

队控号

仅供办公使用 T1

T2

T3

T4

1922154

问题的选择

C

仅供办公使用 F1

F2

F3

F4

2019

MCM /
ICM

汇总表

阿片类药物的研究现状、发展趋势及对策

总结

美国正在经历一场阿片类药物滥用危机，这对美国的发展前景构成了巨大威胁。基于元胞自动机的思想，我们不仅描述了俄亥俄州、肯塔基州、西弗吉尼亚州、弗吉尼亚州和宾夕法尼亚州报告的合成阿片类药物和海洛因案件的传播和特征，而且还制定了应对阿片类药物危机的可能策略。

我们将一个县及其周围最近的 k 个县定义为一个“环境”。基于 KNN 的思想，我们确定与该县“环境”最相似的 m 个“环境”，然后用细胞自动机根据 m 个“环境”的增长率预测该县明年的病例数。同时，我们定义阿片类事件集中指数(CI)，参照 HHI 指数来表征病例的聚集程度。最后，我们获得了五个州合成阿片类药物和海洛因事件的分布情况。案件仍然集中在交通枢纽，有扩散的趋势。海洛因以列克星敦为中心，向西南的肯塔基蔓延，并有扩散到宾夕法尼亚和弗吉尼亚的趋势。根据历史数据和预测，我们确定了每个州的毒品识别阈值水平。2026 年，俄亥俄州将达到 12 万的阈值，使政府难以控制阿片类药物的使用量和传播速度。

为了确定某些社会经济因素是否对阿片类药物使用趋势有显著影响，我们选取 2010-2016 年所有州案例的前 25% 和后 25% 的数据，如果数据通过方差同质性检验，进行方差分析。对未通过方差同质性检验的数据进行相关分析，以确定社会经济因素与阿片类药物使用趋势之间的相关性。最终选择的显著因素是婚姻状况、受教育程度、血统和家中使用的语言。将上述选择的重要因素加入到“环境”相似性考虑因素中，我们得到了一个考虑社会经济因素的修正模型。

基于上述分析，我们制定了一个包含两种行动的策略来应对阿片类药物危机。第一个是给夫妻减税和抵押贷款利率折扣，鼓励人们在法定年龄结婚。另一种是开办低成本的英语语言培训机构，提高非英语母语人士的英语水平。

关键词:阿片类;元胞自动机;集中指数;传播;特征

备忘录

来自:团队#1922154

收件人:总管理员

数据:2019 年 1 月 27 日

主题:如何应对阿片类药物危机

尊敬的总管理员，我们很荣幸地通知您我们在进行数据分析和建模后取得的成果。

首先，我们介绍了 2010 年至 2017 年 5 个州及其县之间合成阿片类药物和海洛因使用的分布和特征。结合提供的数据和收集的经纬度数据，我们注意到阿片类事件的聚集点主要集中在交通发达的地区，且有向周边扩散的趋势。合成阿片类药物在弗吉尼亚州的分布最广泛，在宾夕法尼亚州的分布最集中。海洛因一度有扩散的趋势。然而，也许出于某种原因，这一趋势被遏制了。如今，海洛因在一些州又开始传播，比如弗吉尼亚。

然后，我们预测了 2017 年到 2026 年每个县的合成阿片类药物和海洛因的使用情况。根据预测，合成阿片类药物将在未来蔓延到肯塔基州。而且从预测来看，华盛顿周边县的合成阿片类药物使用量正在增长。

根据我们对提供的数据和 computed 数据的观察，我们认为美国政府正在关注两点：

“阿片类药物的使用应该限制在一定的水平。

- 阿片类药物的传播应控制在一定范围内。

根据历史数据和预测，我们可以识别药物识别阈值水平，预测政府的关注会在何时何地发生。例如，俄亥俄州的门槛是 12 万。政府的担忧发生在 2026 年的俄亥俄州。

通过分析人口普查的社会经济数据，我们注意到一些重要的变量，如婚姻状况、受教育程度、血统和所使用的语言，将影响每个县的阿片类药物的使用。

基于上述分析，我们提出了一项策略，其中包括两项行动。• 给予夫妻税收和房贷利率优惠，鼓励人们在法定年龄结婚。

“开设低成本的英语语言培训机构，提高非英语母语人士的英语水平。

我们的策略可以有效减少阿片类药物的使用。

采取行动 1，阿片类案件将从 257496 件减少到 231073 件采取行动 2，阿片类案件将从 257496 件减少到 225873 件。

以上是我们的研究总结。衷心希望能为大家提供有用的信息。

谢谢！

内容

1 介绍1

 1.1 背景1

 1.2 计划的方法1

2 术语、符号和假设2

 2.1 条款2

 2.2 符号2

 2.3 一般假设3.

3 阿片类事件的传播和特点 3.

 3.1 预处理数据3.

 3.1.1 缺失值处理3.

 3.1.2 地理坐标采集4

 3.1.3 毒品案件分布的概述4

 3.2 基于 CA 模型的阿片类事件传播 4

 3.2.1 介绍方法的想法4

 3.2.2 属性的细胞5

 3.2.3 自我界定的规则5

 3.2.4 浓度指数(CI)7

 3.3 结果与分析7

 3.3.1 传播和特点7

 3.3.2 关心和发生9

 3.4 模型的敏感性分析10

4 模型修改考虑社会经济因素 11

 4.1 预处理数据 11

 以下 4.4.1 数据概述 11

 4.1.2 数据选择及分析12

 4.2 重要因素选择12

 4.2.1 的总体想法准备选择的因素12

 4.2.2 分组12

 4.2.3 双重的过滤器13

 4.3 模型的修改14

 4.4 结果与分析14

 4.5 模型的评估16

 4.6 战略17

 4.6.1 策略的原则17

 4.6.2 行动17

5 的优点和缺点18

 5.1 强度18

 5.2 缺点18

6 结论18

7 参考19

附录20.

有道文档翻译
pdf.youdao.com

1 介绍

1.1 背景

目前，美国阿片类药物成瘾和滥用现象严重。阿片类药物滥用不仅给美国政府带来了沉重的经济负担，也影响了美国劳动力的数量和质量，影响了美国经济的前景。

美国毒品管制局(DEA)办公室的 DEA/国家法医实验室信息系统(NFLIS)每年发布一份关于药物鉴定结果和毒品案件相关信息的报告。具体而言，他们需要：“描述五个州及其县之间长期报告的合成阿片类药物和海洛因事件的传播和特征，并确定五个州可能已经开始使用特定阿片类药物的可能地点。

“从美国人口普查的社会经济数据中分析影响阿片类药物使用或使用的重要因素。

“应对阿片类药物危机的一种可能策略。

大量文献追踪了阿片类药物在美国的滥用情况:例如，Cicero、Inciardi 和 Muñoz[1]根据研究的滥用、转移和成瘾相关监测(RADARS®)系统，专门描述了 2002 年至 2004 年美国阿片类药物滥用的趋势;沃尔科夫、琼斯、爱因斯坦和沃戈[2]分析了引发阿片类药物危机的因素及其进一步演变，以及管理和预防阿片类药物使用障碍的干预措施。

然而，大部分文献并没有根据各县阿片类药物鉴定案例的数据，科学地总结其传播模式和随时间的分布特征，以至于未来的预测无法准确地指示出一种药物鉴定可能传播的时间和地点。此外，过去的工作也未能提出有效的应对阿片类药物危机的策略。

1.2 有计划的方法

基于以上分析，我们提出了如图 1 所示的框架模型，可归纳为以下步骤:

