

Searching for the Extreme Value of Opioid Events Based on BP Neural Network and Genetic Algorithm

Mingchao Zhang Xinyi Chu Zhongrui Zhao

Qingdao University of Technology, Qingdao, Shandong, 266000, China

Abstract

In this paper, a comprehensive model of BP neural network and genetic algorithm is established and analyzed and solved according to a group of common socio-economic factors collected by the U.S. Census Bureau in the counties of five states in 2010-2016. Thus, the characteristics and sources of opioid drug transmission in five states of the United States are obtained, which is conducive to the further alleviation of the opioid drug crisis.

Keywords

opioids; BP neural network; genetic algorithm

基于 BP 神经网络和遗传算法的阿片类药物事件极值寻找

张明超 储心怡 赵中睿

青岛理工大学, 中国·山东 青岛 266000

摘要

本文建立 BP 神经网络和遗传算法的综合模型, 根据美国人口普查局关于 2010-2016 每年中五个州的县收集的一组共同的社会经济因素数据, 进行分析和求解, 从而得到美国五个州的阿片类药物传播特征与药物的传播源, 有利于阿片类药物危机的进一步缓解。

关键词

阿片类药物; BP 神经网络; 遗传算法

1 背景

阿片类药物是一种镇痛效果最好的药物, 其唯一的来源是罂粟。反复使用阿片类药物可能会导致身体对毒品上瘾, 这是有害的, 甚至致命的。

随着阿片类药物在治疗、镇痛、娱乐等领域的应用日益广泛, 阿片类药物的滥用已成为美国严重的国家危机。根据美国疾病控制与预防中心的数据, 美国阿片类药物的消费量远远超过其他发达国家。当谈到阿片类药物使用和成瘾的威胁, 一方面, 滥用阿片美国政府将加载一个沉重的经济负担, 另一方面, 滥用阿片类药物会影响美国的劳动力的数量和质量, 阻碍美国的经济的发展。因此, 美国政府迫切需要寻求缓解这一局面的解决方案。

DEA/ 国家法医实验室信息系统 (NFLIS) 作为缉毒局

(DEA) 转移控制办公室的一部分, 发布了一份数据密集的年度报告, 涉及“联邦, 州分析的药品鉴定结果和相关信息和当地的法医实验室”。NFLIS 内的数据库中包括处理犯罪实验室的数据超过全国估计的每年 120 万州和地方毒品案件的 88%。对于这个问题, 我们关注美国五个州的个别县: 俄亥俄州, 肯塔基州, 西弗吉尼亚州, 弗吉尼亚州和田纳西州。在美国, 一个县是每一个拥有权力的州的下一级政府。

2 任务

根据 NFLIS 提供的数据, 其中包括 2010-2017 年麻醉镇痛药 (合成阿片类药物) 和海洛因的药品鉴定计数, 我们需要建立一个数学模型来描述五个州及其县之间的合成阿片类药物和海洛因事件 (病例) 的传播和特征。使用所建立的模型, 确定在五个州中可能已开始使用特定阿片类药物的地区。

3 BP 和 GA 综合模型的建立

3.1 数据分析

首先，我们对提供的 NFLIS 数据进行预处理，对数据进行审查的过程中，针对题目附件原始数据中出现的异常数据，如数据缺失，数据格式错误等，采取插值法、取平均值等方法，消除这些奇异数据给统计结果造成误差的可能性。

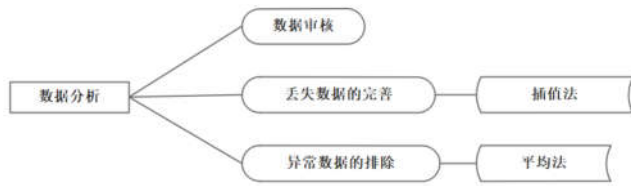


图1 数据分析的描述

3.2 BP 神经网络模型

首先，我们对五个州及其县的经纬度进行查询，根据所查询的数据绘制出五个州及其县的地理位置图。我们引入一个指标：

$$\text{药物传播度} = \frac{\text{药物鉴定计数}}{\text{地区面积}}$$

我们用这个指标来描述每个地区的药物传播程度，以此来研究两种药物在五个州及其县的传播特征。由于州县的数量过多，逐一求出各个县的地区面积太过于繁杂，根据我们的假设，我们可以用每一个地区的药物鉴定计数代表各个地区的药物传播程度。

