

Analysis of Addictive Drug Transmission Based on Artificial Neural Network

Yuan Sun, Zan Yang*, Ke Xu, Yi Zhang, Dan Li

Tongji Zhejiang College, Jiaxing Zhejiang
Email: *yangzan953@163.com

Received: Feb. 11th, 2019; accepted: Feb. 27th, 2019; published: Mar. 6th, 2019

Abstract

The use of addictive drug is a crisis that the U.S. government has to face. Federal organizations such as the Centers for Disease Control (CDC) are struggling to “save lives and prevent negative health effects of this epidemic”. NFLIS publishes a data report, which addresses “drug identification results and associated information from drug cases analyzed by federal, state, and local forensic laboratories.” After a series of data analysis and statistical chart fitting, according to addictive drug quantity proportion of each year with the corresponding year and continents, by the algorithm of BP neural network, with normalized after year and opioids for rate as input, with binary code after the state number as output, training, we can draw the conclusion: the earliest occurrence of addictive drug in five states is B state. According to the analysis of the model established by us, counties with a large number of drug reports are likely to drive the increase of drug reports in surrounding counties, thus leading to the increase of drug reports in the whole state.

Keywords

BP Neural Network, Fitting

基于人工神经网络分析成瘾性 药物传播

孙源, 杨赞*, 徐柯, 张仪, 李丹

同济大学浙江学院, 浙江 嘉兴
Email: *yangzan953@163.com

收稿日期: 2019年2月11日; 录用日期: 2019年2月27日; 发布日期: 2019年3月6日

*通讯作者。

摘要

滥用成瘾药物是美国政府必须面对的危机。美国疾病控制中心(CDC)等联邦组织正在努力“拯救生命，防止这种流行病对健康造成负面影响”。国家法医实验室信息系统(NFLIS)发布了一份数据报告，其中涉及“联邦，州和地方法医实验室分析的药品案例中的药品鉴定结果和相关信息”。经过一系列的数据分析和统计图表拟合，根据成瘾性药物每年与相应年份和各州的数量比例，采用BP神经网络算法，以年份后的标准化和成瘾性药物作为输入，以二进制代码后的状态数作为输出，训练，我们可以得出结论：五种状态下最早出现的成瘾药物是B状态。根据我们建立的模型分析，得出结论：有大量药品报告的县可能会推动周边县药品报告的增加，从而导致全州药品报告的增加。

关键词

BP神经网络，数据拟合

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

1.1. 问题背景

成瘾性药物的使用是美国政府必须面对的危机。美国疾病控制中心(CDC)等联邦机构正努力“拯救生命，防止这种流行病对健康的负面影响”。国家法医实验室信息系统(NFLIS)发布了一份数据报告，其中涉及“联邦、州和地方法医实验室分析的药品案件的药品鉴定结果和相关信息”。这些数据来自犯罪实验室，它们处理了美国估计每年 120 万州和地方毒品案件中的 88%以上。对于这个问题，我们将重点放在美国五个州的及其所属的 462 个县，我们将 Kentucky 州，Ohio 州，Pennsylvania 州，Virginia 州，West Virginia 州简记为：A 州、B 州、C 州、D 州和 E 州。我们通过分析成瘾性药物的数据特征来达到分析成瘾性药物。

1.2. 问题提出

- 1) 利用 NFLIS 提供的数据，建立数学模型，描述五州及其所属的 462 个县间报告的成瘾性事件随时间的传播和特点。
- 2) 根据我们的模型，确定成瘾性药物在这五种状态中可能开始最早使用的州的编号。
- 3) 描述在什么药品的识别阈值内，政府应该关注什么问题，将会面临怎样的困境？

1.3. 问题分析

这个问题要求我们使用 NFLIS 的数据来描述这 5 个州及其县之间的成瘾性药物的传播情况和特点。根据成瘾性药物的比例在所有药物在州、县事件从 2010 年到 2017 年的同比变化事件，我们可以发现，成瘾性药物变化的趋势是增加或减少的州县，看到他们的传输和特点。