特征与传播

使用 NFLIS 数据和地理数据(纬度和经度)绘制热图和其他可视化图像，并分析报告的合成阿片类药物和海洛因事件在五个州及其县之间的传播和特征随时间的推移。

元胞自动机模型(Cellular Automata Model)

以元胞自动机的思想，即下一时刻的状态是由周围环境和自身状态决定的，结合聚类和 KNN 的思想，构造了一种新的元胞自动机模型。该模型将充分利用历史数据的信息，实现更准确的模拟。

(分析社会经济因素)

我们计划使用统计单因素方差分析和相关分析来寻找对模型有显著影响的社会经济因素，并对模型进行修正。

“识别一个可能的策略

我们将考虑元胞自动机模型的结果和分析中有影响的社会经济因素，然后制定应对阿片类药物危机的可能策略。该模型还将用于验证该策略的有效性。

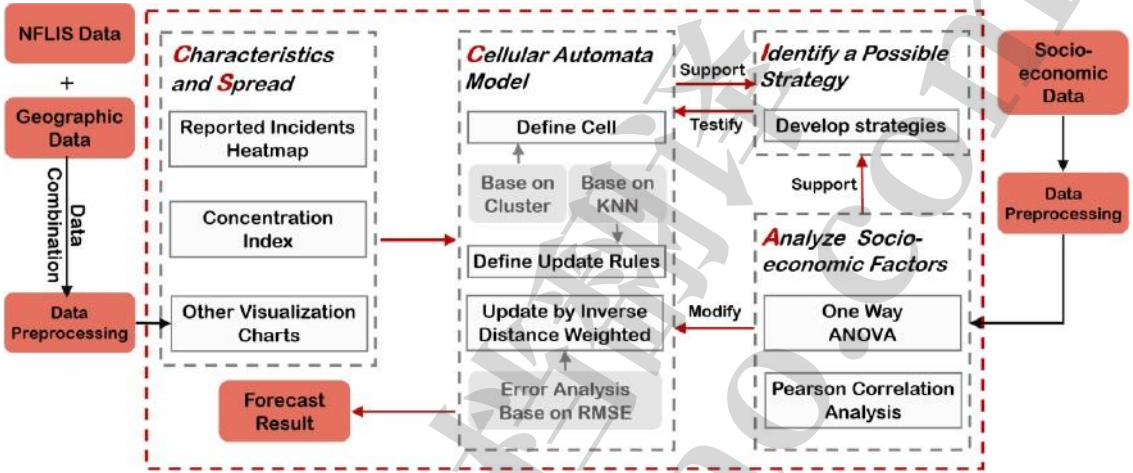


图 1:模型框架

2 术语、符号和假设

2.1 条款

- ▮ 阿片类药物^[3]:医学上主要用于镇痛，包括麻醉，也经常用于非医学上的欣快效果或防止戒断。
- ▮ 海洛因^[4]:一种阿片类药物，因其欣快作用而最常被用作娱乐性药物。无证制造、拥有或出售海洛因通常是非法的。
- ▮ HHI^[5]:赫芬达尔-赫希曼指数(Herfindahl-Hirschman index)是一种浓度的统计度量。例如，它可以用来衡量市场集中度。它的计算方法是将一个市场中所有公司的市场份额平方，然后将这些平方相加。

2.2 符号

表 1:变量描述

象征	定义
	我的 th 县
	与 i 相关的环境(向量) th n 中的县 th 一年
	阿片类药物在美国的使用增长率 th n 的县 th 一年

(,)	参数值为 k 和 m 时，□□h 年的误差
CI	浓度指数
女士	婚姻状况
EA	受教育程度
一个	祖先
LS	家中使用的语言

2.3 一般假设

“假设 1: 一个县阿片类事件数量的变化受周边县影响较大，历史数据能在一定程度上反映阿片类的发展情况。

原因:做这个假设是为了保证我们构建的元胞自动机模型的有效性。

| 假设二:政府即日起至 2026 年不会对阿片类药物进行过度整治，各县阿片类药物变化将遵循 2010-2017 年的历史规律。

理由:之所以做这个假设，是为了在一定程度上保证模型预测结果的有效性。

| 假设 3:本文使用的数据具有一定的真实性和准确性。理由:虽然数据不完整，统计中也存在一些可容忍的误差，但我们做出这样的假设是为了保证有效的解决方案。

| 假设 4:我们的模型不考虑 nfl 数据中没有涉及的县。

原因:我们认为 nfl 数据中没有涉及的县对研究问题的意义不大

阿片类事件的传播和特征

3.1 预处理数据

3.1.1 缺失值处理

第 1 个文件(MCM_NFLIS_Data.xlsx)包含该县 2010-2017 年的大部分药物识别计数，但一些县仍有特定年份甚至年份的数据缺失。我们怀疑该县在缺少这些年份的情况下有药品鉴定计数，但无法确定所鉴定的药品名称。因此，我们将所有已鉴定物质的可变县总计数的缺失值填写如下：

$$X_i = \frac{X_{i-1} + X_{i+1}}{2}$$

(1)

在 \square_i 表示 I 中该县认定的所有物质的缺失总数^[6] 年, \square_{i-1} 表示前一年确认的所有物质的总数, \square_{i+1} 表示下一年确认的所有物质的总数。

注:如果该县有 3 个以上缺失数据, 我们认为该县数据无效。我们放弃填充缺失值, 放弃缺失值。

3.1.2 地理坐标获取

为了了解提交病例的地理分布, 我们从美国城市数据库网站获得了 5 个州:俄亥俄州、肯塔基州、西弗吉尼亚州、弗吉尼亚州和宾夕法尼亚州所有县的经纬度数据^[6]。然后我们用哈弗辛公式计算每个县之间的距离^[7]。假设两个县的经纬度分别为 (\square_1, \square_1) 和 (\square_2, \square_2) , 分别。

$$d = 2r \cdot \arcsin(\sqrt{\sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right) + \cos\varphi_1 \cdot \cos\varphi_2 \cdot \sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)})$$

(2)

其中 d 是两个县之间的距离, r 是地球的半径。

3.1.3 毒品案件分布概况

我们利用各个州的经纬度数据和五个州的麻醉镇痛药(合成阿片类药物)和海洛因的药品鉴定计数数据绘制热图, 对报告病例的分布有一个大致的了解。

以 2010 年报告的病例为例(如图 2 所示), 从热图中大致可以看出, 交通枢纽和沿湖沿海地区的比例较高。

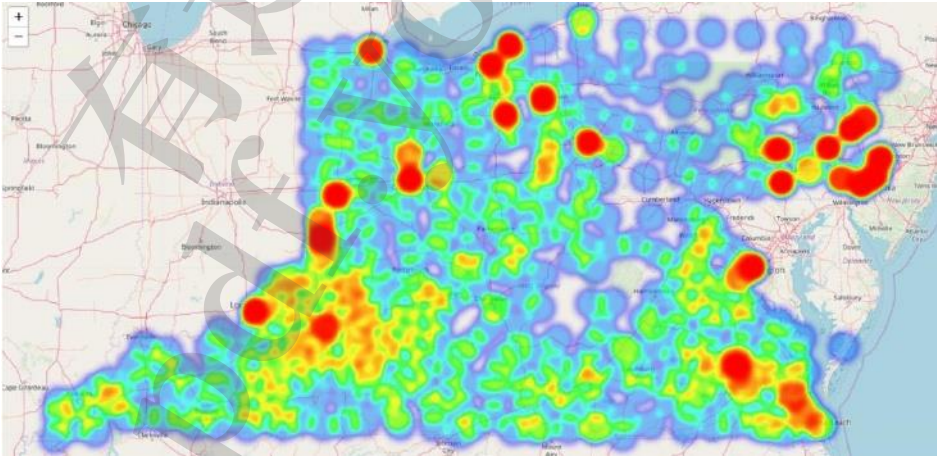


图 2:2010 年药品鉴定案件分布情况

3.2 基于 CA 模型的阿片类事件的传播

3.2.1 方法思想介绍

为了帮助我们了解过去阿片类药物在五个州及其县之间的使用情况, 我们提出了一个模型来模拟过去 8 年阿片类药物的使用情况

各个地区的年代。然后，利用模型的模拟结果来确定五个州中可能已经开始使用特定阿片类药物的任何可能位置。

基于对问题和数据的分析，我们总结出以下挑战：

- “模型应该能够反映每个县阿片类药物使用之间的相互作用。
- “模型应该能够反映每个县的历史发展对其未来的影响。
- 模型必须能够模拟所有县阿片类药物病例数量的变化。

