Opioid Crisis solutions based on linear Model

Xudong Yang

Hainan University, Danzhou, China

2456006733@qq.com

Keywords: Data Processing; Linear Model; Feature Extraction

Abstract: Opioid crisis has contributed to loads of damages in both life expectancy and economical development in U.S. In order to tackle this extremely urgent issue, we need clear purposes and direct efforts, which means the necessity to uncover influential factors.In this paper, we are encouraged to build suitable mathematical models with loads of statistical data provided and required processed, which includes drug reported amount and some economical indexes of five state. First, we establish a linear propagation model and get three kinds of equation: county-county, county-state and state-state, and we evaluated the possibility of some counties as sources from the county-state weight matrix. Then we use the value of the weight matrix to extract the main influencing features and perform a simple analysis.Finally, we come up with the conclusion including the possible start of pointed opioid use, and the key feature of addicted people.

1. Introduction

According to the actual needs, we do the following work:

For Part 1. First, we establish a linear propagation model and get three kinds of equation: county-county, county-state and state-state, and we evaluated the possibility of some counties as sources from the county-state weight matrix. Finally, we use different models for different states and make a 10-year prediction to get some threshold years. For Part 2. Based on features in the database, we build a linear model to describe the contributions to the events counts of every features after standardization. Then we use the value of the weight matrix to extract the main influencing features and perform a simple analysis. For Part 3. Based on the number of feature quantities, we couple model one to model two linearly. The coupling coefficient is the ratio of feature counts of the two models, and in order to control the stability of the fitted values, we assume that the sum of the coupling coefficients is one.

2. Related work

2.1 Overview of Our Work

In order to assist the government to analyze the situation objectively, we are encouraged to build suitable mathematical models for five states (with their FIPS number): Kentucky (21), Ohio (39), Pennsylvania (42), West Virginia (51) and Virginia (54), with loads of statistical data provided, which includes drug reported amount and some economical indexes.

At first, we use drug reported amount only to build models with following intentions:

•Reveal the spread pattern and locate the possible start of pointed opioid use of each states,

•Make reasonable prediction of drug identification threshold levels and make speculations of when and where it will happen

Then, we take socio-economic factors into consideration and perfect the model above to tackle issues as follows:

•Main factors that contributes to the growth in opioid use and addiction,

•Features of people who is likely to overdoes,

•Combine the results of two models to identify a possible strategy for countering this drug crisis.

  DOI: 10.25236/iclsmh.2019.033

2.2 Assumptions and Justifications

We make the following assumptions which enable us to perform modeling:

• Obviously, the spread of opioid crisis in different county of the five states is related to the number of drug reports in this county last year. We assume the mathematics model of the spread of opioid crisis in different counties of one state is first-order Markov model,

• Given data for 8 years, we treat the wave of spread coefficient in and between counties as infinitesimal element. In order to simplify the model, we decide to assume the weight(spread) matrix of the reported synthetic opioid and heroin incidents from 2010 to 2017 is stable.

We make the following assumptions about emergency evacuation process in this paper.Without using additional exits, the four main exits can satisfy the evacuation requirements.

3. Symbol Description

Table 1. Symbol Description

Symbol	Description
n	events count in a certain county at a certain year of a certain drug
N	events count in a certain state at a certain year of a certain drug
B	weight matrix of county-county model
W	weight matrix of county-state model
R	weight matrix of state-state model
Y_i	i-th element of the output column vector, the value of the drug flooding event that occurred in the county in that year
$X_{k,i}$ $\alpha_{k,i}$ N^{new}	represents the k-th feature represents the weight of this feature events count in a certain state at a certain year of a certain drug in coupling model
μ	coupling coefficient of spreading
v	coupling coefficient of socio-economic features
Acc	the accuracy of coupling model

4. Data Processing

4.1 Data Merging

Due to large scale of variable quantities, we should perform data merging according to the similarity and integrity of the information.

•We use FIPS numbers to represent for the state and county names and remove them from our database,

•We replace the names of drugs by opioid except heroin,

•In some counties, number of opioid abusers is near zero and that is fairly small amount of data compared with the total drug reported abusers. So those data will have little influence on our mathematics modeling. Therefore, we treat those data (the number of opioid abusers in some counties) as zero,

•We form a new dataset that includes all the U.S. Census socio-economic data provided. The row of the table is sorted by all of the county numbers and the year. And the column is the combination of all the statistical indicators. Finally, we complete a table of 3244*760, shown partly in Figure 1.