- 针对问题一：我们需要知道这五个州和 462 个县的使用率和使用趋势。确定成瘾性药物使用最广泛的状态，然后确定成瘾性药物最早出现在这五个状态的具体位置。根据各州的年龄分布，我们可以

确定药物被使用的可能性。

- 针对问题二：分析美国政府对此类成瘾性药物持续传播的担忧。成瘾性药物的使用倾向于上升或保持平稳，这对美国政府不利。药物危及生命。如果不下降，就会威胁到政府。
- 针对问题三：分析这些事件发生时药物鉴定的阈值水平。预测它们将在何时何地发生。利用成瘾性的传播趋势，并估计它们在美国和其他国家的概率，我们可以预测它们将在何时何地发生。

2. 算法介绍

BP 学习算法，是对多层网络进行研究而提出的一种有效监督学习算法，该算法基于最小均方差准则，由计算正向输出和误差反向传播组成，通过由比较网络的实际输出与期望输出来不断地调节网络权值，直到收敛为止，如文献[1] [2] [3] [4]所述，网络中每个节点的输入输出存在如下非线性关系：

$$O = \left(1 + e^{-\sum W_{ji} O_{ji} + \theta_j}\right)^{-1}. \quad (1)$$

在该式中， O_{pi} 为模式 P 点至网络节点 j 的输出， W_{ji} 为节点 i 到 j 的连结权， θ_j 为节点 j 的阈值。

定义网络误差函数：

$$\varepsilon = \varepsilon_p = \frac{(T_{pj} - O_{pj})^2}{2}. \quad (2)$$

在该式中， T_{pj} 期望的输出。相应的代价函数为：

$$J = E[\varepsilon] = E[\varepsilon_p] = J_p. \quad (3)$$

网络的最佳权值为使(3)式取得极小值时的解。为此，利用非线性规则中的梯度下降算法来求解最佳权值。训练集中的每个样本输入至网络时，网络的权值都要作相应的调整。其改变量为：

$$W_{ji} = \frac{\partial \varepsilon_p}{\partial w_{ji}}, \quad (4)$$

从而有

$$W_{ji} = \eta \delta_{pj} (1 - O_{pj}).$$

在上式中 η 为学习速率， δ_{pj} 为 j 节点的误差信号。

对于输出层节点 j ，有

$$\delta_{pj} = (T_{pj} - O_{pj}) O_{pj} (1 - O_{pj}).$$

对于隐含层节点 j ，有

$$\delta_{pj} = \sum \delta_{pk} W_{kj} (1 - O_{pj}) O_{pj}.$$

在上式中， δ_{pj} 为节点 j 上一层节点 k 的误差； W_{kj} 为节点 j 到其上一层节点 k 的连接权。

从以上公式可以得出，通过误差反向传播，调整权值，最终的输出就会接近所要求得期望值。这个过程称为训练。当达到所要求的误差，就认为网络已经能在某种程度上能近似表示输入与输出的关系。

3. 基于 BP 神经网络的模型的建立与求解

3.1. 基本假设

为了简化问题，我们做了以下四个假设，并且所提出的假设基本上与事实一致：

- 假设提供的位置数据以及药物分布数据是正确的。
- 假设数据来源方测试了大量药物，药物报告的数据来自其中的药物数量。
- 假设若药物报告的价值低于 50，则认为该县数据量过小而忽略该县。
- 假设五个州的年龄分布稳定，年龄不均对于模型带来的影响可忽略。