由于定性分析合成阿片类药物和海洛因事件的传播和特征需要对美国社会和经济状况有较深刻的理解，同时，提供的 NFLIS 数据中的变量不多（变量仅经度、纬度和药物鉴定计数），统计学方法又不足以完全揭示变量与药物传播特征的关系。所以，我们最终决定选择具有非线性、自适应信息处理系统的模型——神经网络，一方面，人工神经网络具有较强的自组织、自适应与自学习能力，能够在未完全了解阿片类药物传播原理的情况下，完成自变量、变量间与药物传播度之间的非线性映射；另一方面，仅通过神经网络的输入输出数据难以准确寻找阿片类药物鉴定计数的极值（即传播源），而遗传算法具有全局的非线性寻优能力。综合上述考虑，我们建立神经网络结合遗传算法求解药物传播源的数学模型。

3.2.1 BP 神经网络模型概述

BP 神经网络^[1]是一种多层前馈神经网络，一般由单层输

入层、单层或者多层隐含层以及单层输出层构成，该网络的主要特点是信号前向传递，误差反向传播。在前向传递中，输入信号从输入层经隐含层逐层处理，直至输出层。每一层的神经元状态只影响下一层神经元状态。如果输出层得不到期望输出，则转入反向传播，根据预测误差调整网络权值和阈值，从而使 BP 神经网络预测输出不断逼近期望输出。BP 神经网络的模型结构如图 2 所示。

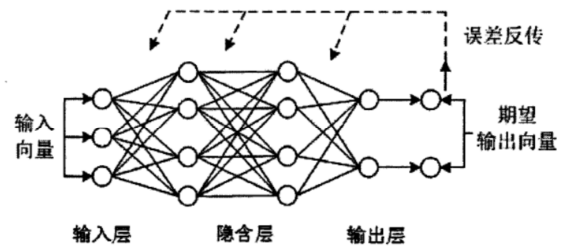


图2 BP神经网络模型结构

3.2.2 数据预处理

原始的 391 组数据显然不能满足神经网络的训练要求，因此，我们从问题一中插值过后的 6000 组数据中随机选择 10000 组数据。前 9900 组作为 BP 网络的训练数据集，后 100 组数据作为 BP 网络的检验数据集。

数据归一化可以方便后面数据的处理，并保证程序运行时收敛加快。我们将数据集按如下公式进行归一化处理：

$$y = \frac{(y_{\max} - y_{\min})(x - x_{\min})}{x_{\max} - x_{\min}} + y_{\min}$$

其中， $y_{\max} = 1$ ， $y_{\min} = -1$ ， x 为需要归一化的数据集。

3.3 建立 BP 网络

提供的数据经过插值可以比较准确的反应药物事件传播随地理位置的变化，我们建立 2-N-1 的 BP 网络结构，其中，2 表示输入项（分别为经度 x ，纬度 y ）；N 为隐藏层神经元个数；1 表示输出项（药物传播度）。结构图如下：

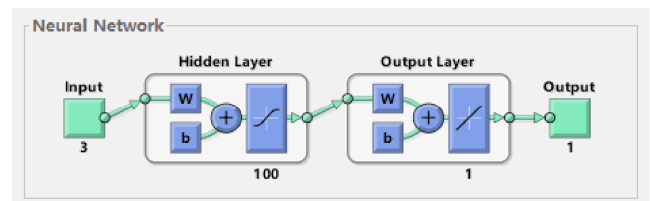


图3 BP网络结构图

在实际问题中，隐含层节点数的选择首先是参考公式来确定节点数的大概范围，然后用试凑法确定最佳的节点数。

经过多次实验,我们选择 N=100,此时 BP 神经网络达到了较高的精度。

学习速率同样对 BP 神经网络具有重要影响作用,学习速率如果选得太小,网络学习缓慢,需要增加训练次数;学习速率如果选得太大,网络学习迅速,但容易导致网络不收敛,影响训练的精度,我们最终决定学习速率为 0.01,训练次数为 300。