针对这些挑战，我们采用了元胞自动机^[8] (CA)，这是一种网格动力学模型，其中时间、空间和状态都是离散的，具有模拟复杂系统演化过程的能力。元胞自动机是分析[9]传播问题的常用模型。在这种情况下，地图被划分为单元，每个县占用一个单独的单元。一个单元格记录了该县的药品报告总量。细胞的状态根据其当前状态和周围细胞的当前状态进行更新。我们应用自定义的更新规则来模拟我们模型中每个县阿片类药物使用的演变。

3.2.2 一个细胞的属性

在我们的模型中，每个 cell 最多只能代表一个县。一个单元格有 3 个属性：一个整数 c(0 或 1)表示单元格状态，0 表示没有县，1 表示一个县该县当前阿片类案件数量□□。

该县当年阿片类案件数量的变化率

其他两个性质只有在 c 非零时才有意义。在模拟的每一步中，□□都由□□更新。下面介绍自定义规则

部分。

3.2.3 自定义规则

可用于描述每个县阿片类药物使用情况的细胞自动机，关键在于制定接近现实的规则。在这种情况下，我们根据给定的历史数据开发更新规则。

首先，我们通过每个县的经纬度计算出它们之间的距离。然后我们把每个县及其最近的□县作为一组向量，基于聚类的思想我们称之为“环境”。每一个的数学表达式

向量的集合如下：

$$\overrightarrow{C_i^n} = (r_i^n, C_i^n(0), C_i^n(1), C_i^n(2), C_i^n(3), \dots, C_i^n(k))$$

其中 □□□为第 n 年与第 I 个县相关的环境(向量)，□□□为第 Ith 第 n 年县th 年(增长率可为负)。(0)为自身阿片类药物使用量 (1)表示与 I 距离最短的县的阿片类药物使用量th n 中的县th 的一年。可以生成一组向量

基于每个县的年度数据。

于是我们有了 3003 组环境(向量)。我们认为，一个县阿片类案件数量的增长是确定的。在历史数据中可以找到类似的环境来确定该县阿片类案件的增长速度。在寻找相似的环境之前，我们要用环境（向量）做一些事情。我们将环境向量 $(c_1, c_2, c_3, \dots, c_n)$ 以升序排序。如果我们不对环境向量进行升序排序，我们将放大误差。例如,我们认为 $(c_1, 1, 2, 3, 4, 5)$ 应该等于 $(c_1, 1, 5, 4, 3, 2, 1)$ 。但当我们计算它们之间的欧氏距离时，结果表明它们是不同的。

我们可以计算每个环境(向量)之间的欧氏距离。的

欧氏距离计算公式如下：

$$d_{ij} = \sqrt{\sum_{i=0, j=0}^k (c_i^n - c_j^n)^2}$$

(3)

基于 KNN 的思想，我们用公式找到最接近 $(c_1, c_2, c_3, \dots, c_n)$ 的 m 个向量。基于 m 向量中的 $(c_1, c_2, c_3, \dots, c_n)$ ，我们可以通过逆距离加权计算，确定 c_i 县 2010 年阿片类案件数量的年增长率。反距离加权公式如下：

$$w_j = \frac{\frac{1}{d_{ij}}}{\sum_{j=1}^m \frac{1}{d_{ij}}}$$

(4)

在 w_j 为各向量对 c_i 县增长率影响的权重。

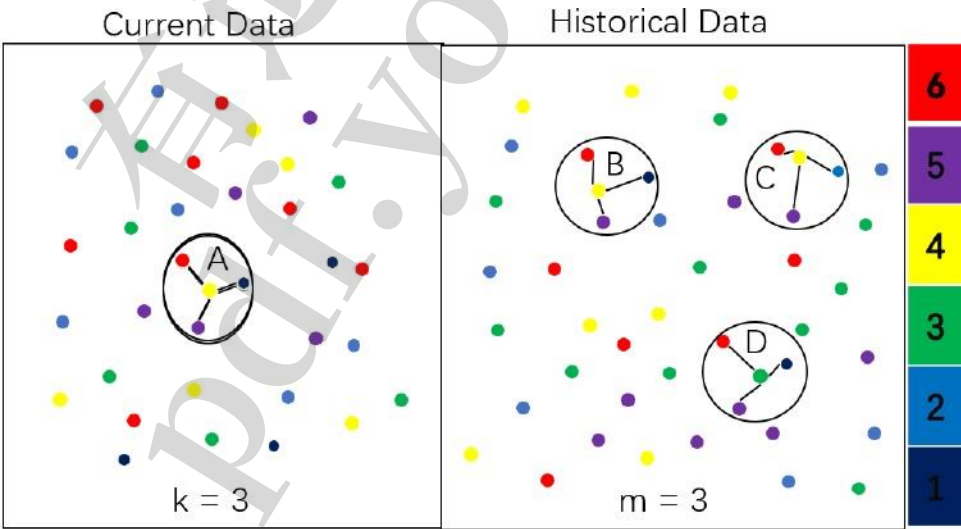


图 3:更新规则 (Updating Rule)

图 3 展示了该规则在实践中的工作原理。点表示县，点的颜色表示不同的值。在本例中，我们选择 $k=3, m=3$ 。我们可以找到环境 $c_i = (c_1, c_2, c_3, \dots, c_n)$ 。从历史数据来看，三者最为相似

模型。2017 年和 2020 年合成阿片类药物的扩散情况如图 4 所示：

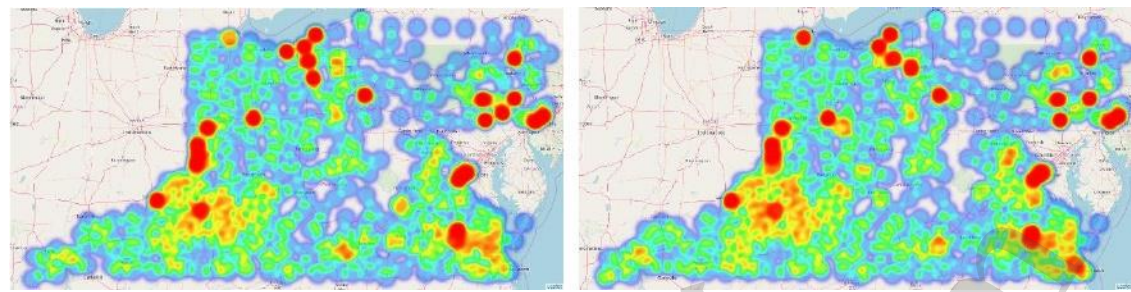


图 4(a) 2017 年合成阿片类图 4(b) 2020 年合成阿片类

通过对比图 3(a)和图 3(b)，我们发现部分区域已经开始形成新的聚集点。与此同时，一些旧的聚集点被转移或消失。聚集点仍主要集中在交通发达地区，有向周边扩散的趋势。根据预测，合成阿片类药物将在未来蔓延到肯塔基州。而且从预测来看，华盛顿周边县的合成阿片类药物使用量正在增长。

