

基于多种预测模型的阿片药物分布研究

彭巧娟¹,周毅²,刘娟³

(1. 南华大学计算机学院,衡阳 421000; 2. 南华大学资源环境与安全工程学院,衡阳 421000;

3. 南华大学化学化工学院,衡阳 421000)

摘要:

近年来美国正在经历关于使用合成和非合成阿片类药物的国家危机,通过 GM(1,1)得到五个州合成类阿片药物随时间的变化规律函数;通过时间序列预测某县海洛因总量的值;通过灰色关联度得到阿片类药物使用趋势与美国人口普查提供的社会因素的相关性,利用神经网络进行修改验证,提出应对阿片危机的可能策略。

关键词:

阿片危机; GM(1,1)模型; 时间序列; 灰色关联度; BP 神经网络

0 引言

(1)问题背景

阿片类药物^[1]就是指从罂粟中提取的树脂渗出物,加上人工半合成的麻醉、镇痛性药物,具有一定的成瘾性。它是癌症患者和其他重症患者遭受极度痛苦时为缓解疼痛采用的手段,本应当被严格限制使用,如今在美国却唾手可得,由此引发出了一场关于使用阿片类药物的全国性危机。考虑到药物的传播具有速度快、范围广等特点,同时阿片类药物的滥用存在多个方面的原因,如何正确估算和预测阿片类药物随时间的传播和特点以及分析阿片类药物的使用如何达到目前的水平,将为药物管理部门监控和管理阿片类药物和缉毒署推进禁毒情报工作等提供理论依据。

(2)已有的研究

2011年4月,白宫提出国家处方药滥用预防计划^[2],通过宣传教育和大数据追踪监控试图改善阿片类药物滥用的现状。2016年3月,美国医学会杂志(JAMA)发布了 CDC^[3]最新版阿片类药物治疗慢性疼痛使用指南,对药物利弊等作了详细说明,一定程度上能提高病人的用药风险意识^[4]。但阿片类药物滥用问题的控制并非一朝一夕,在各方利益博弈中,阿片类药物滥用问题的解决之路依旧困难重重。

(3)研究目标

①目标一

建立一个数学模型,描述五个州(俄亥俄州、肯塔基州、西弗吉尼亚州、弗吉尼亚州、宾夕法尼亚州)及其县之间报告的合成阿片和海洛因事件(案例)随时间的传播和特点。使用该模型,确定五个州中每一个州可能已经开始使用特定阿片类药物的任何可能位置。

②目标二

关于阿片类药物的使用如何达到目前的水平,是谁在使用/滥用它,以及为什么阿片类药物的使用尽管存在已知的危险,但仍然持续存在,使用或趋势是否与美国人口普查提供的任何社会经济数据有某种关联?

③目标三

找出一个应对阿片危机的可能策略。

1 符号与名词解释

表 1

符号	定义
x_t	某个州在第 t 年的药物总量
D_i	i 县海洛因的总量
T_i	i 县药物总量
r_j	县药物总量与影响人口的第 j 个因素之间的相关性
x	存储 i 州每年药物总量的矩阵
$y_{F,t}$	F 县在第 t 年的药物总量
s_i	i 县药物总量与该州药物总量的比值

2 模型的建立与求解

2.1 目标一:模型的建立与求解

(1)各州合成类阿片随时间变化的趋势

①变化趋势

为了研究五个州中每一个州合成类阿片类使用量的变化,我们将 2010-2017 年各州合成类药物的使用量筛选出来,进行散点分析,得到它们的变化趋势图,如图 1 所示:(数据来源于 <https://www.cdc.gov/features/confrontingopioids/index.html>)。

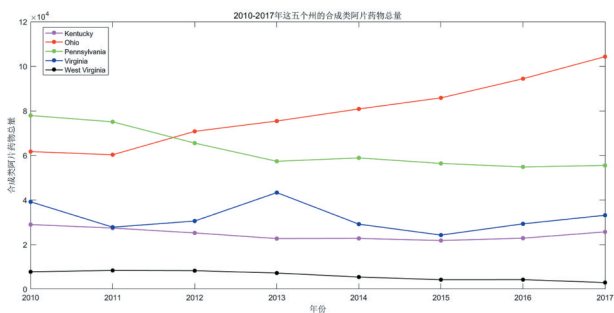


图 1 2010-2017 年各州合成阿片类药物的综合计数

②图像分析

我们可以发现俄亥俄州的阿片药物总使用量随时间呈上升趋势,宾夕法尼亚州的阿片药物总使用量随时间呈稳步下降趋势,肯塔基州和西弗吉尼亚州的药物使用量趋势相对稳定,弗吉尼亚州的药物使用量随时间呈上下波动趋势,我们对各州药物总量随时间的变化进行 GM(1,1)模型预测,得到各州的药物总量的变化规律。

