n'(m) \end{bmatrix} \tag{22}$$

Where m is the number of metrics, $X_i' = (x_i'(1), x_i'(2), ..., x_i'(m))^T, i = 1, 2, ..., n$

Determine the reference data column
The reference data column should be an ideal comparison standard, which can be composed of the optimal value (or the worst value) of each index, or other reference values can be selected according to the evaluation purpose. Recorded as

$$X_0' = x_0'(1), x_0'(2), ..., x_0'(m)) \tag{23}$$

Determine the absolute difference between the evaluated object and the reference column So,

$$|x_0(k) - x_i(k)|(k = 1, 2, ...m.i = 1, 2, ...n) \tag{24}$$

where n is the number of the object being evaluated.

Determine minimum and maximum

$$\min_{i=1}^{n} \min_{k=1}^{m} |x_0(k) - x_i(k)| \tag{25}$$

$$\max_{i=1}^{n} \max_{k=1}^{m} |x_0(k) - x_i(k)| \tag{26}$$

Calculate the correlation coefficient between the comparison column and the corresponding element of the reference column

$$\zeta_i(k) = \frac{\min_{i=1}^{n} \min_{k=1}^{m} |x_0'(k) - x_i'(k)| + \max_{i=1}^{n} \max_{k=1}^{m} |x_0'(k) - x_i'(k)|}{|x_0'(k) - x_i'(k)| + \max_{i=1}^{n} \max_{k=1}^{m} |x_0'(k) - x_i'(k)|} \tag{27}$$

Where ρ is the resolution coefficient, $0 < \rho < 1$. The smaller the ρ is, the greater the difference between the correlation coefficients is, and the stronger the discrimination ability is. Usually ρ take 0.5.

Computational order
For each evaluation object (comparison sequence), the mean value of the correlation coefficient between the respective indicators and the corresponding elements of the reference sequence is calculated to reflect the relationship between each evaluation object and the reference sequence, and is referred to as an associative sequence, which is recorded as:

$$r_{0i} = \frac{1}{m} \sum_{k=1}^{m} \zeta_i(k) \tag{28}$$

Weighted mean of correlation coefficients
The weighted mean of the correlation coefficients is:

$$r_{0i}^{'} = \frac{1}{m}\sum_{k=1}^{m} W_k \zeta_i(k) \tag{29}$$

Where W_k is the weight of each indicator.

4 Model

Based on the improved SIS model, we use H value as a measure of drug abuse in the region, and h value as a measure of the extent of diffusion in the region. Through the entropy weight method, we can find the weights corresponding to I and F, which are 0.73 and 30.63. then:

$$H = 0.73I + 30.63F \tag{30}$$

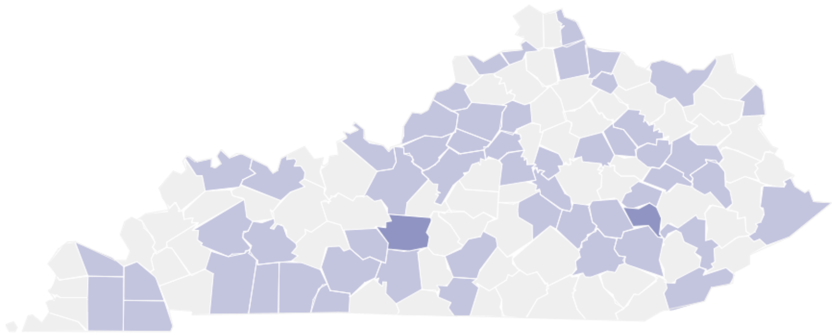
Thus, we can obtain the corresponding H and h values of each county from 2010 to 2017. The H values and h values of the counties in the five states were plotted by python. The higher the H value, the more serious the drug abuses in the region. The higher the h value, the more serious the region is being affected by the spread.

For future related predictions, due to the excessive number of counties and the possible influence between non-adjacent counties, we simplify the situation as the independent variable of each county, that is, the longitude and latitude. Independent variable. Since the number of drug incidents in the second year was affected by the previous year, the number of drug incidents in the previous year was used as the dependent variable. And because there is a nonlinear relationship between the independent variable and the dependent variable in time, we choose to use BP neural network to predict the change trend of H value and h value by training the historical data, and based on BP neural network. The value of the value is also very good to avoid the inaccuracy caused by linear programming.