2017 年和 2020 年海洛因的扩散结果如图 5 所示。

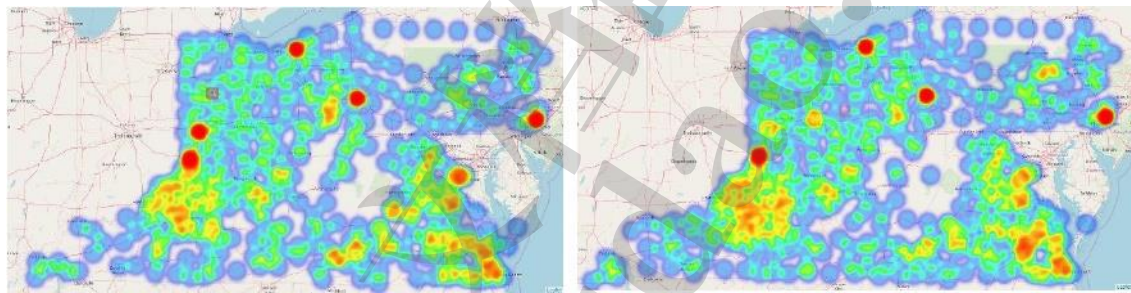


图 5(a) 2017 年海洛因类阿片图 5(b) 2020 年海洛因类阿片

通过对比图 5(a)和图 5(b)，我们发现海洛因按照预测以列克星敦为中心向肯塔基州西南部扩散。根据预测可以看出，聚集点在俄亥俄州逐渐消失。这可能是因为俄亥俄州开始打击毒品。但在宾夕法尼亚和弗吉尼亚，海洛因仍然倾向于在这两个州的广大地区传播。

□特点

为了描绘合成阿片类药物和海洛因的特征，我们根据历史数据计算了它们的浓度指数。

通过计算，我们可以得到合成阿片类药物的浓度指数。结果如图 6 所示。我们注意到浓度指数的值是周期性的。根据浓度指数公式，我们知道，数值越小，扩散程度越强。

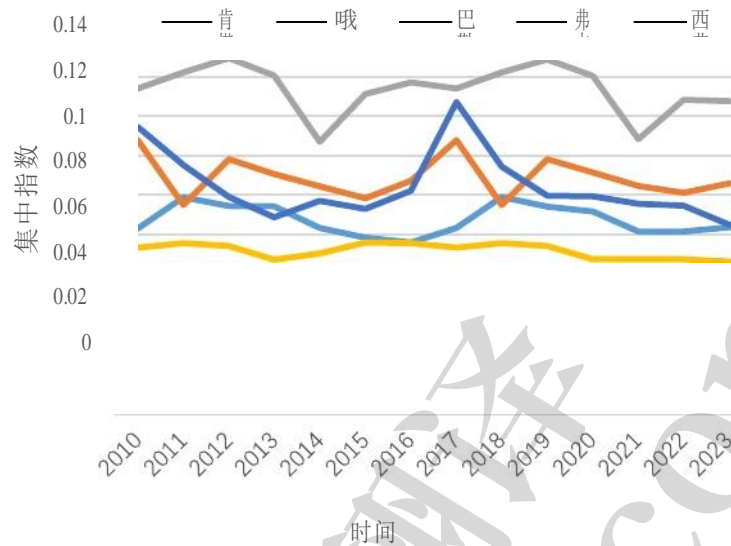


图 6:合成阿片类药物的 CI

从图 6 中，我们发现弗吉尼亚州的合成阿片类药物分布最广泛。此外，宾夕法尼亚州的分布最集中。从图 5 中我们还可以发现，阿片类药物的病例主要发生在宾夕法尼亚州的 metropolis 地区。

海洛因浓度指数的结果如图 7 所示。

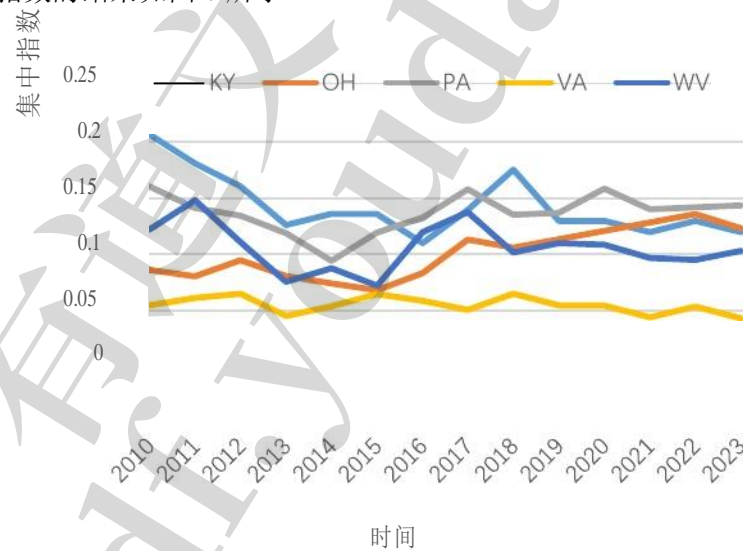


图 7:海洛因的浓度指数

从图 7 中，我们注意到海洛因曾经有扩散的趋势。然而，也许出于某种原因，这一趋势被遏制了。从图 7 可以看出，海洛因在一些州又开始传播了，比如弗吉尼亚。

3.3.2 Concern and occurrence

根据我们对提供数据和计算数据的观察，我们认为美国政府关注的是两个方面：

“阿片类药物的使用应该限制在一定的水平。

- 阿片类药物的传播应控制在一定范围内。

当报告病例数量达到一定水平时，政府将采取行动限制其发展。

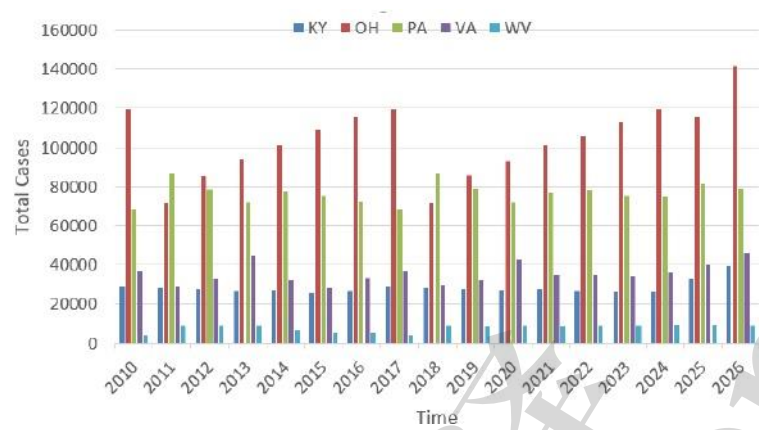


图 8:2010 - 2026 年阿片类药物的使用情况

根据历史数据和预测，我们注意到每个州的报告病例都局限在一定范围内。例如，2010 年至 2023 年，俄亥俄州报告的病例被限制在 12 万例以下。因此，我们可以确定药物识别阈值水平，以预测政府的担忧将在何时何地发生。从图 8 中，我们可以确定俄亥俄州的阈值水平是 120,000。但根据预测，2026 年的事件数量超过了俄亥俄州的阈值水平。那么我们认为政府的担忧已经发生在 2026 年。