A	B	C	D	E	F	G	H	I	J	K	L	M
HC01_VC0	HC01_VC0	HC01_VC0	HC01_VC0	HC01_VC0	HC01_VC0	HC01_VC0	HC01_VC1	HC01_VC1	HC01_VC1	HC01_VC1	HC01_VC1	HC01_
7272	5004	8864.259	1945	4124	1506	377	160	64747.17	6211.707	81271.05	10894.46	822
7811	5526	8864.259	2363	4011	1519	476	222	64747.17	6211.707	81271.05	10894.46	822
8319	6148	8864.259	2771	4848	1891	267	168	64747.17	6211.707	81271.05	10894.46	822
3363	2408	8864.259	1001	2001	725	193	137	64747.17	6211.707	81271.05	10894.46	822
16727	11722	8864.259	4825	9189	3586	470	134	64747.17	6211.707	81271.05	10894.46	822
4338	3088	8864.259	1396	2426	872	93	81	64747.17	6211.707	81271.05	10894.46	822
10902	7575	8864.259	2900	5268	1724	705	281	64747.17	6211.707	81271.05	10894.46	822
41612	30519	8864.259	15658	24522	11987	1777	1027	64747.17	6211.707	81271.05	10894.46	822
8132	5548	8864.259	2508	3916	1689	280	115	64747.17	6211.707	81271.05	10894.46	822
19542	13036	8864.259	5080	9728	3224	798	385	64747.17	6211.707	81271.05	10894.46	822
10787	7543	8864.259	3299	5587	2066	437	258	64747.17	6211.707	81271.05	10894.46	822
3177	2244	8864.259	896	1725	631	172	80	64747.17	6211.707	81271.05	10894.46	822
5257	3338	8864.259	1185	2357	690	375	99	64747.17	6211.707	81271.05	10894.46	822
7474	5423	8864.259	1790	4303	1298	354	175	64747.17	6211.707	81271.05	10894.46	822
27266	20418	8864.259	9405	16106	6664	1269	798	64747.17	6211.707	81271.05	10894.46	822
5175	3603	8864.259	1423	2875	1025	168	76	64747.17	6211.707	81271.05	10894.46	822
5261	3418	8864.259	1176	2775	899	201	102	64747.17	6211.707	81271.05	10894.46	822
14844	9108	8864.259	3659	7474	2902	454	198	64747.17	6211.707	81271.05	10894.46	822
35300	22457	8864.259	9833	16917	6903	1297	573	64747.17	6211.707	81271.05	10894.46	822
2062	1510	8864.259	661	1203	536	106	43	64747.17	6211.707	81271.05	10894.46	822
4222	2984	8864.259	1266	1961	610	339	225	64747.17	6211.707	81271.05	10894.46	822
10577	7510	8864.259	2907	6180	2155	459	307	64747.17	6211.707	81271.05	10894.46	822
6012	4215	8864.259	1551	3491	1199	269	187	64747.17	6211.707	81271.05	10894.46	822
25923	18746	8864.259	10253	13324	6729	1138	511	64747.17	6211.707	81271.05	10894.46	822
14535	10365	8864.259	4491	7717	3019	808	404	64747.17	6211.707	81271.05	10894.46	822
6520	4492	8864.259	1646	3162	1232	446	90	64747.17	6211.707	81271.05	10894.46	822
4184	2847	8864.259	1172	2000	653	338	185	64747.17	6211.707	81271.05	10894.46	822
3749	2452	8864.259	905	1915	650	158	41	64747.17	6211.707	81271.05	10894.46	822
2660	1765	8864.259	681	1381	488	107	17	64747.17	6211.707	81271.05	10894.46	822

ACS

Figure 1: collection of economical statistical indicators for all counties and years

It is noticeable that the dimensions of data in the following table are not exactly the same, in order to facilitate the subsequent modeling process, we need to standardize the data in the table, which we will explain later in the process of modeling.