4.1 Part 1

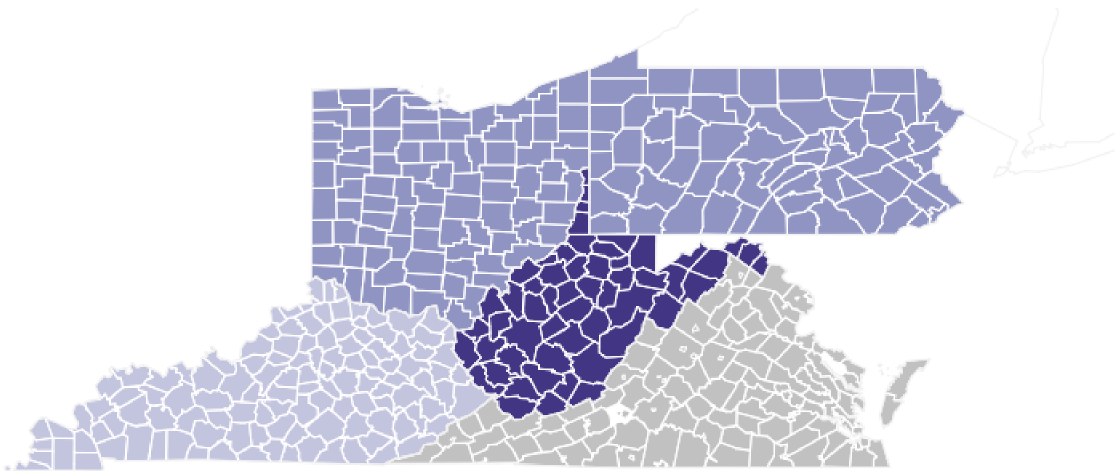
4.1.1 Describe the situation of diffusion

We first describe the case of the spread of the state of Kentucky as an example. The map of the state's h value for 2010-2011 is as follows:



It can be seen that the dark county in the picture is being spread by drug abuse, and the direction of spread is spread to the surrounding. Based on the 2010-2017 data, we can see that in West Virginia, Parkersburg, Clarksburg, and Uniontown spread around. in Ohio, spread by columbus and cleveland. in Kentucky, by Louisville, Lexington Cincinnati spreads around. in Pennsylvania, spread by philadelphia, laureldale, and York. in Virginia, spread by richmond, roankoke, and Bedford.

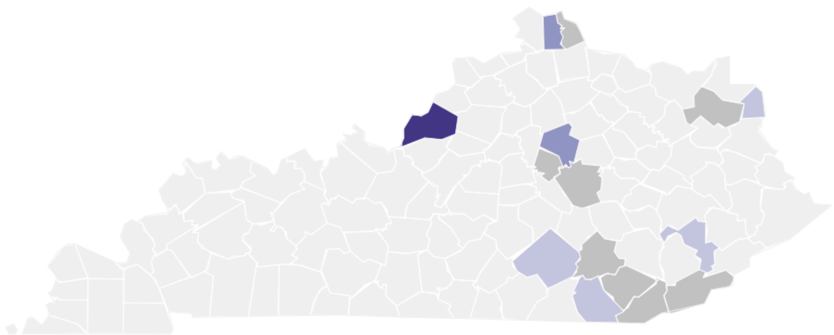
Considering the overall diffusion of the five states, the overall 2010-2011 h-value map is as follows:



It can be seen that the direction of diffusion between states is from a state with a light color to a state with a deep color. Based on all data for 2010-2017, drug abuse 2010-2011 was spread from Virginia to West Virginia, spread from West Virginia to Ohio and Virginia in 2011-2012, and the same in 2012-2013 and the previous year. , 2013-2014 was spread by West Virginia to Ohio and Pennsylvania, 2014-2015 was spread by Kentucky and West Virginia to the remaining three states, and 2015-2016 was spread by Kentucky to Ohio and Virginia The state, 2016-2017 was spread by West Virginia to the remaining four states.