3.4 模型敏感性分析

基于一个简单的分析，我们注意到我们的模型对特定参数是敏感的。我们根据提供的数据，用不同的参数值运行模型。以 2010 年的数据作为模型的初始状态。

然后我们将分析我们模型中的参数是如何影响我们的结果的。最后，我们会根据灵敏度分析的结果来确定参数的值。

首先，选取 RMSE(root-mean-square error, RMSE)评估不同参数值下 CA 模型结果与实际值之间的误差;RMSE 公式为:

$$RMSE^y_{(k,m)} = \sqrt{\frac{\sum^n_i (\hat{c}_i - c_i)^2}{n}}$$

(8)

在□□□□吗(,) represents 的误差在 h 年的时候参数的值都是 k

而 m, □□为估计值, □□为实际值。

为了从全局的角度评估 CA 模型的准确性，我们需要考虑每年的 RMSE。最终用来估计模型精度的数学表达式是:

$$RMSE_{(k,m)} = \frac{\sum_{y=1}^a RMSE^y_{(k,m)}}{a}$$

(9)

然后我们可以得到集成的□□□□来描绘模型的精度。

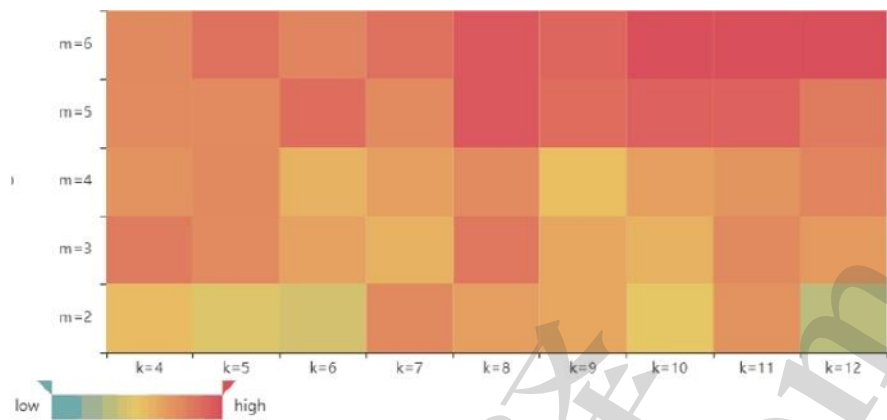


图 9:RMSE 的热图

由图 9 可知，当 k 为 12,m 为 2 时，RMSE 最小，为 104.21。因此，我们确定 k 和 m 的值，具体数值见附录 I。

4 考虑社会经济因素的模型修正

4.1 预处理数据

4.1.1 数据概述

我们目前从美国人口普查局收集了 2010-2016 年每一年这五个州的县的一组共同的社会经济因素。除了“GEO”的三个特殊识别属性。id "， "GEO。id2 和 GEO。Display-label”， 其余属性包括每年社会经济因素数据集中的关系、婚姻状况、祖父母、学历、祖先等。每个属性有四个值(估计值、估计值误差范围、百分比、百分比误差范围)。整体数据帧如图 10 所示。

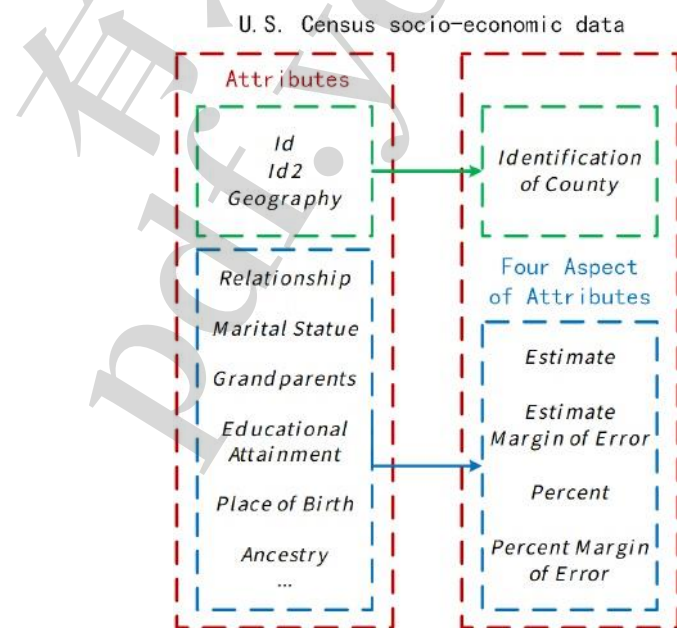


图 10:数据帧

4.1.2 数据选择与分析

这里我们只研究估计值和百分比这两个值，暂不考虑估计误差。在选择了对模型有重大影响的属性后，可以用误差幅度来衡量属性是否有效。此外，我们根据以下考虑，删除部分数据：“在误差范围列中有 ‘(X)’ 的属性

我们可以直接放弃该属性，因为一个 ‘(X)’ 表示估算不适用或不可用。

- 估计数据 (不含百分比)

我们正在考虑的问题是人口普查社会经济数据中某些重要因素对阿片类药物使用趋势的影响。总的来说，某些社会经济因素的估计值之所以会增加或减少，是因为整体形势在变化，此时的使用趋势也会随着整体趋势而上升或下降。因此，我们认为估算的数量不可能是一个有效的因子，而更应该关注其比例结构的变化。

在对数据进行分析后，我们发现 2010-2012 年的数据属性与 2013-2017 年的属性略有不同。对比如下表所示：

表 2:数据对比

一年	县数	属性数量	被研究的属性数量
2010 - 2012	464	599	125
2013 - 2016	463	611	121

基于以上分析，我们将 7 年的人口普查社会经济数据分为 2010-2011 年和 2012-2017 年分别进行分析。

4.2 重要因素选择

4.2.1 因素选择的总体思路

考虑到我们要从大量的属性中找出对每个县的毒品案件有重大影响的属性，我们认为可以按照事件报道的数量进行分组。

采用单因素 ANOVA 分析同一属性在组间的差异，并初步筛选出属性。但是，选定的属性并不一定对案例数量有显著影响。因此，我们考虑使用相关分析来进一步筛选和确定最终的重要因素。

4.2.2 分组

为了讨论不同因素对阿片类药物使用趋势的影响，我们可以分析阿片类药物使用频率高低县的社会经济因素差异。因此，为了更清楚地看到差异，我们选取了两组极端数据。具体描述如下：

表 3:各县药品报告总量

指数	2010 -	2013 -
----	--------	--------

Mean	501.16	534.45
Std	1717.67	1646.28
Min	0	0
25%(Q1)	59	51
50%(Q2)	151	160
75%(Q3)	387	412
Max	33513	21761

我们提取前 25% (即 0~25%) 作为一组数据，最后 25% (即 75%~100%) 作为另一组数据。通过方差分析，我们可以分析两组数据之间的差异，达到初步筛选的目的。

4.2.3 双重筛选

我们知道，方差分析是为了使两组样本数据有显著性差异，样本数据需要满足方差齐性检验。如果数据没有通过同质性检验，可以用相关分析来分析相关性。

ANOVA[10]

方差分析(ANOVA)又称“变异系数”，是由统计专家 R.A. Fisher 发明的，用于对两个或两个以上样本均数的差异进行显著性检验。其基本思想是通过分析不同来源的变异对总变异的贡献来确定可控因素对研究结果的影响。

相关性分析(Correlation Analysis)[11]

相关分析是指对两个或两个以上相关的变量元素进行分析，以衡量两个变量因素之间的密切程度。相关性元素需要有一定的联系或概率，才能进行相关分析。

通过方差分析和相关分析，我们筛选对阿片类药物使用趋势有显著影响的社会经济因素，如表 4 所示：

表 4:重要因素

因素组	解释
婚姻状况 (MS)	百分比;男性 15 岁及以上-未婚(Mn)
	百分比;女性 15 岁及以上-未婚(Fn)
教育 达标(EA)	百分比;本科及以上学历百分比(Ph)
血统(AN)	百分比;阿拉伯 (Ar)
	百分比;希腊 (Gr)
	百分比;爱尔兰 (Ir)
	百分比;意大利(It)