以俄亥俄州为例。

对原始数据 $x^{(0)}$ 作一次累加,得:

$$x^{(1)} = (70999, 142881, 227696, 321443, 422866, 532016, 647292, 766641)$$

构造数据矩阵 B 及数据向量 Y,有:

$$B = \begin{bmatrix} -1/2(x^{(1)}(1) + x^{(1)}(2)) & 1 \\ -1/2(x^{(1)}(2) + x^{(1)}(3)) & 1 \\ \vdots & \vdots \\ -1/2(x^{(1)}(7) + x^{(1)}(8)) & 1 \end{bmatrix}, Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(8) \end{bmatrix}$$

计算:

$$\hat{u} = \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = (B^T B)^{-1} B^T Y = 1.0e + 0.4 * \begin{bmatrix} -0.0000 \\ 6.9626 \end{bmatrix}$$

于是得到 $\hat{a} = 0, \hat{b} = 1.0e + 0.4 * 6.9626$

建立模型:

$$\frac{dx^{(1)}}{dt} + \hat{a}x^{(1)} = \hat{b}$$

求解,得:

$$\hat{x}^{(1)}(k+1) = \left(x^{(0)}(1) - \frac{\hat{b}}{\hat{a}} \right) e^{-\hat{a}k} + \frac{\hat{b}}{\hat{a}} = 977915.0 * \exp(0.0767746 * t) - 906916.0$$

③变化趋势方程

最终得到各州的合成类阿片总量随时间的变化规律(其中,为了保证预测模型的准确,我们舍弃了弗吉尼亚州 2013 年的突增点):

$$\text{俄亥俄州(OH): } x_t = 723614.0e^{0.0831486t} - 661916.0$$

$$\text{肯塔基州(KY): } x_t = 1532144 - 1503188e^{-0.0169621t}$$

$$\text{西弗吉尼亚州(WV): } x_t = 67156.8 - 59390.8e^{-0.164836t}$$

$$\text{弗吉尼亚州(VA): } x_t = 5114977 - 5075811e^{-0.00625324t}$$

$$\text{宾夕法尼亚州(PA):}$$

$$x_t = 1512155 - 1434277e^{-0.0499548t}$$

(2)形成变化趋势的分析

我们研究这五个州的地理位置发现,这五个州以靠近水源的俄亥俄州为起点,向四周发散。其中,俄亥俄州是两大交通动脉 70 号和 75 号公路交汇枢纽,是墨西哥毒枭分销毒品和禁药最理想的地点,同时也是阿片类药物流通的极佳地点。宾夕法尼亚州同样作为一个靠近水源的州,它拥有十分发达的交通运输系统,该州阿片类药物的使用量相对较高。肯塔基州、西弗吉尼亚州和弗吉尼亚州由于地处山区,交通相对来说不太便利,因此,这三个州阿片类药物的使用量相对较低。

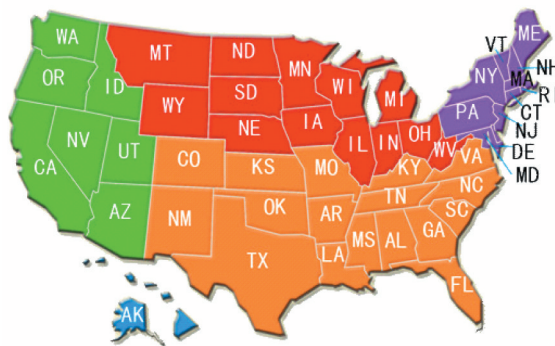


图 2 五个州的地理分布

(3)各州及其县之间海洛因随时间的传播和特点。

①模型的建立与求解

以肯塔基州海洛因的使用量为例。我们先对肯塔基州每个县的海洛因总数和对应县的阿片类药物总量进行标准化处理:

