Advances in Computer, Signals and Systems (2019) 3: 21-24
Clausius Scientific Press, Canada

DOI: 10.23977/acss.2019.31004
EISSN 2371-8838

Opioid Crisis solutions based on big data processing and linear regression

Hanqi Wang[1], Ye Wang[2]

[1]*Department of Electronic Information and Electrical Engineering , Dalian University of Technology, Dalian, Liao Ning, China, 116024*
[2]*Department of Chemical Engineering, Dalian University of Technology, Dalian, Liao Ning, China, 116024*

Keywords: Thermal protective clothing design, One-dimensional heat conduction model, Difference equation, Heat balance analysis, Dichotomy search

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.

1. Problem Background

Life expectancy in U.S., reported by the centers for disease control and prevention, has fallen for two years in a row since 2016 and the main reason is opioid abuse. Synthetic and non-synthetic opioids used for pain control or pastime purposes can result in chronic drug addiction, which will damage both physical and psychological health of individuals. Even worse, it may cause the constraint for economy development, for employees who abuse drugs will suffer social function degradation and not be able to complete sophisticated and high-demanded work.

Therefore, it is of great importance for the government of each states to analyze their current situation, focus enforcement efforts on punish drug trafficking and carry out efficient policies to deal with their issue accordingly.

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

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

2) 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:

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

2) Features of people who is likely to overdoes,

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

2. Data preprocessing and qualitative analysis

In data merging, considering the policy and production differences, we combine the opioid into two categories-heroin and synthetic opioid. Now, we use the time (year) and county number as the index to add and classify the number of the reported drug to get a new data set.

Based on this data set, we make a histogram of the number of the reported synthetic opioid and heroin in all counties in each state over time. Then, we only use the 21st state as an example for a brief analysis.

According to the analysis of above figures, we can make the conclusions as following:

1) There is a stable and large number of opioid abusing cases in some counties,

2) Meanwhile, for some counties, although there is a large number of opioid abusing cases, the number of cases has change is a lot.

3) As for those counties where there is a small and stable number of cases, we believe that those data are the result of statistical data fluctuations

According to our assumption the mathematics model of the spread of opioid crisis in different counties and states is first-order Markov model and the spread matrix (weight matrix) of the reported synthetic opioid and heroin cases in and between the five states and their counties from 2010 to 2017 is stable. So, in the process of propagation, the size of the matrix element in the transfer matrix indicates the contribution of the county corresponding to the matrix element to the total amount of propagation. Therefore, the size of the matrix element in the transfer matrix can be used to judge whether the county is one of the sources.

It is worth noting that since the transfer matrix is in a stationary and independent parameter, the propagation source predicted by the transfer matrix is convincing regardless of whether there is interstate propagation. Thus, when constructing a discriminant model for states 39, 42 and 51, we can also use this model, which has no meaning for predicting the total number of cases in subsequent years, and is only used as a means of judging the source of the propagation.

3. 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:

$$Yi = \Sigma\alpha k,i \cdot Xk,i \tag{1}$$

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.

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 1 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.

4. Model Realization

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

$$N_k^{new} = \mu \cdot N_k + \nu \cdot \Sigma\alpha_i \cdot F_i \tag{2}$$

Whereμandνare 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:

$$Nknew = 0.37756 \cdot Nk + 0.62244 \cdot \Sigma\alpha i \cdot Fi \tag{3}$$

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 2.

Table 2 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

5. 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.

6. Strengths and weaknesses

Strengths: The model we use is simple and easy to understand. It is conceived fromthe 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.

Weaknesses: Cause we use the multivariate linear regression model, its fluctuation range affected by singular value is very large, which may lead to huge errors in prediction.

References

[1] Matthew B J, Becker W C, Keith H. Big Data and the Opioid Crisis: Balancing Patient Privacy with Public Health [J]. The Journal of Law, Medicine & Ethics, 2018, 46(2):440-453.

[2] Wickramatilake S, Zur J, Mulvaney-Day N, et al. How States Are Tackling the Opioid Crisis [J]. Public Health Reports, 2017, 132(2):171-179.

[3] Mark J E, Phoenix D, Gutierrez C A, et al. Tackling the opioid crisis: Implementation of an ultra-restrictive opioid prescription protocol in patients undergoing major gynecologic surgery radically decreased dispensed opioid without reducing pain control [J]. Gynecologic Oncology, 2018, 149:5.

[4] Gray H J, Rind C, Wu E S, et al. Tackling the opioid crisis: Reduced postoperative oral and intravenous opioid use after implementation of an enhanced recovery after surgery (ERAS) program in gynecologic oncology patients [J]. Gynecologic Oncology, 2018, 149:37.

[5] Wickramatilake S, Zur J, Mulvaney-Day N, et al. How States Are Tackling the Opioid Crisis [J]. Public Health Reports, 2017, 132(2):171-179.