4.3 模型修改

通过分析人口普查的社会经济数据，我们注意到一些变量将影响每个县的阿片类药物的使用。经过前一节的分析，我们发现重要的变量是关于婚姻状况、受教育程度、血统和所使用的语言。这些变量都会对阿片类药物的使用产生影响。相应地，我们随后将它们添加到元胞自动机模型中，以提高我们模型的准确性。具体变量如表 4 所示。

在元胞自动机模型中，迭代规则的基本思想是依靠类似的环境来确定当前的增长速度。因此，我们将 MS、EA、AN 和 LS 引入环境。然而，我们发现一个问题是变量的值非常小。这会导致我们在计算中忽略这些新变量的影响。为了解决这个问题，我们重新定义了变量的公式：

$$MS = \frac{1}{Mn} + \frac{1}{Fn} \tag{10}$$

$$EA = \frac{1}{Ph} \tag{11}$$

$$AN = \frac{1}{Ar} + \frac{1}{Gr} + \frac{1}{Ir} + \frac{1}{It} + \frac{1}{Ru} + \frac{1}{Uk} \tag{12}$$

$$LS = \frac{1}{Lo} \tag{13}$$

然后，我们将这些变量添加到环境中(向量)。每组新向量的数学表达式如下：

$$\overrightarrow{C_i^n} = (r_i^n, C_i^n(0), MS_i, EA_i, AN_i, LS_i, C_i^n(1), C_i^n(2), C_i^n(3), \dots, C_i^n(k))$$

在□□, □□, □□, □□ 是□县的婚姻状况、受教育程度、血统和使用的语言。

最后，我们按照 3.2.3 节解释的更新规则，完成了元胞自动机的迭代。

4.4 结果与分析

我们使用两种显著性分析方法，方差分析和相关分析，来检验人口普查社会经济因素对阿片类药物使用的显著影响。通过方差齐性检验的数据采用 ANOVA，其余数据采用相关分析。

有趣的是，当我们试图关联通过方差显著性检验的因素时，相关性处于较低水平。经过仔细分析，我们发现两种方法选取的数据样本是不一致的;我们在方差分析中提取前 25%和后 25%数量的数据集作为样本，而在相关分析中则使用所有样本。因此，可以解释两种方法得到的结果是不同的，因为这些属性是显著本地化的，当样本量增加时，显著性就减弱了。

我们从以下四个方面分析了显著因素对阿片类药物使用趋势的影响:

“婚姻状况

15 岁及以上从未结婚的男性和女性使用阿片类药物之间存在正相关。

出现这种情况的原因可能是，目前未婚的男性和女性是阿片类药物的主要使用者。因为他们还在追求个人的娱乐和享受，缺乏家庭责任和社会责任。没有抵抗阿片类药物的动力，所以更容易滥用药物。

教育程度

在受教育程度方面，我们发现结果与我们的认知相反。一般来说，我们认为学历越高的人越能意识到阿片类药物滥用的危害，所以他们自然会与之相悖。

相反，研究结果显示，本科及以上学历的人与阿片类药物的使用呈正相关。究其原因，可能是学历越高的人，责任越重，学习和工作压力越大。这导致他们会寻求令人兴奋的阿片类药物来保持注意力集中或缓解压力。

祖先

祖先是阿拉伯人、希腊人、爱尔兰人、意大利人、俄罗斯人和乌克兰的人对阿片类药物的使用有显著影响，并且呈正相关。

我们怀疑这是由于混合人口造成的。当地人口来源复杂，不稳定因素多，是阿片类药物非法贩运的绝佳环境。

家中使用的语言

在美国，如果英语不是很好，你就不能很好地与他人交流。因此，交朋友不容易。所以，在与人际交往的过程中，很容易产生自卑，也往往容易被人批评，染上毒瘾。

传播

通过 CA 模型和改进的 CA 模型求解未来人工合成阿片类药物传播的结果。原模型和修正模型预测的 2020 年合成阿片类药物扩散结果如图 11 所示。

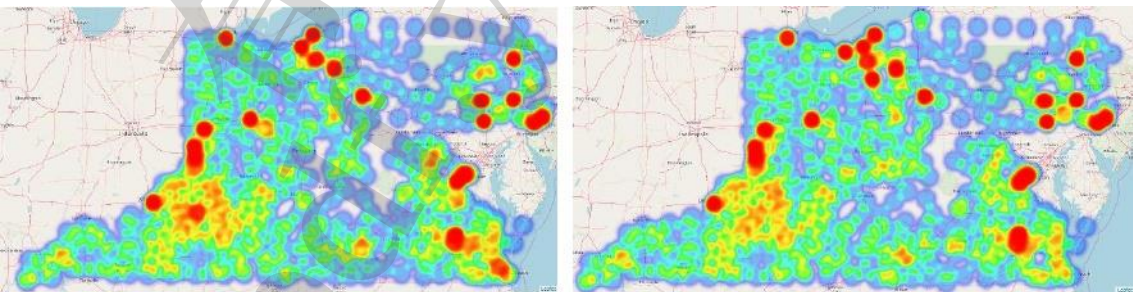


图 11(a) 2020 年合成阿片类原图 11(b) 2020 年合成阿片类修改后

对比图 11(a)和图 11(b)，聚集点仍主要集中在交通发达地区，有向周边扩散的趋势。俄亥俄州出现了更多的聚集点。这意味着我们观察到了更多的传播

规则。从图 11(b)中，我们注意到合成阿片类药物的使用在弗吉尼亚州更为集中。

两种模型预测 2020 年海洛因的扩散结果如图 12 所示：

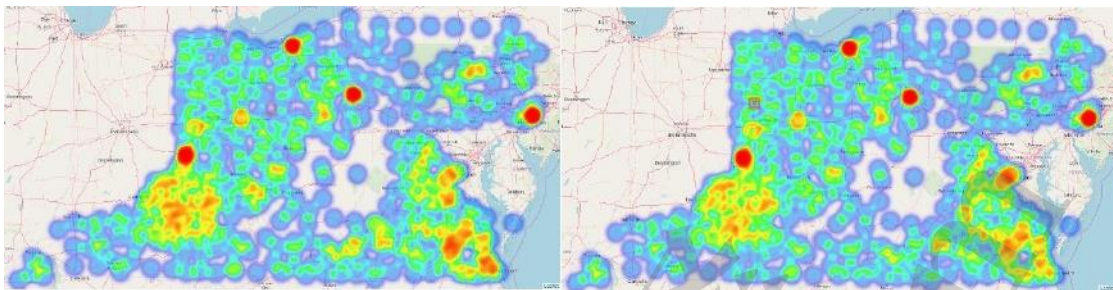


图 12(a) 2020 年海洛因阿片原图 12(b) 2020 年海洛因阿片修正图

对比图 12(a)和图 12(b)，我们发现修改后的模型预测的海洛因在肯塔基州西南部的扩散速度比原模型要慢。两个模型对俄亥俄州和宾夕法尼亚州的海洛因传播有相同的预测。