4.1.2 Speculative source of diffusion

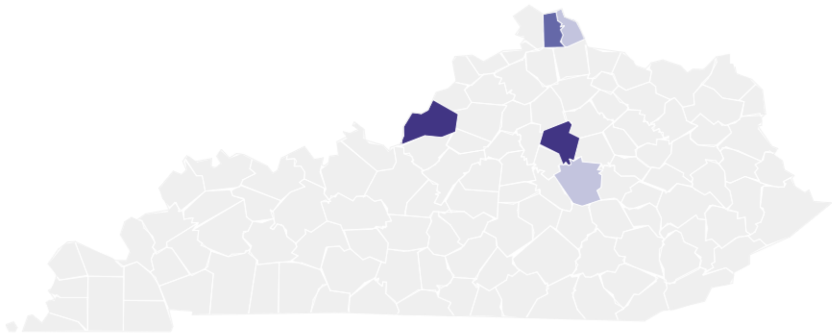
Through the BP neural network, we take the chronological order as an independent variable into the model for calculation, and speculate on the situation after 2010. When most of the counties have smaller H values and smaller h values, but some counties have smaller H values and larger h values, it can be concluded that these areas are the origin points, and no longer move forward. Speculated. We make a map of the corresponding h-values, as in the case of Kentucky:



Then we can learn that lexington in Kentucky is the source of proliferation. In the same way, we can see that in West Virginia, Parkersburg is the source of proliferation. in Ohio, cleveland is the source of proliferation. in Pennsylvania, laureldale is the source of proliferation. in Virginia, Bedford is the source of proliferation.

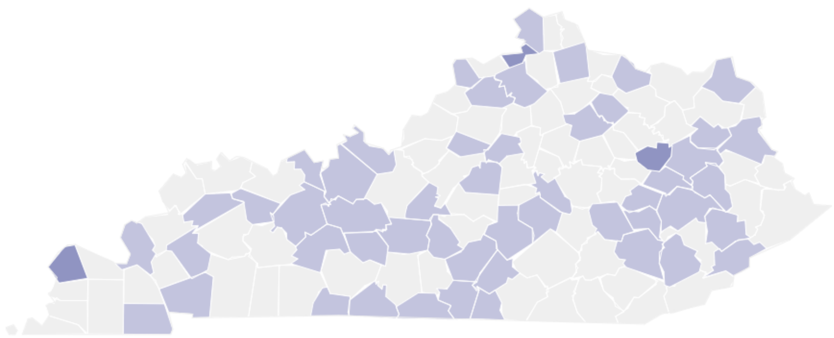
4.1.3 Predicting future situations and recommendations

According to the data, it can be known that the number of drug abuse incidents in counties with a large number of drug incidents reaches a certain value, and then starts to decrease from a higher level. It can be seen from the data that Nanticoke and other counties have started to decrease after reaching a certain higher value, and the data can be found by referring to the data. Take the mean of these data and find 201.3. If the value is greater than 201.3, it is recommended that the government implement relevant measures in the area to curb drug abuse. The BP forecast can be used to predict the state of Kentucky in 2020 as shown in the following figure:



Therefore, it is necessary to warn that cincinnati and frankfort counties in Kentucky need to control drug abuse. The same method can be found in Washington, West Virginia, in Cleveland, Ohio, in Pennsylvania, Pennsylvania, in Lynchburg, Virginia, where drug abuse control is required.

The h values of the counties in the five states are plotted, and the higher the color, the more severe the impact is affected by the diffusion, so that it can be seen from the figure which areas are diffusing areas. For example, the BP neural network predicts the spread of Kentucky from 2019 to 2020 as shown in the following figure:



From the data we can know that when h is greater than 87.6, the H value of the region has a rapid increase in degrees. Therefore, we set 87.6 as the threshold. When $h > 87.6$, it indicates that the region is receiving the spread of drug abuse. It is necessary to suggest that the government implement corresponding measures to prevent its spread. The proliferation of the rest of the states is similar to that of Kentucky, so we can draw regional conclusions from the data that need prevention.

According to the data, it is necessary to warn that cincinnati and frankfort counties in Kentucky need to control drug abuse. The same method can be found in Washington, West Virginia, in Cleveland, Ohio, in York, Pennsylvania, in Lynchburg, Virginia, to prevent drug spread in 2020.