$$n_i = \frac{D_i}{T_i}$$

我们对标准化后的数据与县代码进行处理,以县代码为 x 轴,以标准化数据为 y 轴,得到海洛因在每个县内的波动情况如图 3。

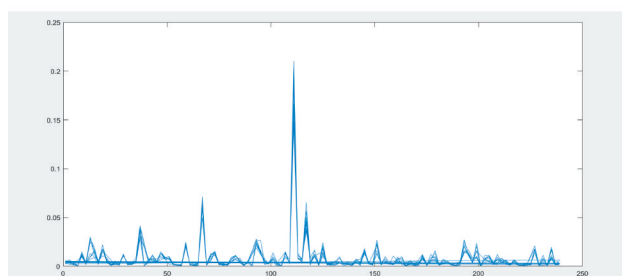


图3 某县(用县代码表示)海洛因的比例

通过图像我们可以发现肯塔基州的前 125 个县内海洛因所占比重波动特别明显,后面 100 多个县虽然也存在这种情况,但波动很小,我们可以猜测海洛因主要存在于编号为 75-125 的县内,并以此为中心,向四周扩散。

我们将年份、县代码、县海洛因物质总量建立一个数据集,对肯塔基州各县的数据用一个矩阵进行存储,当我们需要进行预测时,只需输入该县的县代码,即可得到该县近几年海洛因数据的预测值。

记原始时间序列为 $a_{Fi}(t=2010, 2011, \dots, 2017)$, 对于时间序列样本 $a_{F2010}, a_{F2011}, \dots, a_{F2017}$, 记 a_{Fi} 的秩为 $R_i = R(a_{Fi})$, 考虑变量对 $(t, R_i), t=1, 2, \dots, 8$ 的 Spearman 相关系数 q_s , 有:

$$q_s = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n (t - R_i)^2$$

为了构造平稳序列,对序列 $a_{Fi}(t=2010, 2011, \dots, 2017)$ 作一阶差分运算 $b_{Fi} = a_{Fi(t+1)} - a_{Fi}$, 得到序列 $b_{Fi}(t=2010, 2011, \dots, 2017)$ 。从时间序列 b_{Fi} 的散点图,我们可以对每个县建立其预测模型:

$$y_{Fi} = c_{F1}y_{F(t-1)} + c_{F2}y_{F(t-2)} + \varepsilon_{Fi}$$

F: 表示县编号, t: 表示时间, c_{F1}, c_{F2} : 各县的待定参

数, ε_{Fi} : 各县的随机扰动项。

我们随机选取两个县进行预测,可以得到的预测值如表 2(以肯塔基州 047 号县为例)。

表 2 肯塔基州 047 号县的时间序列预测值

年份	2011	2012	2013	2014	2015	2016	2017
原始值	2	1	2	7	2	2	3
预测值	2	1	1.92	舍去	1.92	2.56	2.56
相对误差	0	0	0.04	舍去	0.04	0.28	0.1467

②结论和分析

预测可知,肯塔基州中 Laurel 等县可能已经开始使用特定阿片类药物,美国政府应该对这五个县加大监察力度。同理,我们也将俄亥俄州、西弗吉尼亚州、弗吉尼亚州和宾夕法尼亚州的相关数据全部存入一个矩阵中,当我们需要预测任何地点、任何时间的海洛因总量,只需要输入该县的县代码和预测年份,即可得到该县所需年份对应的数据预测值。

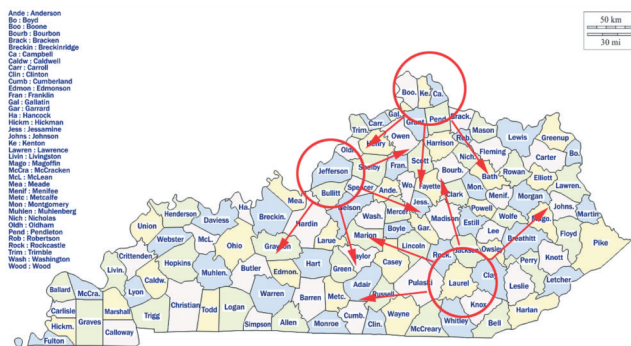


图4 肯塔基州各县的位置图

图中红色圆圈里的这些县都是肯塔基州阿片类药物使用量偏高的位置,这些县都有一个共同的特点,就是交通十分便利,有的是交通枢纽,有的是靠近水源。阿片类药物有以这些县为中心,向其他县扩散的趋势。因此,对于各大交通枢纽及临河城市,美国相关部门应该加大监察力度,从源头控制非法阿片类药物的扩散。

③模型的进一步验证

为了增加预测的准确性,我们对筛选出的数据采用灰色预测,对各州各县的海洛因总量进行预测,我们得到已知年份对应数据的预测值,通过与原始数据进行比对,我们可以得到表 3。

表3 肯塔基州047号县的灰色预测值

年份	2011	2012	2013	2014	2015	2016	2017
原始值	2	1	2	7	2	2	3
预测值	2	1.0901	1.9733	Exclude	1.9188	2.3402	2.854
相对误差	0	0.0901	0.0133	Exclude	0.0406	0.1701	0.0487

上面的结果是对前面结果的预测,同时也是一个升华,使得预测的数据更加精确。

(4) 药物识别阈值水平

① 问题分析

通过对背景分析,我们可以得知每个州的犯罪实验室提供的合成类阿片和海洛因的药物识别计数的数据并不一定是该州的真实记录,当样本中没有提供确定的位置数据时,该位置数据由市/县/州调查执法机构的位置数据代替。

为了降低由各州药物识别计数数据水平差异引起的误差,我们根据给定的每个县的合成类阿片和海洛因的药物识别计数的数据,对县阿片类药物总量和州阿片类药物总量进行标准化处理,得到二者之间的比值,通过下列流程图的方法,对其进行处理,得到一个比例阈值,低于该阈值时,则认为该县的阿片是由于没有提供位置数据导致的,高于该阈值时,则认为该县的阿片类药物是由于已经开始使用导致的。

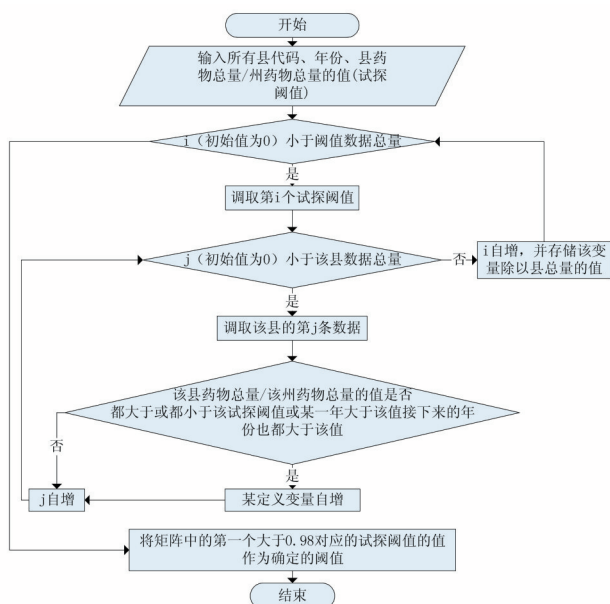


图5 确定阈值的流程图

② 阈值的确定

我们将初始阈值定为 0.001,每次以 0.001 的幅度

递增,用各县标准化处理后的值与试探阈值进行比较,当该县所有年份的值都大于试探阈值、都小于试探阈值、其中某一年份的值大于且该年份之后的值也大于试探阈值这三种情况发生时,则认为可以判断该阈值正确的频数增加一个,当频数与县的总数之比大于 0.98 时,我们最终得到符合要求的阈值为 0.08。

2.2 目标二:模型的建立与求解

(1) 模型的建立与求解

为了测得每个县的阿片类药物的总使用量和这 23 种因素的相关性强度,我们拟采用灰色关联分析法^[5]来进行求解,将每个县的阿片类药物总使用量存入数组 $Y = \{y_1, y_2, \dots, y_n\}$, 将我们分类完的 23 种人口因素存入数组 $X_j = \{x_{1j}, x_{2j}, \dots, x_{nj}\}$, 我们采用下面的公式分别对每一种因素与阿片类药物总使用量进行 Pearson 相关性分析。

$$r = \frac{N \sum x_{ij} y_i - \sum x_{ij} \sum y_i}{\sqrt{N \sum x_{ij}^2 - \left(\sum x_{ij} \right)^2} \sqrt{N \sum y_i^2 - \left(\sum y_i \right)^2}}$$

$j=1,2,\dots,23$

通过计算,得到每年县阿片类药物鉴定总数和美国人口普查提供的社会经济数据之间的相关性表格,分析结果如柱状图所示。

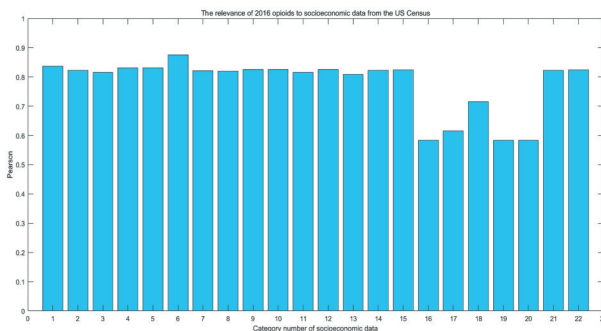


图6 2016年阿片类药物与美国人口普查的社会经济数据的相关性

(2) 评价结论

由数据分析结果以及表 4 可知,阿片类药物的使用和趋势与多种社会经济数据都有极强的相关性,如教育程度、生育力等,其中,与教育程度的关联度最大。

表 4 不同值范围变量的相关性强度

每个值范围确定变量的相关强度					
相关系数	0.8-1.0	0.6-0.8	0.4-0.6	0.2-0.4	0.0-0.2
相关性	极强相关	强相关	中等程度相关	弱相关	极弱相关或无相关

(3)阿片类药物的使用对象

分析可知,和祖父母一同生活的未成年人是使用或滥用阿片类药物的主要群体,当然,青年和中年人群滥用阿片类药物的现况同样也让人担忧。

(4)阿片类药物的使用和成瘾增长的原因

阿片类药物的滥用和成瘾问题已经成为社会讨论的焦点,执法机构和戒毒团体经常将滥用和成瘾的原因归结为药企盲目扩张市场以及药品监管部门的不作为,但深入分析不同滥用行为和滥用程度的人群,其滥用原因也不尽相同。

以和祖父母一同生活的未成年人为例,他们的判断力明显弱于成年人,同时在成长的过程中,往往得不到正确的引导,因此,缺乏基本常识的非治疗性娱乐用药,是他们对阿片类药物成瘾的重要原因。

而对于大多数成年人来说,他们往往十分清楚服用这类药物的潜在危害,却依旧选择使用这些药物,原因来自两个方面。一方面,本着“免除疼痛是全人类的权力”的原则,他们选择服用医生开出的大剂量阿片类药物来缓解身体上的疼痛。另一方面,现代人面临的工作压力十分大,为了逃避生活的不如意,他们只要谎称身体上有疼痛,就能轻而易举的得到这些药物。

(5)对第一部分模型的修改

我们筛选出相关性最强的三个因素,与县阿片类药物总量进行 BP 神经网络预测,采用下式分别对样本的输入、输出数据进行规格化处理:

$$\tilde{k} = \frac{2(k - k_{\min})}{k_{\max} - k_{\min}} - 1$$

K 为规格化前的变量; k_{\max} 和 k_{\min} 分别为 k 的最大和最小值; \tilde{k} 为规格化后的变量。

通过 MATLAB 执行算法:

$$k = \frac{(\tilde{k} + 1)(k_{\max} - k_{\min})}{2} + k_{\min}$$

通过上式,我们对 2016 的数据进行预测处理,得到部分县相对误差值如表 5 所示。

表 5 神经网络预测值表

县代码	年份	原始值	预测值	相对误差
21009	2016	289	274.559	0.05
21047	2016	390	367.5152	0.0577
39009	2016	698	719.683	0.0311
39065	2016	186	307.2374	0.6518
42031	2016	125	134.2653	0.0741
42077	2016	796	918.5262	0.1539
51047	2016	419	303.0658	0.2767
51041	2016	1994	1.75E+03	0.122
54057	2016	26	18.7092	0.2804
54069	2016	88	115.5035	0.3125

2.3 目标三:应对阿片危机的策略

翻阅以往的文献,我们发现,遏制阿片类药物滥用的干预措施多集中于患者,但缺少对不同滥用行为和滥用程度的人群的滥用原因的深入分析以及对不同程度滥用者药物来源的差异性研究。结合我们前面建立的模型和得出的结论,我们提出了一个方案:

第一步,根据每个州每年提交的合成类阿片和海洛因的药物识别计数数据,利用我们第 1 问的模型确定每个州及其县可能已经开始使用特定阿片类药物的时间和位置。

第二步,根据美国人口普查局提供的社会经济数据,划分出不同年龄阶段、不同职业的人群,用第 2 问的 Pearson 分析确定最有可能滥用阿片类药物的人群类别,结合 PDMPs 记录的每个患者开具处方和配药信息的数据,对涉嫌滥用或转移处方的人群类别的相关信息进行了监控。

最后,结合调取数据的程序反馈出来的信息,筛选出各州各县涉嫌滥用阿片类药物的人群类别,对这些人群的特点进行分析,找到他们滥用阿片类药物的原因,对症下药,加大对这些人群的教育宣传力度,控制和酌情减少他们阿片类药物的使用量。

3 模型的评价与推广

3.1 模型的评价

(1)优点

①能够很好地通过县代码把该县的有关数据提取出来,从而能够对该县的数据单独进行分析预测,使之不受州的阿片类药物使用水平限制;

②剔除了突变过大或者过小的量,在一定程度上

减少了由于未提供位置数据造成的影响;

③通过对程序多次循环试探得出了一个合适的阈值,可信度达 98%;

④利用多种预测模型进行对比分析,增大了分析结果的准确度;

⑤通过对数据的标准化处理,不同州的所有的数据都可以在同一水平进行讨论。

(2) 缺点

①处理的方法有些采用人工合作处理,数值的精确性有待加强;

②由于存在未提供位置数据的情况以及部分数据缺失严重,从而导致部分县内的数据变化并不是特别

规律,以致结果存在一定的误差。

3.2 模型的应用与推广

(1) 应用

可用于预测每个县的阿片药物总量和各州的药物总量,并判断其是否已经开始使用特定阿片类药物,为各州/市/县针对阿片危机政策起指示作用,能够针对性地采取措施,减少了不必要的花费。

(2) 推广

该模型同理可以应用于预测人口的增长趋势和其他的受地区影响程度较大的数据模型,如各国的经济增长、商品的进出口等方面问题。

参考文献:

- [1]丁香园. 急诊科危机:过量使用阿片类药物[J]. 临床合理用药杂志,2014,7(33):12.
- [2]孙丽. 美国阿片类药物滥用及其管控措施[J]. 中国药物滥用防治杂志,2018,04.
- [3]杨玉慧,许秀丽,朱珠. 美国阿片类药物滥用及其治理举措概述[J]. 中国药物警戒,2017,12.
- [4]常玲,张瑜. 从阿片类药物研究的今昔得到的启示[J]. 中国药学杂志,1989,02.
- [5]美国修订阿片类药品的风险管理计划[J]. 中国医药导刊,2018,20(11):699.

作者简介:

彭巧娟(1997-),女,湖南邵阳人,本科

周毅(1997-),男,湖南娄底人,本科

刘娟(1997-),女,湖南株洲人,本科

收稿日期:2019-02-28

修稿日期:2019-03-08

Opioid Distribution Based on Multiple Predictive Models

PENG Qiao-juan¹, ZHOU Yi², LIU Juan³

(1. School of Computer Science, Nanhua University, Hengyang 421000; 2. School of Resources, Environment and Safety Engineering, Nanhua University, Hengyang 421000; 3. School of Chemistry and Chemical Engineering, Nanhua University, Hengyang 421000)

Abstract:

In recent years, the United States is experiencing a national crisis on the use of synthetic and non-synthetic opioids. This paper uses GM (1, 1) to obtain the regularity function of five-state synthetic opioids over time; the time series predicts the total amount of heroin in a county. Value; the correlation between the trend of opioid use and the social factors provided by the US census is obtained through the grey correlation degree, and the neural network is used to modify and verify, and the possible strategies for dealing with the opioid crisis are proposed.

Keywords:

Opioid Crisis; GM(1,1) Model; Time Series Model; Grey Correlation; BP Neural Network