□特点

通过计算相关公式，我们可以得到修正模型预测的合成阿片类药物的浓度指数。结果如图 13 所示。

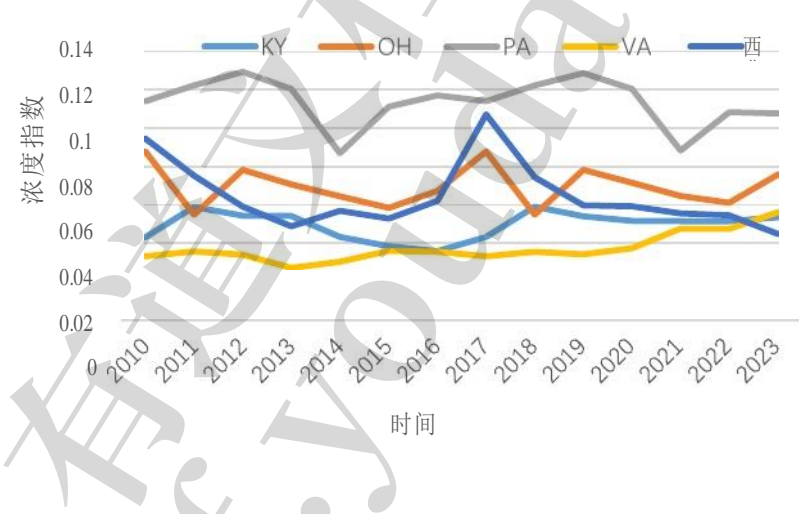


图 13:合成阿片类修饰后的 CI

从图 13 中，我们发现修正模型预测的合成阿片类药物在弗吉尼亚州的分布比原始模型更加集中。他们对宾夕法尼亚和俄亥俄州的浓度指数有相同的预测。

4.5 模型评价

从元胞自动机的建模过程中，我们可以知道，提高模型模拟和预测能力的关键是更有效的更新规则。我们认为在 CA 模型的更新规则中引入新的变量，会使模拟结果更好。

来评估我们的模型是否比未修改的模型表现更好。如 3.4 节所述，我们使用 RMSE 计算修改后模型的性能。在 3.4 节中，我们知道当 k 为 12,m 为 2 时，原始模型的平均 RMSE 最小。有趣的是，k=12 和 m=2 并不是修改后模型的最优组合。

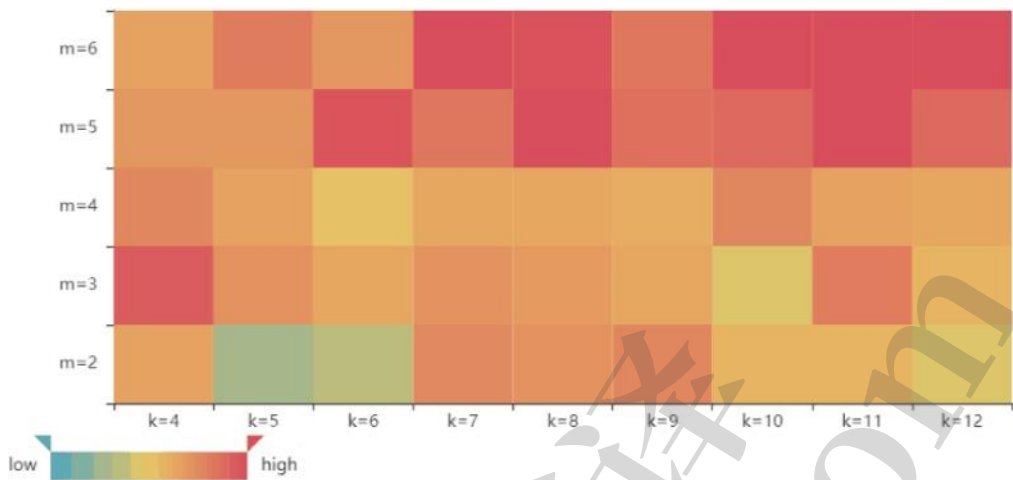


图 14:修改后模型的热图

由图 14 可知，当 k 为 5,m 为 2 时，修正模型的平均 RMSE 最小。最小的平均 RMSE 为 98.70。

4.6 战略

4.6.1 战略原则

- 我们将策略拆分为单独的操作，使用控制变量方法，只考虑一项行动对阿片类药物使用的影响；
- “行为的影响将直接反映在相关变量值的变化中；
- “当变量发生变化时，我们选择在 4.3 节中已经确定的参数组合，模拟 2010 年至 2017 年阿片类药物的使用情况。

基于上述原则，我们提出了以下两种行动，并定量解释了它们有希望的效果。模拟数据为 2017 年累计报告的阿片类事件的模拟结果。真实数据为 2017 年报告的阿片类事件总数。

4.6.2 行动

▮ 行动一:给予夫妻税收和抵押贷款利率折扣，鼓励人们在法定年龄结婚。

这一行动将影响婚姻状况(MS)。假设变化率 α 分别为 5%、10%、15%、20%。根据表 1，我们注意到，当 α 为 5%时，模拟结果超过 2017 年的真实数据。我们认为这可能是造成误差的原因。当 α 为 10%、15%、20%时，很明显模拟数据低于真实数据。说明动作一是有效的。

表 5:比较结果

数据	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 15\%$	$\alpha = 20\%$
实际的数据	257496	257496	257496	257496

▮ 行动二:开办低成本的英语语言培训机构，提高非英语母语人士的英语水平。

这个行动会影响英语说得低于“很好”的比例(Lo)。假设变化率 \square 为-5%，-10%，-15%，-20%。我们从模拟中注意到，当 r 为-10%时，模拟数据超过了 2017 年的实际数据。当 r 在其他值时，模拟数据低于实际数据。说明动作二是有效的。

表 6:比较结果

数据	= - 5%	= - 10%	= 15%	= 20%
实际的数据	257496	257496	257496	257496

综上所述，我们可以得出结论，采取行动一和行动二作为我们的策略可以有效减少阿片类案件的数量。

5 优势与劣势

5.1 强度

基于元胞自动机的思想，巧妙地结合聚类和 KNN 思想对本文模型进行了创新。

“我们的模型充分利用了历史数据的信息，从历史发展的角度模拟了每个县阿片类药物的变化。

虽然基于聚类和 KNN 的模型对 k 值很敏感，但我们尝试了多个不同的 k 值集，并选择 RMSE 最小的 k 值作为模型参数。

我们的模型是可扩展的，可以在构建自身周围的“环境”的过程中添加新的影响变量。

我们严格按照优化理论设定目标。

5.2 缺点

“我们的模型预测过去一两年的结果误差很小，但对长期结果的预测会非常大。

“模型结果可能会过度拟合，因为用于验证模型的真实数据数量太少。

“该模型只考虑了原始数据集中讨论的县，而没有考虑到县向没有滥用阿片类药物的县的扩散。

6 结论

在本文中，我们首先收集了五个州每个县的经纬度数据。在数据预处理后，我们结合热图和浓度指数(CI)，分析了五个州及其县之间以及在一段时间内报告的合成阿片类药物和海洛因事件的传播和特征。接下来，基于细胞自动机的思想，我们建立了基于聚类思想的细胞自动机模型

和 KNN，并对病例数进行了演化，分析了模型的误差。根据模型模拟的结果，我们然后分析了五个州中具体的阿片类药物使用可能是从哪里开始的。然后，我们用方差分析和皮尔逊相关分析(Pearson Correlation Analysis)找到了一些对模型有显著影响的社会经济因素，并对模型进行了修改。此外，我们提出了一些应对阿片类药物危机的可能策略，并通过我们的模型进行了验证。最后，讨论了该模型的优缺点。

7 参考

[1] Cicero, T., Inciardi, J.和 Muñoz, A.(2005)。美国奥施康定®和其他阿片类镇痛药滥用趋势:2002-2004。The Journal of Pain, 6(10), 第 662-672 页。