4.2 Part 2

4.2.1 Choosing factors with more information

Through the analysis of the second part of the data, we can know that there are more than a dozen political and economic factors in total, but the types of factors investigated each year are not the same. Therefore, in order to make the analysis of relevant trends more scientific, we only take the same factors every year, and through analysis, we know that there are 12 kinds of the same. Since some values in the data are either lossy or unavailable, we preprocess and filter the data to arrive at the available data.

Through the gray correlation analysis method, the order of the influence of these 12 factors on the H value and the h value is obtained. The weighted mean of the correlation coefficient of each factor is as follows:

$$0.78 > 0.72 > 0.61 > 0.30 > 0.22 > 0.20 > 0.19 > 0.17 > 0.11 > 0.10 > 0.09 > 0.06$$

The characteristics of the sequence are that the first three coefficients are significantly larger than the values of the latter nine coefficients, indicating that the information contained in these three factors is larger than the latter nine factors. It is assumed that we only consider the three factors that have the greatest impact.

4.2.2 Judging the population with serious drug abuse

We extracted the data from the seventeen categories below the three factors obtained above, and tested the correlation between the number of people and the number of events by the spearman correlation. By calculation, the spearman correlation coefficients are: -0.82, 0.12, 0.06, 0.76, -0.45, 0.23, 0.22, 0.11, 0.02, 0.04, 0.06, 0.76, 0.48, 0.91, -0.85, 0.67, -0.68.

So we can learn about relationship - Population in households - Nonrelatives, Other relatives , educational attainment - Population 25 years and over - Less than 9th grade ,place of birth - Total population - Native - Born in Puerto Rico, US Island areas, or born Abroad to American parent(s) There are four types of people who are positively related to the number of incidents, so they are predicted to be more serious in drug abuse. Educational attainment - Population 25 years and over - Associate's degree, Graduate or professional degree, relationship - Population in households - Householder, relatives, a total of four categories of people and the number of events are negatively correlated, predicting that they hardly abuse drugs.

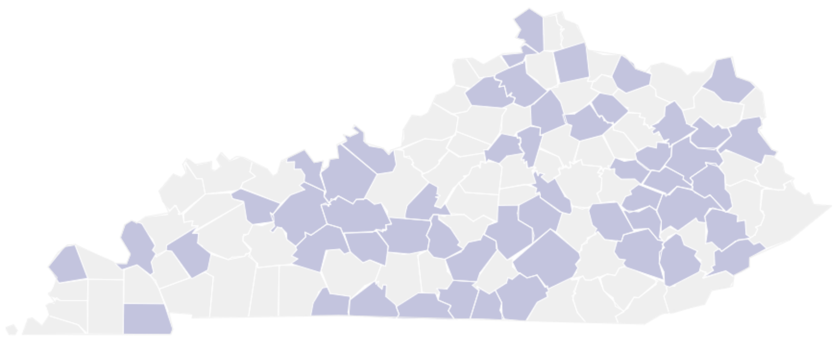
4.2.3 The improved model

From the above analysis, we can only analyze the three factors with the most information in the sequence: relationship, place of birth, education attainment. Since the number of the above eight populations is highly correlated with the number of drug abuse incidents, the number of these populations is used as an independent variable, and the predicted values of H and h values of each region can be obtained through the BP neural network.

4.3 Part 3

The improved model in the second part combines the model of the first part with the factors that need to be considered in the second part, so we use this improved model to predict the diffusion of future opioids, and then give the corresponding strategy.

By analyzing the number of the above eight populations and the number of drug abuses at the same time, the threshold of H and h can be changed to 197.2 and 76.2, respectively. With the improved model, data of H values and h values of each county are obtained. The data is plotted as follows:



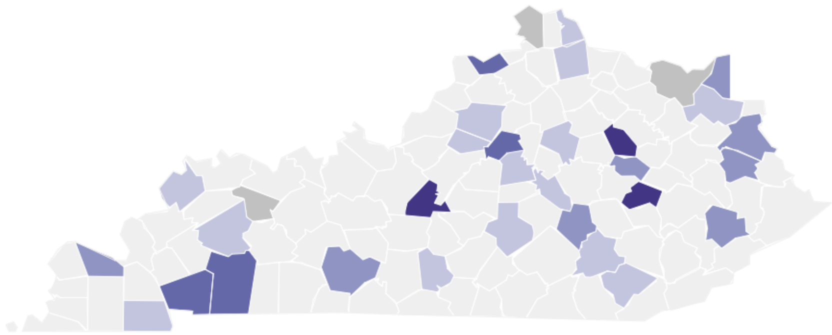
In West Virginia, spread by Parkersburg, Clarksburg, and Uniontown. in Ohio, spread by columbus, cleveland. in Kentucky, by Louisville, Lexington, Cincinati. in Pennsylvania, by philadelphia , laureldale, York spread around. in Virginia, spread by richmond, roankoke, Bedford.