BP 神经网络采用梯度修正法作为权值和阈值的学习算法,从网络预测误差的负梯度方向修正权值和阈值,没有考虑以前经验的积累,学习过程收敛缓慢。对于这个问题,可以采用附加动量方法来解决,带附加动量的权值学习公式为:

$$\Delta w(t+1) = -\eta \cdot \frac{\partial E}{\partial w} + \alpha \cdot \Delta w(t)$$

3.4 神经网络结果分析

3.4.1 神经网络训练结果

以 2010 年合成阿片类药物的传播为例,训练结束的神经网络性能如图 4 所示:

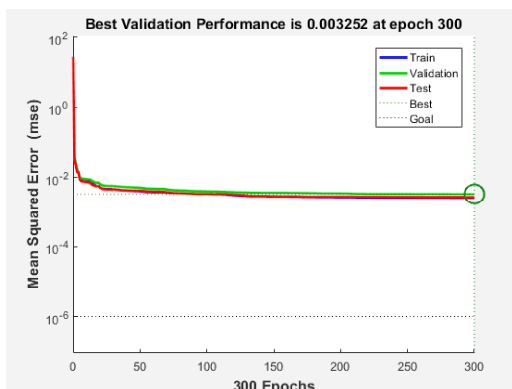


图 4 BP 神经网络训练性能

用训练好的神经网络预测药物的传播,如图 5、图 6。

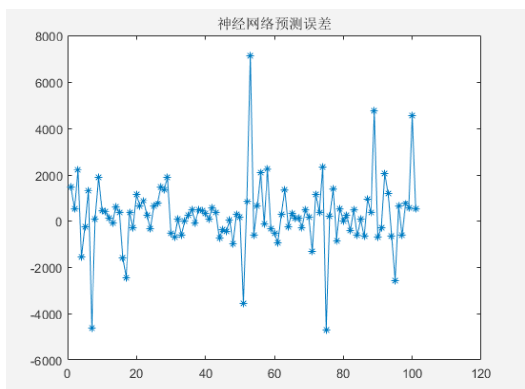


图 5 BP 神经网络预测输出

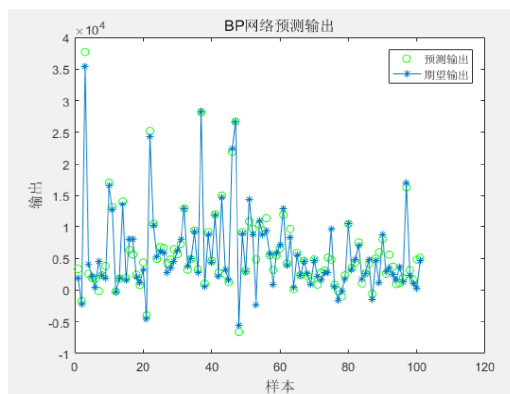


图 6 神经网络预测误差

在图 5 中,可以看出 BP 神经网络的预测输出与期望输出相差不大,基本达到了我们要求的效果。在图 6 中,可以看出神经网络预测误差不大,说明神经网络的训练结果满足我们的要求。

3.4.2 阿片类药物的分布特征

我们用训练好的神经网络进一步模拟出了 2010-2017 年 5 个州合成阿片类药物在不同地理位置鉴别计数的等高线图,如图 7:

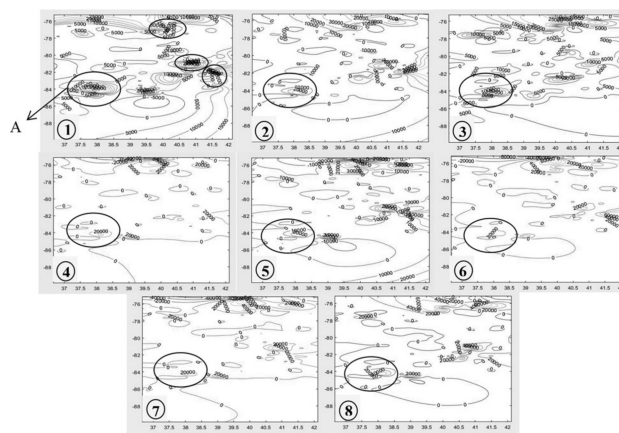


图 7 合成阿片类药物的分布特征

在图 7 中我们清晰地看到五个州及其县阿片类药物在地理位置上地分布和阿片类药物随年份传播的分布特征,从整体趋势上我们可以发现,这五个州及其县的阿片类药物浓度在 2010 年至 2017 年之间呈下降趋势。

在①中可以明显观察到有很多高值中心,我们随机取其中一个为研究对象,我们把该区域命名为区域 A。区域 A 的阿片类药物浓度大体上呈下降趋势,2010 年至 2011 年之间区域 A 的阿片类药物浓度大幅度下降,2011 年至 2012 年之间区域 A 的阿片类药物浓度有小幅上升,2013 年至 2017 年

之间区域 A 的阿片类药物浓度变化幅度很小。

3.4.3 阿片类药物的传播

同时，我们用训练好的神经网络模拟出了 2010–2017 年 5 个州合成阿片类药物在不同地理位置鉴别计数的空间分布图，如图 8：

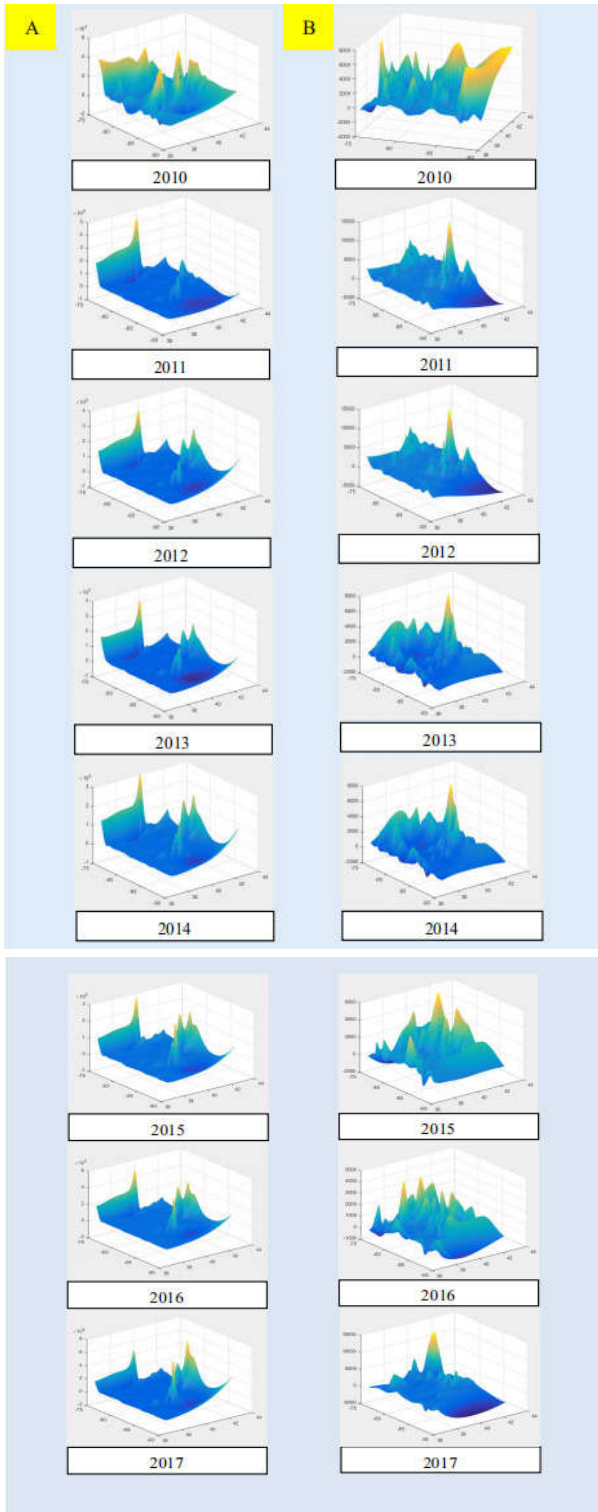


图 8 合成阿片类药物的空间分布

在图片 8 中，A 列图片表示合成阿片类药物在 2010 年至 2017 年在不同地理位置鉴别计数的空间分布图，B 列图片表示海洛因事件在 2010 年至 2017 年在不同地理位置鉴别计数的空间分布图。

3.5 遗传算法模型^[2]

训练好的神经网络可以准确地模拟两种药物传播的特征，但是由于神经网络是一个“黑箱”，仅通过网络的输入输出数据难以准确寻找药物事件传播度的极值（即传播源），我们在此引入遗传算法，通过区域分割的方法（某些县可能不止一个传播源）寻找传播源。算法流程图如图 9：

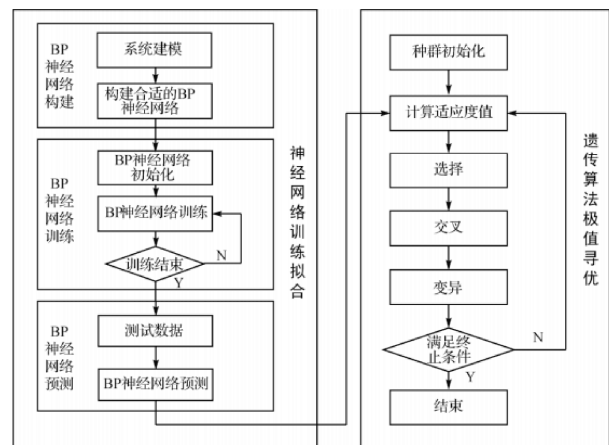


图 9 遗传算法流程图

3.5.1 遗传算法模型概述

遗传算法^[3]是一类借鉴生物界自然选择和遗传机制而发展起来的自适应全局优化随机搜索算法。遗传算法以编码空间代替问题的参数空间，以适应度函数为评价依据，以编码群体为进化基础，对群体中个体位串的遗传操作实现选择和遗传机制，建立起一个迭代过程。在这一过程中，通过随机重组编码位串中重要的基因，使新一代的位串集合优化于老一代的位串集合，群体的个体不断进化，逐渐接近最优解，最终达到求解的目的。

3.5.2 遗传算法模型构建

(1) 选择编码策略，用二进制的编码方法来表示每一个地理位置的经纬度，并将其看作一个个体；

(2) 定义适应度函数 $f(x)$ ，我们令训练后的 BP 神经网络的输出值为每一个对应个体的适应度；

(3) 确定遗传策略，我们确定群体大小 $N = 100$ ，交叉概率 $P_C = 0.25$ ，变异概率 $P_m = 0.01$ ；

- (4) 随即初始化生成群体 P ;
- (5) 计算群体中个体位串解码后的适应值 $f(x)$;
- (6) 按照遗传策略, 运用选择、交叉和变异算子作用于群体, 形成下一代群体;
- (7) 判断群体性能是否满足某一指标, 或者已完成预定迭代次数, 不满足则返回步骤(6), 或者修改遗传策略再返回步骤(6)。
- (8) 最终可以得到最优种群和目标值, 即我们可以得到药物传播源的位置。

3.5.3 模型结果

表格为将遗传算法与综合分析相结合的药物传播源, 如表1所示。“S”是药物传播源的缩写。其他年份的数据见表格。

使用我们的模型, 我们确定了在这五个州的每个州可能已经开始使用特定阿片类药物的位置。如果我们的团队所确定的模式和特征继续存在, 那么在各个州的边境地区可能会发生越来越多的阿片类药物事件。极值点所在的县很可能带动周边地区阿片类药物事件的发生。在极端点所在的县, 阿片类药物事件的数量很高。

表 1 2010 年药品传播源与中心坐标

海洛因数件数					合成阿片类药物				
经度	纬度	州	县	总计	经度	纬度	州	县	总计
-87.59	37.76	KY	Robards	6076	-84.12	38.07	KY	Winchester	40063
-80.40	38.27	WV	southern	2762	-75.98	36.79	VA	Suffolk	37102
-75.24	40.54	PA	Springfield	5289	-77.52	39.24	VA	Catocain	34314
-84.79	41.66	OH	Williams	6167	-81.16	40.93	OH	Flushing	37549

表 2 2011 年药品传播源于中心坐标

海洛因数件数					合成阿片类药物				
经度	纬度	州	县	总计	经度	纬度	州	县	总计
-84.07	39.62	OH	Sugarcreek	1574	-84.19	38.18	KY	Bourbon	33575
-75.82	39.71	PA	Franklin	4301	-77.32	39.71	PA	Franklin	35104
-80.34	41.63	PA	Township	9875	-75.88	39.74	PA	Elk	31492
-84.79	41.06	OH	Polding	7812	-80.68	42.12	PA	Erie	44592

表 3 2012 年药品传播源于中心坐标

海洛因数件数					合成阿片类药物				
经度	纬度	州	县	总计	经度	纬度	州	县	总计
-84.34	39.46	OH	Monroe	2876	-83.93	38.21	KY	sharpsburg	24166
-77.45	39.05	VA	Lawton	3981	-77.60	39.24	VA	Lawton	30720
-77.18	41.72	PA	Covington	4392	-75.90	39.74	PA	Elk	28338
-85.41	39.75	OH	Prebule	2177	-83.30	39.98	OH	Canaan	12774

表 4 2013 年药品传播源于中心坐标

海洛因数件数					合成阿片类药物				
经度	纬度	州	县	总计	经度	纬度	州	县	总计
-84.19	36.81	KY	Wheatley	2901	-84.01	38.21	KY	Bourbon	33599
-76.92	39.30	VA	Charles	595	-75.86	39.72	PA	Chester	64260
-77.20	41.76	PA	Covington	4347	-75.74	39.91	PA	Chester	75115
-84.67	39.66	OH	Prebule	2191	-87.84	41.72	OH	Williams	28293

表 5 2014 年药品传播源于中心坐标

海洛因数件数					合成阿片类药物				
经度	纬度	州	县	总计	经度	纬度	州	县	总计
-83.82	36.72	KY	Bell	2406	-84.15	38.11	KY	Bourbon	27669
-77.35	39.04	VA	Dranesville	3615	-77.35	37.72	VA	Catoctin	55706

-78.54	41.98	PA	Foster	6610	-75.76	39.85	PA	Square	63066
-82.75	40.94	OH	Auburn	2912	-82.12	41.42	OH	Loren	40974

表 6 2015 年药品传播源于中心坐标

海洛因数件数					合成阿片类药物				
经度	纬度	州	县	总计	经度	纬度	州	县	总计
-84.04	36.69	KY	Wheatley	4073	-84.42	38.36	KY	Harrison	26786
-76.66	39.60	WV	Burrick	2954	-75.83	39.78	PA	Chester	89819
-78.00	41.42	PA	Grove	5429	-75.71	39.85	PA	Chester	96921
-81.56	41.39	OH	Cuyahoga	4026	-81.23	41.69	OH	Lake	69411

表 7 2016 年药品传播源于中心坐标

海洛因数件数					合成阿片类药物				
经度	纬度	州	县	总计	经度	纬度	州	县	总计
-84.76	39.64	OH	Israel	1710	-83.08	38.79	OH	Kennett	13227
-77.01	40.32	PA	Rye	2913	-75.71	39.87	PA	Square	111386
-78.17	38.30	VA	Madison	2823	-75.77	39.72	PA	Township	100908
-82.86	41.52	OH	Danbury	1748	-84.75	41.10	OH	Benton	44599

表 8 2017 年药品传播源于中心坐标

海洛因数件数					合成阿片类药物				
经度	纬度	州	县	总计	经度	纬度	州	县	总计
-88.49	37.31	KY	Livingston	1332	-84.25	37.80	KY	Madison	61766
-78.45	37.68	VA	Buckingham	3081	-75.83	39.78	PA	Chester	82600
-75.24	42.12	PA	Wayne	10537	76.14	41.38	PA	Lake	90619
-84.76	40.07	OH	Dak County	1988	-81.13	40.98	OH	Stark	68656

4 结语

本文将 BP 神经网络算法与遗传算法巧妙结合，利用神经网络拟合阿片类药物传播分布图，遗传算法寻找不同区域的极值，并且利用训练好的神经网络解决了遗传算法中适应度的计算问题。此模型易于解决拟合传播方向、极值寻优一系列问题。

参考文献

- [1] 闫凌飞. 基于神经网络与遗传算法土壤成分含量研究 [D]. 吉林大学, 2006.
- [2] 陈晓桐. 基于改进单亲遗传算法的车辆路径优化问题研究 [J]. 山东工业技术, 2018(19):212.
- [3] 苏艳娇. 基于遗传算法的建筑节能多目标优化 [D]. 深圳大学, 2017.