[2]沃尔科夫, N., 琼斯, E., 爱因斯坦, E.和沃戈, E.(2018)。阿片类药物滥用和成瘾的预防和治疗。JAMA 精神病学。

[3]阿片类药物。(2019)。检索自 <https://en.wikipedia.org/wiki/Opioid>

[4]海洛因。(2019)。检索自 <https://en.wikipedia.org/wiki/Heroin>

[5] Rhoades, S. A.(1993)。herfindahl-hirschman 指数。美联储决议看涨。， 79、188。

[6]美国城市数据库。检索自 https://simplemaps.com/data/us-cities?tdsourcetag=s_pcqq_aiomsg

[7]哈弗辛公式。(2019)。从检索 https://en.wikipedia.org/wiki/Haversine_formula

苏,[8],J. J. ,陈,F. ,谢,P,和通用电气,问:(2016)。局部土地利用竞争元胞自动机模型及其应用。国际地理信息学报, 5(7), 106。doi: 10.3390 / ijgi5070106

[9] Pfeifer, B.(2008)。传染病传播模拟的细胞自动机框架。开放医学信息学杂志, 2(1), 70-81。doi: 10.2174 / 1874431100802010070

[10] faraway, j .(2002)。实用回归和方差分析使用 R.[巴斯]:[巴斯大学]。

[11] Knapp, T.(1978)。典型相关分析:一个通用的参数显著性检验系统。《心理学通报》, 85(2), 410-416。Doi: 10.1037//0033- 2909.85.2.410

附录

k 和 m 不同组合的 RMSE

	K4m2 k4m3 k4m4 k4m5 k4m6 k5m2 k5m3 k5m4 k5m5 k5m6 k6m2 k6m3 k6m4														
的意															
2011	0.702974035	0.660725927	0.651338937	0.641141611	0.631438872	0.689582827	0.654207955	0.640512615	0.631438872	0.618291503	0.686192193	0.64954709	0.618291503		
2012															
2013	6.382522756	6.227991553	6.993836781	7.969931839	9.086992163	4.858846736	7.678869381	7.768503739	8.055460439	9.061818257	6.812678437	6.942830587	7.560941222		
2014															
2015	34.5552495	43.26408309	51.06440469	66.29665969	73.79199646	35.09461072	103.3068956	58.58633778	67.97781758	72.9797526	95.67006971	42.10246686	54.46951201		
2016															
2017	118.5778098	133.7225235	173.3128395	205.9417903	231.547958	106.5340157	138.4434617	189.2682819	211.0507622	232.5925888	118.6640475	148.4046161	177.7839959		
	K4m2 k4m3 k4m4 k4m5 k4m6 k5m2 k5m3 k5m4 k5m5 k5m6 k6m2 k6m3 k6m4														
的意															
2011	166.2801999	242.8589327	214.3292094	224.959704	225.0248144	135.8870985	224.7961209	226.7296587	224.0752928	197.719093	126.253364	197.3480577	178.1509649	181.0268378	283.2203141
2012	223.5731777	220.9499227	228.8786141	229.1343946	209.9527292	206.8617715	99.72905351	203.4261458	104.82919						
2013															
2014															
2015															
2016															
2017															
	K6m5 k6m6 k7m2 k7m3 k7m4 k7m5 k7m6 k8m2 k8m3 k8m4 k8m5 k8m6														
2011	0.610704817	0.618291503	0.69295489	0.67936622	0.675926369	0.654907955	0.647750276	0.699650262	0.668993608	0.636685184	0.64954709	0.651338937			
2012	8.630325456	9.028703258	4.464049611	5.923561267	7.193463752	8.454182011	9.475769615	5.501218344	6.570223144	7.409909575	8.553819741	9.530468867			
2013	71.43347436	92.56316435	27.78748859	42.73791218	54.88778551	78.72786624	86.23140656	45.83951431	90.11657821	92.82732045	107.6172966	116.9397171			
2014															
2015	194.4008247	206.4638508	81.23659653	160.0603252	190.9938429	230.9795297	251.8435245	144.6112464	219.3248115	233.5453915	247.5924823	260.5668601			
2016															
2017															
	K6m5 k6m6 k7m2 k7m3 k7m4 k7m5 k7m6 k8m2 k8m3 k8m4 k8m5 k8m6														
2011	261.6187528	230.2924519	225.2393222	176.7875745	198.9067185	224.7901744	252.079791	201.0809019	246.2813523	222.6504823	286.7257795	283.5586009	253.9350887		
2012	211.5791515	332.8141107	173.4923546	200.0109218	205.2761728	235.100024	223.2626137	239.0649512	199.2557212	274.4817078	261.4978886				
2013															
2014															
2015															
2016															
2017															
	K9m2 k9m3 k9m4 k9m5 k9m6 k10m2 k10m3 k10m4 k10m5 k10m6 k11m2														
2011	0.684493585	0.686194193	0.689582827	0.656685184	0.656685184	0.702974035	0.682788742	0.674199862	0.658457617	0.656685184	0.69295489				
2012	7.120835183	14.1057445	14.0904504	14.20979996	14.54604457	8.450734786	14.08879599	12.28165953	14.34319604	14.52487618	8.444664251				
2013	59.07880636	84.64556571	99.20734802	111.3241032	122.6053612	61.89817942	75.34132123	84.18581679	98.70752162	137.9641266	85.14030652				
2014	127.0279598	166.5318102	182.0729381	208.7606899	251.9515669	103.818786	159.3326116	181.8095843	211.9986421	243.8577292	128.3012735				
2015															
2016															
2017															
	K9m2 k9m3 k9m4 k9m5 k9m6 k10m2 k10m3 k10m4 k10m5 k10m6 k11m2														
2011	189.6825937	191.7753611	161.4448427	259.9986001	268.4550734	145.4702853	174.4068161	200.8947694	274.2048056	299.0286286	215.870939	262.2116446			
2012	193.6968913	128.2049746	246.2202988	235.1875612	156.803166	173.0293453	186.6657714	274.2056883	282.2549797	232.2630743					
2013															
2014															
2015															
2016															
2017															
	K11m3 k11m4 k11m5 k11m6 k12m2 k12m3 k12m4 k12m5 k12m6														
2011	0.668993608	0.668993608	0.653125889	0.644141611	0.702974035	0.684493585	0.672468923	0.667249163	0.658457617						
2012	7.456790838	8.110829044	9.927710741	10.80781466	3.19564097	4.891837071	6.155900007	6.82344785	8.610585993						
2013	68.2665829	79.52509097	129.7653522	145.0459054	31.67462788	60.44919268	74.86483547	83.11639829	107.6360851						
2014	196.4820792	184.6359614	207.0583006	240.8498163	83.36576525	169.4449323	200.8082445	226.0490437	255.054727						
2015	300.2621271	259.1984282	330.2023749	385.1362336	137.3245437	250.5763008	331.2169656	354.8736011	355.0737538						
2016	509.1044035	449.8267136	436.4830278	811.8192583	208.9219191	458.2980677	480.8881546	450.7891992	665.1452319						
2017	506.6644407	505.8198681	831.256539	760.9899331	264.3423647	490.0159053	535.5806926	591.6666851	681.5557088						
	K11m3 k11m4 k11m5 k11m6 k12m2 k12m3 k12m4 k12m5 k12m6														
2011	226.9864883	212.5408407	277.906633	336.4704433	104.2182622	204.9086756	232.8838945	244.8550892	296.2477929	219.3608208					
2012	203.9800991	291.4689724	334.9536692	103.5350184	204.702353	221.4074139	230.0692101	287.340526							
2013															
2014															
2015															
2016															
2017															

有道文档翻译
pdf.youdao.com

有道文档翻译
pdf.youdao.com