According to the data, in 2021, the h value of uniontown in West Virginia, cleveland in Ohio, and richmond in Virginia reached a threshold, so these areas need prevention of drug abuse. in 2021, Laureldale, Pennsylvania, and The H value of cincinnati in Kentucky has reached a threshold, so these areas need to be controlled by drug abuse.

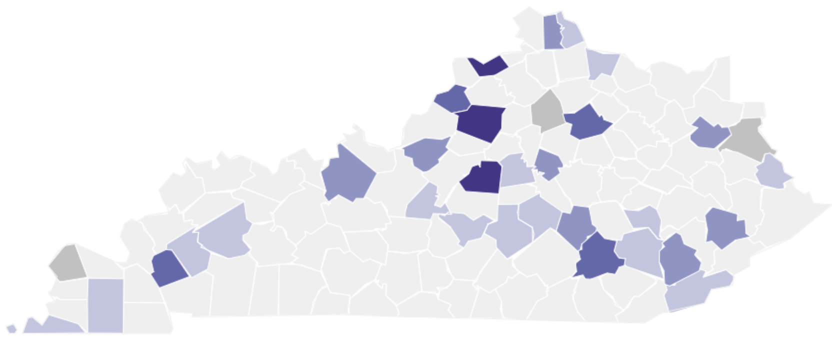
5 Sensitivity analysis

In the improved SIS model, we obtain the appropriate a_1, a_2 by the entropy weight method to find H, then a_1, a_2 have a greater impact on the results of subsequent predictions. Where a_1 is the weight of I and a_2 is the weight of F. Considering the effect of a_1, a_2 alone change on H is meaningless, because the role of a_1, a_2 is to balance the amount of information containing H and I in H. So we need to look at the effect of the size relationship between a_1, a_2 on the prediction. It can be seen from the data that the value of I is much larger than the value of F. The larger a_1/a_2, the greater the impact of H by I, so H contains more information about I and less F. Therefore, in the sensitivity analysis, we consider how the influence of a_1/a_2 changes on the prediction results. In order to simplify the impact, we may wish to assume that the value of a_1 is unchanged.

The value of a_1/a_2 obtained by the entropy weight method is about 0.0243. We may take a_1/a_2 as a smaller value, for example, 0.01, then a_2 is 73. The plots of the five states predicted by the BP neural network are as follows:



Taking a_1/a_2 to a larger value, for example 0.1, a_2 is 7.3. The prediction results obtained by BP neural network are shown in the following figure:



The prediction results of the five states are as follows:

It can be seen from the figure that the prediction result after changing a_1/a_2 is larger than that of the previous prediction, so we can know that the prediction result is sensitive to the change of a_1/a_2.

6 Strengths and weaknesses

6.1 Strengths

- The model established by our team is centered on the SIS propagation dynamic model, supplemented by BP neural network, grey relational analysis, and entropy weight method, which makes up for the defect that the basic central model cannot predict future data and cannot describe the drug diffusion trend.
- The model established by our team takes into account the correlation of data between counties, that is, the characteristics of all data can be recorded when using neural networks for training, ensuring that data correlation between counties and counties is not excluded.
- The model established by our team uses the gray correlation analysis method, which fully considers the correlation between the independent variables, and also reduces the pressure for the second part of the data analysis.
- The model established by our team successfully passed the test of sensitive data analysis, that is, the reasonable value of the abnormal value, the resulting data results have no large fluctuations relative to the normal value, thus ensuring the correctness of the conclusion.

6.2 Weakness

- The model established by our team excludes external objective factors such as the size of the adjacent area and other factors in the hypothesis of the first part. Therefore, in the training of BP neural network, the lack of support of certain independent variables leads to a certain error between the predicted results and the ideal analysis results.
- The model established by our team In the hypothesis of the second part, the gray correlation analysis method is used to screen out several major factors that have a great influence on the diffusion, and then it is used as an independent variable to BP neural network for prediction, but in the calculation. In the correlation process, the mantissa of the floating point number is not limited, so there is a calculation error of the floating point number.