4.2 Fit Missing Values

In the dataset, missing values may cause some biases in final conclusion. Therefore, fit missing values, which means replacing missing data with substituted values, is of great necessity in eliminating error and obtaining more authentic results.

- Through preliminary observation on the raw dataset for part 1, we discover that people in some counties use synthetic opioid only. In order to guarantee the convenience of establishing our mathematics modeling, we assign the unknown number of heroin abusers is zero while the number of synthetic opioid abusers is provided.
- In the new table that we form for part 2, missing data is obviously unavoidable. To facilitate subsequent modeling process, we use the method of mean imputation, which means missing data of a county is replaced by the average of the eigenvalues of other counties. In this case, the average of the eigenvalues in the column is unchanged.

5. Model Building and Parameter Calculation

5.1 Spreading Model of County-County in State 21

Considering the relationship with other counties of the events that occurred in a certain county. We use symbol nNO.,typet,k to describe this count, which means the count of events of two types occurred in k-th county of No.X state at a certain time t(in years). For example,means the count of heroin events occurred in 1st county of No.21 State at the first year. Here we use labels from the data table for the year, state and county names.

We take the modeling process of the 21st state as an example.

We believe that the propagation process is first-order Markov, and the number of specific types of events occurring in a particular county in No.21 state is contributed by two parts. One is the number of events caused by the countys own downward year. And the other is the spread of the downward year from other counties.For the k-th county of heroin model, we have:

$$n_{t,k}^{21,h} = n_{t-1,k}^{21,h} \cdot B_k^{21,h} + \Sigma_{m \neq k} n_{t-1,m}^{21,h} \cdot B_m^{21,h} \quad (1)$$

The $\Sigma_{m \neq k} n_{t-1,m}^{21,h} \cdot B_m^{21,h}$ means the sum of the propagation effects of other counties, where m is the labels of counties according to the data table.

We are surprised to find that these two items can be combined as following:

$$n_{t,k}^{21,h} = \Sigma_{minlabels} n_{t-1,m}^{21,h} \cdot B_m^{21,h} \quad (2)$$

This can be written as a form in which a column vector is multiplied by a row vector:

$$n_{t,k}^{21,h} = (n_{t-1,1}^{21,h}, \quad \dots \quad , n_{t-1,120}^{21,h}) \cdot \begin{pmatrix} B_1^{21,h} \\ \dots \\ B_{120}^{21,h} \end{pmatrix} \quad (3)$$

The labels of k and m is from 1 to 120 cause there are 120 counties in No.21 state according to the data table.

Now we turn the left side of the equation into a column vector indexed by t. Due to the first-order Markov model, t in the left of the equation is from 2 to 8, and in the right it will be from 1 to 7 cause there's data of 8 years' in the data table. We assume the column vector B is stable since it's a short time period model. So we can generalize the change of the events counts belongs to a county with the year in one equation:

$$\begin{pmatrix} n_{2,k}^{21,h} \\ \cdots \\ n_{8,k}^{21,h} \end{pmatrix} = \begin{pmatrix} n_{1,1}^{21,h}, & \cdots & , n_{1,120}^{21,h} \\ \cdots & & \\ n_{7,1}^{21,h}, & \cdots & , n_{7,120}^{21,h} \end{pmatrix} \cdot \begin{pmatrix} B_1^{21,h} \\ \cdots \\ B_{120}^{21,h} \end{pmatrix} \tag{4}$$

This is a standard multiple linear regression model. We can fit the coefficient column vector B in MATLAB.

For each identified county, their independent variables are the same, that is, the matrix consisting of the events counts in all counties, that is, the second 7*120 matrix on the right side of the equation.So we can directly summarize the Markov model of inter-county transmission in one equation:

$$\begin{pmatrix} n_{2,1}^{21,h}, & \cdots & , n_{2,120}^{21,h}, \\ \cdots & & \\ n_{8,1}^{21,h}, & \cdots & , n_{8,120}^{21,h} \end{pmatrix} = \begin{pmatrix} n_{1,1}^{21,h}, & \cdots & , n_{1,120}^{21,h} \\ \cdots & & \\ n_{7,1}^{21,h}, & \cdots & , n_{7,120}^{21,h} \end{pmatrix} \cdot \begin{pmatrix} B_{11}^{21,h}, & \cdots & , B_{1,120}^{21,h} \\ \cdots & & \\ B_{120,1}^{21,h}, & \cdots & , B_{120,120}^{21,h} \end{pmatrix} \tag{5}$$

We can write it as following:

$$n21,h = n21,h \cdot B21 \tag{6}$$

Matrix B is called a weight matrix.The coefficients of the diagonal line are called the self-propagation coefficient, and the others are called the transfer coefficient.

We can similarly give other equations as following:

$$n21,o = n21,o \cdot B21 \tag{7}$$

$$n39,h = n39,h \cdot B39 \tag{8}$$

$$n39,o = n39,o \cdot B39 \tag{9}$$

$$n42,h = n42,h \cdot B42 \tag{10}$$

$$n42,o = n42,o \cdot B42 \tag{11}$$

$$n51,h = n51,h \cdot B51 \tag{12}$$

$$n51,o = n51,o \cdot B51 \tag{13}$$

$$n54,h = n54,h \cdot B54 \tag{14}$$

$$n54,o = n54,o \cdot B54 \tag{15}$$

Where 'o' means model for opioid.

We use the following steps to find the solution of the above weight matrix:

Step 1: Import the number of drug reports in each county of the five states in 2010-2016 as an independent variable and regard those data in 2011-2017 as the amount to be fitted.

Step 2: Select the first column of the quantity to be fitted as the index of the first county, and calculate the transfer column vector b1 of the county by multiple linear regression fitting.

Step 3: Assign the transfer column vector b1 to the first column of the transfer matrix.

Step 4: Update the amount to be fitted to the second column (second county) and subsequent indicators, and repeat steps 2 and 3.

Step 5: After the end of the loop, the transfer matrix b is obtained.

Based on the above results, we obtain the transfer matrix is a square matrix with a size equal to the number of counties and the data of the transfer matrix is sparse (there are many matrix elements with a value of 0 or close to 0). The matrix element bij reflects the degree of contribution from No. i county to No. j county. That is to say, if a row has more non-zero items or extreme values, then the county may be the main source/report of some areas. We collectively consider it as the transmission source of drugs, and the specific divisions are considered when building the state model later.

Thus, we draw RGB images of the transfer matrix and pick out the lines with more prominent or rich colors. Take State 21 as an example, the images of heroin and synthetic-opioid transfer matrix model are showed as following:

We select from the above two figures and take counties[8 19 56 105]and [34 48 60 100 118] accordingly for further fitting.

5.2 Spreading Model of County-State between State 21 and 54

Obviously the total number of events in the state should be contributed by the heroin model and the opioids model. Thus, we give the equation as following:

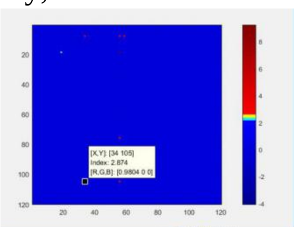
$$N_t^{21} = \Sigma n_{t-1,i}^{21,o} \cdot W_i + \Sigma n_{t-1,j}^{21,h} \cdot W_j \tag{16}$$

i, j respectively represent the labels of counties which may be as sources or destinations that picked up from the opioids and heroin models.

Combine the row vector W and write the equation as following:

$$N_t^{21} = [(\ldots \quad , n_{t-1,i}^{21,o}, \quad \ldots), (\ldots \quad , n_{t-1,j}^{21,h}, \quad \ldots)] \cdot \begin{pmatrix} W_1^{21} \\ \ldots \\ W_n^{21} \end{pmatrix} \tag{17}$$

Where n = i+j. Similarly, we can also write the left side of the upper equation as



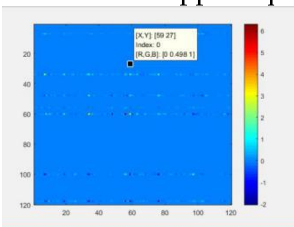


Figure 2 heroin transfer matrix Figure 3 opioid transfer matrix

a column of time-indexed column vector. Then the equation will be as following:

$$\begin{pmatrix} N_2^{21}, \\ \ldots \\ , N_8^{21} \end{pmatrix} = [\begin{pmatrix} \ldots & , n_{1,i}^{21,o}, & \ldots \\ \ldots & & \\ \ldots & n_{7,1}^{21,o} & \ldots \end{pmatrix}, \begin{pmatrix} \ldots & , n_{1,j}^{21,h}, & \ldots \\ \ldots & & \\ \ldots & n_{7,j}^{21,h} & \ldots \end{pmatrix}] \cdot \begin{pmatrix} W_1^{21}, \\ \ldots \\ , W_n^{21} \end{pmatrix} \tag{18}$$

We write it in the form of a matrix equation:

$$\mathrm{N}^{21} = [\boldsymbol{n}^{21,h}, \boldsymbol{n}^{21,o}] \cdot \mathrm{W}^{21} \tag{19}$$

Similarly, we can write equations for other states:

$$\mathrm{N}^{39} = [\boldsymbol{n}^{39,h}, \boldsymbol{n}^{39,o}] \cdot \mathrm{W}^{39} \tag{20}$$

N42 = [n42,h,n42,o] · W42 (21)

$$\mathrm{N}^{51} = [\boldsymbol{n}^{51,h}, \boldsymbol{n}^{51,o}] \cdot \boldsymbol{W}^{51} \tag{22}$$

$$\mathrm{N}^{54} = [\boldsymbol{n}^{54,h}, \boldsymbol{n}^{54,o}] \cdot \mathrm{W}^{54} \tag{23}$$

Note, however, that the equations for these states have only a similar form, and the values of i and j (that is, n) are not the same, and we only use these equations for No.39,No.42 and No.51 state to find the sources without using them for prediction.

$$\begin{pmatrix} N_2^{39}, & N_2^{42}, & N_2^{51} \\ \ldots & & \\ N_8^{39}, & N_8^{42}, & N_8^{51} \end{pmatrix} = \begin{pmatrix} N_1^{39}, & N_1^{42}, & N_1^{51} \\ \ldots & & \\ N_7^{39}, & N_7^{42}, & N_7^{51} \end{pmatrix} \cdot \begin{pmatrix} R_{11}, & R_{12}, & R_{13} \\ R_{21}, & R_{22}, & R_{33} \\ R_{31}, & R_{32}, & R_{33} \end{pmatrix} \tag{24}$$

5.3 Model Realization and Prediction

Since the weight matrix of the county-county model is too large, we won’t give specific contents here.

The calculated weight has ’0’ items, here we have removed ’0’ items.

Here we also give the W column vectors of other states cause we'll use it to judge the sources. For convenience, we put them in one matrix and fill the missing part with 0.

$$W_{39,42,51} = \begin{pmatrix} -3.9029, & -719.9887, & -2.8353 \\ -21.3472, & 47.6727, & -21.3543 \\ -86.3669, & 23.2320, & 27.2179 \\ 70.9532, & 134.7188, & -163.9073 \\ 75.0986, & -4.2847, & 173.1292 \\ 102.9821, & 9.2853, & -31.0489 \\ 50.6534, & -6.1744, & 8.9448 \end{pmatrix} \tag{25}$$

The state-state model in No.21 and No.54 state are given as following:

$$\begin{pmatrix} N_2^{39}, & N_2^{42}, & N_2^{51} \\ \dots & & \\ N_8^{39}, & N_8^{42}, & N_8^{51} \end{pmatrix} = \begin{pmatrix} N_1^{39}, & N_1^{42}, & N_1^{51} \\ \dots & & \\ N_7^{39}, & N_7^{42}, & N_7^{51} \end{pmatrix} \cdot \begin{pmatrix} 1.0123, & 0.022725, & 0.15229 \\ 0.14448, & 0.73746, & 0.27208 \\ -0.16381, & 0.44742, & -0.041184 \end{pmatrix} \tag{26}$$

According to the above two models, the predicted events counts in each state in the next decade is as following, the numbers in this table has been transformed into integer value:

From the county-state model we can get the possible source of heroin and opioids from each state.

• No.21 state Heroin Source: JEFFERSON county

No.21 state Opioids Source: FAYETTE, KNOTT county

Table 2. Predict For Next 10 Years

Year	**No.21**	**No.39**	**No.42**	**No.51**	**No.54**
2018	28870	124700	69965	35359	3672
2019	36722	130550	70250	36571	2061
2020	64621	136320	71136	37490	761
2021	81626	142140	72331	38572	243
2022	157310	148030	73829	39739	210
2023	216210	154010	75590	40995	5261
2024	356130	160120	77586	42334	381
2025	573840	176370	79796	43752	5801
2026	843250	172790	82203	45247	79
2027	1331200	179390	84792	46818	17412

•No.39 state Heroin Source: LORAIN, MONTGOMERY, STARK county

No.39 state Opioids Source: MONTGOMERY county • No.42 state Heroin Source: LEHIGH county No.42 state Opioids Source: none • No.51 state Heroin Source: none

No.51 state Opioids Source: PRINCE WILLIAM county • No.54 state Heroin Source: RALEIGH county

No.54 state Opioids Source: NICHOLAS county, MERCER county

Now let's make some judgements about the problem.We can see that events counts in No.21 state have a dramatic increase by 2020.The fluctuation of No.54 state is always very intense.The other three states is growing slowly.

•The possible threshold year for No.21 state is 2020, 2022, 2026, 2027 The possible threshold year for No.54 state is 2023, 2025, 2027 There's no possible threshold year for other states.

So what the US government needs to do is to strengthen the control of No.21 state in 2020 and beyond, including immigration management and local industry inspections.For No.54 state it is necessary to conduct the inspections and controls in the next year in which the report is less.

6. Correlation Model between Drag Abusing and Population Census

6.1 Build Model and Select Main Factors

We use LRM (linear regression model) to construct the relationship between the number of drug abuse incidents in a county and the county's current census characteristics, the equation is as followed:

$$Yi = \Sigma \alpha k,i \cdot Xk,i \quad (27)$$

In the above equation, Yi is the i-th element of the output column vector, which represents the value of the drug flooding event that occurred in the county in that year, Xk,i represents the k-th feature, and αk,i represents the weight (ratio) of this feature.

According to our assumption, we believe that the value of αk,i between different states and counties does not change with the year, because αk,i is determined by the nature of the feature itself: for example, we can objectively believe that the poor and the outside population are more prone to cause drug abuse cases which can be found in almost all time in all counties of the five states.

For this model, we use MATLAB for regression analysis to get the column vector of αk,i and find out the features corresponding to each element in αk,i. Then based on the distribution characteristics of the column vector of αk,i, we make further analysis to find out the main characteristics which cause the opioid crisis.

It is worth noting that we have mentioned above that the dimensions of the features in the data set obtained are not exactly the same, so we normalize each column value to eliminate the influence of the dimension, and the column vector for the output value does not have to be performed, because that only causes the order of magnitude change in the output value.

We use MATLAB function for normalization so that all values are in the [-1,1] range.

The specific process can be summarized as the following steps:

Step 1: Import the independent data set and the dependent variable data set

Step 2: Standardize each column of the argument data set to get a new argument data set

Step 3: Make an image of the weight change with the characteristics, and select the main factor that causes the drug crisis and is related to the absolute value of the weight.

Figure 4 shows an image of weight changes with features:

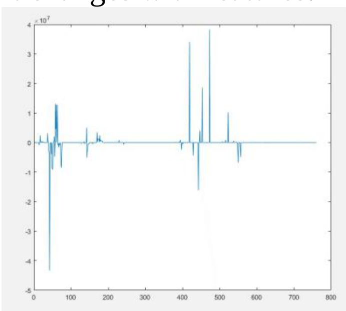


Figure 4: weight changes

We can see the obvious peaks and troughs from above picture, and their corresponding characteristics may be the main characteristics affecting the drug crisis. Therefore, we arrange the weights and their corresponding feature labels according to the weight size as an index. The feature is our main factor, and selects several feature tags corresponding to the largest absolute weight value (10th power of 10).

Table 3. Index Weight

index	42	43	58	60	63	419	443	453	473	523
weights(10^5)	-430	-190	131	127	128	340	-160	185	382	102

We classify the selected feature tags into two categories according to their weights and positives, representing their contribution nature.

Negative: 42 43 443

Positive: 58 60 63 419 453 473 523

Then, we find the corresponding feature and their name as follows:

```
In [16]: nega
Out[16]:
   HC01_VC134  HC01_VC135  HC03_VC157
0         127   1025.4236        20.5

In [17]: pos
Out[17]:
   HC01_VC151  HC01_VC154  HC01_VC157  HC03_VC131  HC03_VC167  HC03_VC186  \
0          56   5172.2755        26.0        75.1        96.3         0.0

   HC03_VC44
0       55.8
```

Figure 5: weight changes

The data in the above table is the main impact feature, of which the first three features are negatively correlated with the number of cases, and the rest are positive correlations.

6.2 Result Analysis

From the features we extracted, we have to believe without any inclination that the immigrant population after 2000, that is, mainly young people, especially those in Europe or Denmark, have a tendency to use/abuse narcotics. We attribute this to the possible stress of life. There is no doubt that excessive use of non-opioid drugs is dangerous, which may cause mental irreversible damage. On the other hand, given the relationship we have analyzed and the pressures of life, these events may be more likely to occur in economically developed and populous areas. So the trend of use is related to the socio-economic factors in the census.

7. Linear Coupling Model Used Model 1&2

7.1 Model Realization

Model that takes into account the influence of socio-economic features will be given as following:

$$N_k^{new} = \mu \cdot N_k + \nu \cdot \Sigma\alpha_i \cdot F_i \quad (28)$$

Whereμandvare the weights,Nkis the original prediction,Σαi ·Fimeans the distribution of socio-economic features.Here we use the size of features to weigh the μ and ν, that is

$$\mu + \nu = 1$$

$$\frac{\mu}{\nu} = 461/760 = 0.60658$$

We can get that ν = 0.62244,μ = 0.37756, so the equation will be given as following:

$$N_k^{new} = 0.37756 \cdot N_k + 0.62244 \cdot \Sigma\alpha_i \cdot F_i \quad (29)$$

Use values of each state in the last eight years for verification, the formula we use is given as following:

$$Acc = \frac{N_k^{new} - N_k^{real}}{N_k^{real}}$$

It can be seen that the error of the model is about 20% on average,so it is suitable model. The accuracy has been put in Table 4.

Table 4. Model Accuracy

Year	No.21	No.39	No.42	No.51	No.54
2011	0.257	0.204	-0.167	0	0.178
2012	0.132	-0.17	0.185	-0.112	-0.192
2013	0.406	0.155	0.201	-0.152	-0.2
2014	0	0.286	-0.188	0.451	0
2015	0.14	0.399	0.193	0.261	0.170
2016	0.274	0.128	0.663	-0.433	0.168
2017	0.366	0.365	0.453	-0.196	0

7.2 Result Analysis

We summary a strategy for confronting the opioid crisis in the following two aspects:

•Strengthen control over the source, especially in the possibly threshold years.

•Increase social welfare appropriately in economically developed areas to alleviate the living pressure of the young and middle-aged population.

The parameters that may affect decision-making are mainly due to the inflow and outflow of the population and economic benefits of the year. When the economic benefits change greatly, it may affect the contribution of socio-economic characteristics, while the immigration of population will have an impact on both the drug spreading process and the socio-economic characteristics.

The model we use is simple and easy to understand. It is conceived from the most basic linear propagation. Assuming that it is a first-order Markov model, we find that it is suitable for multiple linear regression and can get a good result.

We use data preprocessing to make the data look intuitive, and in the subsequent modeling, we also use similar methods, such as making RGB diagrams of weight matrices to observe their contributions and possible sources.

References

[1] Kong X , Qi Y , Song X , et al. Modeling Disease Spreading on Complex Networks[J]. Computer Science & Information Systems, 2011, 8(4):1129-1141.

[2] Xiaoning L , Ancha X . The Dynamic Interaction Between the Interest Rate Of Informal Finance and Formal Finance——Based on Empirical Research of Bayesian VAR Model[J]. Journal of Financial Development Research, 2017.

[3] Shuai Z , Van d D P . Global Stability of Infectious Disease Models Using Lyapunov Functions[J]. SIAM Journal on Applied Mathematics, 2013, 73(4):1513-1532.

[4] http://pandas.pydata.org/pandas-docs/stable/

[5] https://blog.csdn.net/golden_xuhaifeng/article/details/ 79610947