3.2. 模型一的建立与求解

通过分析五种状态下成瘾性药物的比值关系，基于预处理数据进行数据拟合处理，如图 1、图 2 所示。

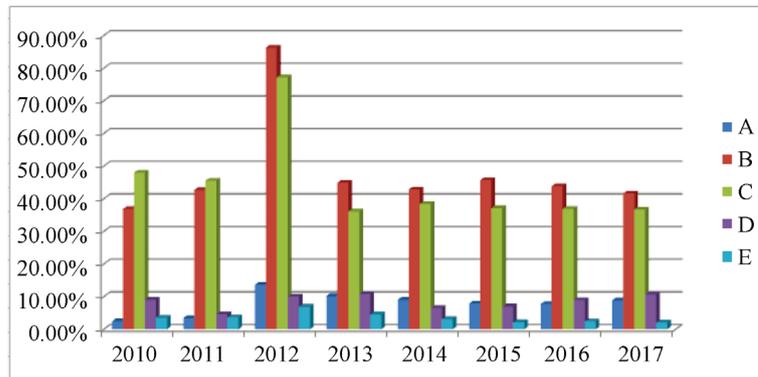


Figure 1. Curve: Addictive drug data analysis histogram
图 1. 成瘾性药物数据分析柱状图

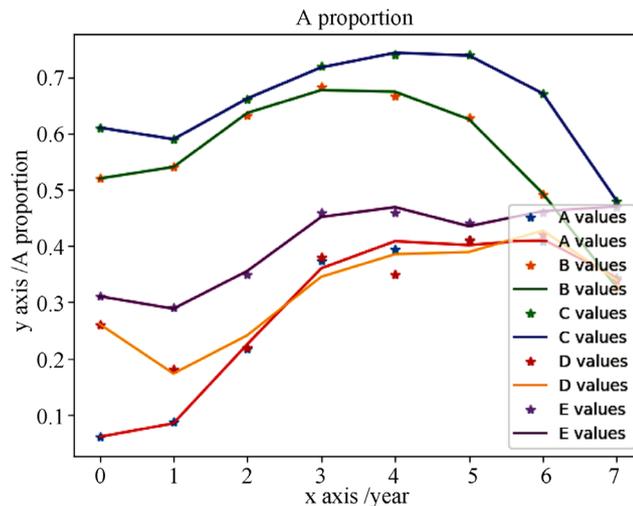


Figure 2. Curve: Addiction drug data analysis scatter plot
图 2. 成瘾性药物数据分析散点拟合图

我们使用基于 Python 的 numpy 与 matplotlib 模块进行六次多项式拟合，所作出的曲线为：

$$\begin{aligned}
 y_1 &= -0.000194x^6 + 0.003489x^5 - 0.02053x^4 + 0.03294x^3 + 0.0569x^2 - 0.04865x + 0.06052, \\
 y_2 &= 0.0001932x^6 - 0.004219x^5 + 0.03562x^4 - 0.1475x^3 + 0.2884x^2 - 0.1521x + 0.5196, \\
 y_3 &= 0.0000625x^6 - 0.001662x^5 + 0.01663x^4 - 0.08224x^3 + 0.1998x^2 - 0.1533x + 0.6101, \\
 y_4 &= -0.0002639x^6 + 0.004843x^5 - 0.03041x^4 + 0.0651x^3 + 0.03116x^2 - 0.1584x + 0.261, \\
 y_5 &= -0.0003194x^6 + 0.006304x^5 - 0.04425x^4 + 0.1266x^3 + 0.1103x^2 - 0.00007051x + 0.3103
 \end{aligned}$$

我们可以通过数据分析图和拟合图得出结论，成瘾性药物随着时间的增加呈增长趋势。

3.3. 模型二的建立与求解

我们根据数据处理的成瘾性药物的药物总比作为神经网络的输入，然后相应的国家数量作为神经网络的输出，我们十隐藏设置，建立 BP 神经网络如图 3 所示，我们在神经网络算法进行数据标准化处理标准。

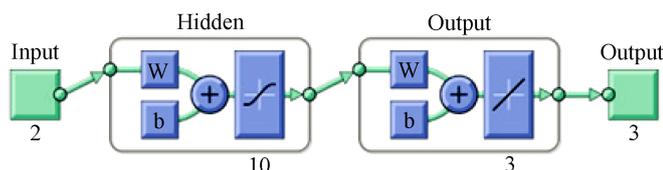


Figure 3. Curve: Schematic diagram of neural network training process
图 3. 神经网络训练过程示意图

经过计算，当我们输入 2010 年和当年的最大药物百分比时，得到的结果是 0.1325、0.8623、0.8635，我们可以推断出是 Pennsylvania 州。然后我们在 2010 年绘制了 Pennsylvania 州的毒品分布图我们得到了所属 Pennsylvania 州的 Lackawanna 县占有的最高百分比。

3.4. 模型三的建立与求解

首先，我们分析了根据表格数据统计显示，我们对数据进行了筛选绘制了统计图，如图 4 所示。

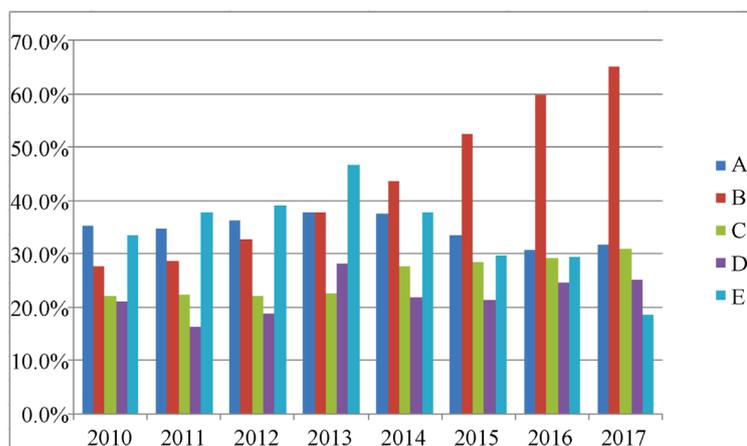


Figure 4. Curve: Schematic diagram of annual growth rate of addictive drugs
图 4. 每年成瘾性药品增长率示意图

经过计算，当我们输入 2010 年和当年的最大药物百分比时，得到的结果是 0.1325、0.8623、0.8635，我们可以推断出 A 州先发生了成瘾性药品传播。此外，我们分析了 B 州，并描述了每个县的毒品分布情况。我们将题目中给出的药物与各县经纬度结合，制作分布图(见图 5)。

从药物分布随年份的分布图 5 可以看出，成瘾性药品随时间的变化较为明显。在成瘾性药品的识别阈值水平内，成瘾性药品会增加其周围地域药物事件的数量。通过数据分析和经纬度比较，我们得出结论：这种失控事件将在 2010 年发生在 B 州某市。毒品报告较多的县容易带动周边县成瘾性药品报告的增加，从而导致全州成瘾性药品报告的增加。如果这种情况继续下去，美国将陷入毒品的困境。

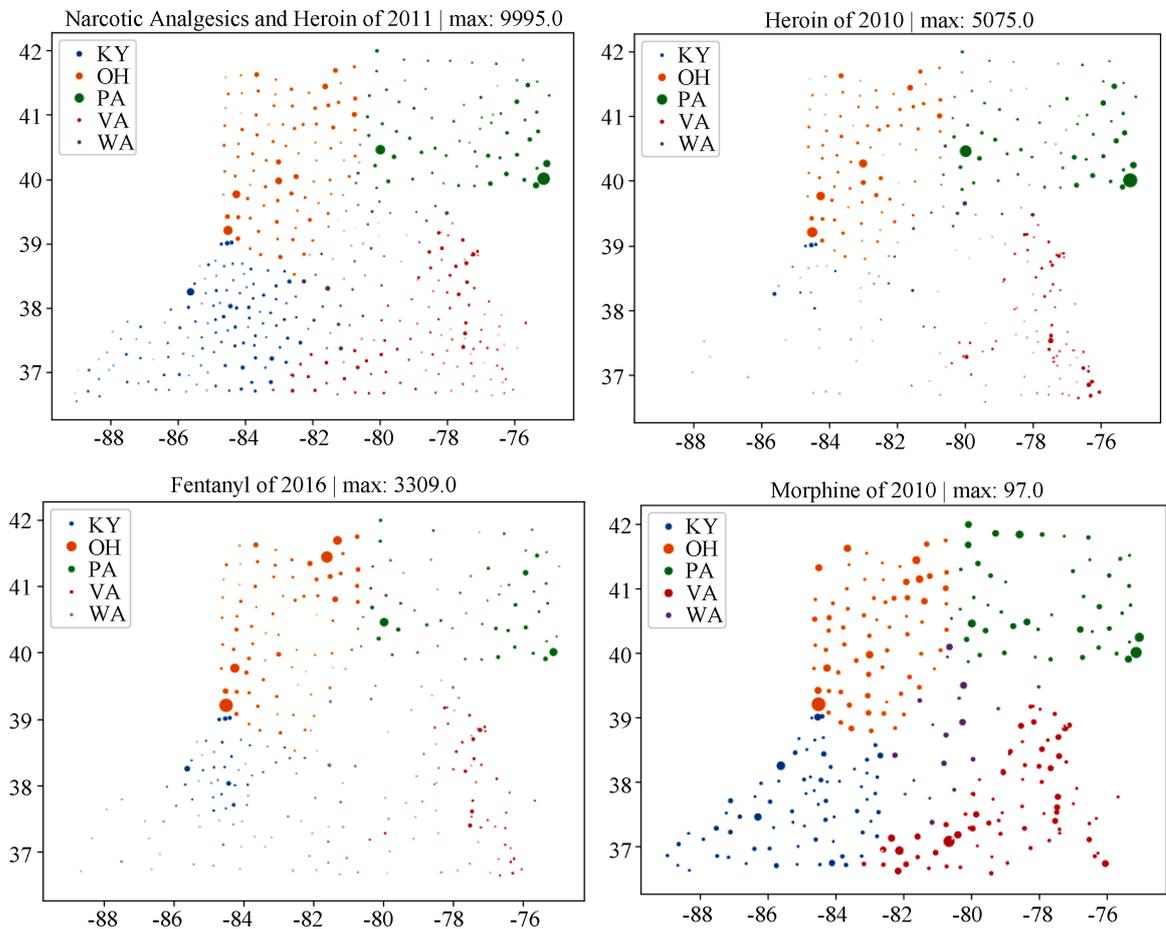


Figure 5. Drug distribution with latitude and longitude
图 5. 毒品分布随经纬度示意图

4. 结论

在本文中，我们建立了 BP 神经网络模型，分析了 2010 年至 2017 年五个州各县的成瘾性药物的药物鉴定计数，2010 年至 2016 年的五个州。我们终于得出了一些的结论。Pennsylvania 州在五个州中成瘾性药物比例最高，五个州最早出现的成瘾性药物是 Kentucky 州某地。有大量药物报告的县可能会推动周边县药物报告的增加，从而导致全州药物报告的增加。

基金项目

同济大学浙江学院第七届教改项目(项目编号: 0118037)

参考文献

- [1] 李文华. 基于人工神经网络的煤矿立井施工工期预测方法研究[J]. 煤炭科学技术, 1999, 27(6): 44.
- [2] 杨荣富. 神经网络模拟降雨径流过程待[J]. 水利学报, 1998(10): 69.
- [3] 约翰·内特, 著. 应用线性回归模型[M]. 张勇, 等, 译. 北京: 中国统计出版社, 1990.
- [4] Kurkova, V. (1992) Kolmogorov's Theorem and Multilayer Neural Networks. *Neural Networks*, 5, 501-506. [https://doi.org/10.1016/0893-6080\(92\)90012-8](https://doi.org/10.1016/0893-6080(92)90012-8)

知网检索的两种方式：

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择：[ISSN]，输入期刊 ISSN：2324-7991，即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入，输入文章标题，即可查询

投稿请点击：<http://www.hanspub.org/Submission.aspx>

期刊邮箱：aam@hanspub.org