7 Memo

After a series of complex modeling and calculation processes, we present the conclusions and recommendations in this MEMO.

Firstly, our team established a comprehensive model based on SIS dynamics diffusion to simulate the spread of drug abuse within and between states, and to predict future diffusion scenarios through this model. Our model successfully passed the test of sensitivity analysis, so the model we established is stable and the conclusion is trustworthy.

Secondly, through the analysis of drug abuse data and the trend of its generated image migration, we can know the characteristics of some diffusion phenomena. For example, the origin is randomly scattered within five states, and each state has 1-2 origins, such as lexington in Kentucky. The origin of each county level has a tendency to spread around, and the phenomenon of drug abuse between states has a tendency to spread progressively. And when the number of events reaches a certain level, the speed of diffusion will be significantly accelerated.

Thirdly, for the spread of the future five states, our model predicts that some regions will reach the threshold in 2020 and need to manage the areas that reach the threshold.

Fourthly, through the data given by the census institutions and the established mathematical models, it can be known that people who are far away from home, those who lack higher education, those who have unrelated relationships at home, and those who are distant relatives, The increase in the number of these four categories will lead to an increase in the number of drug abuse incidents in the region, while those with higher education, those with doctoral degrees, those with independent home ownership, those over 25 years old and those with degrees An increase in the number of people will result in a reduction in the number of drug abuse incidents in the region. From this we can see who has a higher probability of drug abuse and who has a lower probability of drug abuse.

Finally, we combined the work of the first part and the second part, combined with the existing data and future forecast data, and made the following suggestions to the government:

1. When the number of drug abuse incidents reaches the threshold, strong control measures should be used to control the drug in time.

2. The increase in the number of statistics will lead to an increase in the number of four types of drug abuse incidents. When the number of these four types of people increases rapidly, the government should introduce measures to strengthen the prevention and control of drug abuse in the region.

3. The data supervision department shall, according to the data changes over the years, reasonably formulate effective thresholds in real time, and timely curb the excessive diffusion of drugs.

References

[1] 2017 U.S. Gazetteer Files, U.S. Census Bureau
https://www.census.gov/geo/maps-data/data/gazetteer2017.html

[2] Renjing Liu, XiHan Gao,etc. Uncertain multi-attribute decision making method based on grey entropy model and its weight and its application, 2018

[3] Yu Wang, Shanshan Chen, Xinchu Fu, Review and prospect of propagation dynamics models, 2018

[4] XiaoNa Han, etc.Application of infectious disease model in SARS, 2005prevention

[5] ZhiCong Fei, Research on Entropy Weight-AHP and Grey-Analytic Hierarchy Process, 2009

Appendices

- Source code of BP Neural Network:

```
longitude = [longitude1; longitude2; longitude3; longitude4;
        longitude5; longitude6; longitude7];
latitude = [latitude1; latitude2; latitude3; latitude4; latitude5;
        latitude6;latitude7];
poison_number_before = [poison_number_before1; poison_number_before2;
        poison_number_before3; poison_number_before4;
        poison_number_before5; poison_number_before6;
        poison_number_before7];
poison_number_now = [poison_number_now1;poison_number_now2;
        poison_number_now3; poison_number_now4;
        poison_number_now5; poison_number_now6;
        poison_number_now7];

p = [longitude'; latitude'; poison_number_before'];
t = [poison_number_now'];

[pn, inputStr] = mapminmax(p);
[tn, outputStr] = mapminmax(t);

net = newff(pn, tn, [3 7 2], {'purelin', 'logsig', 'purelin'});
net.trainParam.show = 10;
net.trainParam.epochs = 20000;

net.trainParam.lr = 0.05;
net.trainParam.goal = 0.5 * 10^(-6);
net.divideFcn = '';
net = train(net, pn, tn);
p_test = [longitude1';latitude1';poison_number_now7'];
[pn, inputStr] = mapminmax(p_test);
answer = sim(net, pn);

answer1 = mapminmax('reverse', answer, outputStr);

csvwrite('2018_poison_42.csv',answer1');
```

- Map of h of five states in 2017: