Opioid Spread Based on CA-ABM Model

Tao Ji*, Xuehu Zhou and Qianqian Lu
College of Computer Science and Technology, Nanjing Tech University, Nanjing, 211816, China

*Corresponding author e-mail: Jettji0817@gmail.com

Abstract. Issue with opioid always arouses social attention. To deal with the opioid crisis in east America, this paper constructs Cellular Automata-Agent Based Model (CA-ABM) model to analyze the opioid spread of five states and succeed in assessing the present situation along with estimating the future situation. Besides, a strategy has been set up to control the opioid spread. To be specific, with data, this paper summarizes the distribution of two kinds of opioid from year 2010 to 2017, and then establishes a CA-ABM model considering history, neighboring and socio-economic factors to predict the opioid spread next year with highly accuracy. It vividly shows the opioid spread through years. Having a general idea of the spread of opioid, this paper comes up with strategies to ease the opioid crisis, succeeding in offering guidance for those five states.

1. Introduction
When it comes to opioid, diseases and crimes along with concerns about the future of our country may immediately fill our mind. There is no hesitation to say, a decade ago, we would have no idea addressing such issues and what we could do was just letting it go. However, even in modern era, opioid is still an intractable word. Not only many natural factors contribute to drug abuse in America, but economic, government and social elements make a difference.

Nowadays data has penetrated into our lives and work, and has brought great changes. Such tool guides us to work more easily and leads a better life. In commerce, finance and some other fields related to data analysis, data has already become an intelligent assistant. And it is the same case in the field of opioid. With data analysis, intuitive opioid spread can be figured out which directs local government to strictly control it.

2. Opioid Spread
From statistics, we can briefly figure out the changes horizontally, say we have access to the changes among areas. By employing thermodynamic diagrams, we can attain the opioid cases’ vertical changes from the data.

Here, KY stands for Kentucky State, OH stands for Ohio State, PA stands for Pennsylvania State, VA stands for Virginia State and WV stands for West Virginia State.
2.1. Interstate Spread

From the chart, we can see total line is relatively stable with a slight increasing tendency. In terms of the state lines, the case number in PA, KY, WV shows a decreasing trend, case number in VA almost stays in a same level with a little fluctuation while the case number in OH shows a great increasing trend. We can safely conclude, over these years, opioid cases have had a tendency to move from the other four states to Ohio state.

2.2. Instate Spread

From the figure, we can clearly point out the flowing trend of case number in each state:

Ohio: Diffuses from the upper left corner to the upper right corner and presents an explosive upward trend.
Virginia: It’s almost constant with a little fluctuation.
West Virginia: Case number is basically unchanged while the total number of it shows a little downward trend.
Kentucky: The number is basically the same and basically no amplitude through years.
Pennsylvania: The number of the opioid cases in the right side becomes less gradually, and has the circumstance that diffuses to all round, however it doesn’t become a serious situation nevertheless. The total case number has a downtrend.

3. CA-Abm Model
CA-ABM model focuses on helping government supervise and control abuse of opioid so that getting to know the future situation of opioid cases in each state is a must.

This model involves the decision-making interaction of micro subjects in the spatial environment, combining top-down and bottom-up analysis methods together.

![Figure 3. Schematic diagram of various levels in our model](image)

CA is mainly a kind of prediction model while ABM is kind of simulation model. Thus, we establish CA-ABM model to predict the future tendency of opioid cases, which can be expressed as a formula:

\[ E(t) = f(Z_t, A_t, t) \]

In it, \( Z_t \) is the environment of research area at time \( t \), where will be composed by Layer CA model; \( A_t \) is the collection of each microscopic decision-making subject at time \( t \), which will be composed by ABM model; \( t \) is the moment simulated in the model.

3.1. CA Model

3.1.1. CA Model Establishment. CA has two layers formulated below:

\[ Z_t = \{\text{history, neighbor}\} \]

Among it, history stands for historical condition, and neighbor stands for neighborhood condition. Specifically, standard CA is composed of 4 parts: cell, state, neighborhood and conversion rule. Here, we adopt Logistic Regression method as the conversion rule and it can be expressed like follows:
\[
\text{Case}_{i,j} = A \cdot \text{Case}_{i,j-1} + B \cdot \frac{\mu_{j-1} \cdot \lambda_i}{\text{Sum}(\lambda_i)} + \theta
\]

Index \(i\) varies from 1 to 5, which each stands for one state, \(j\) stands for year varying from 2010 to 2017. \(\text{Case}_{i,j}\) represents the case number in state \(i\) of year \(j\).

In the formula,
- \(\text{Case}_{i,j-1}\) reflects the history part, which means the case number in state \(i\) of year \(j-1\), the year before \(j\);
- \(\mu_{j-1} \cdot \lambda_i / \text{Sum}(\lambda_i)\) reflects the neighbor part, which means the case number in all neighboring states of state \(i\) of year \(j-1\). \(\mu\) and \(\lambda\) are both in form of matrix.

Index \(\mu\) is a series of \(5 \times 1\) matrices storing the case number of each state where its subscript determines its year:

\[
\mu_j = \begin{bmatrix}
CN_j(KY) & CN_j(OH) & CN_j(PA) & CN_j(VA) & CN_j(WV)
\end{bmatrix}
\]

Index \(\lambda\) is a set of \(1 \times 5\) matrices describing the relationship between two states using ‘0’ or ‘1’ whether state \(i\) is bordering with the other states.

\[
\lambda_{KY} = \begin{bmatrix}
0 \\
1 \\
0 \\
0 \\
0
\end{bmatrix}, \quad \lambda_{OH} = \begin{bmatrix}
1 \\
0 \\
1 \\
0 \\
0
\end{bmatrix}, \quad \lambda_{PA} = \begin{bmatrix}
0 \\
1 \\
0 \\
0 \\
0
\end{bmatrix}, \quad \lambda_{VA} = \begin{bmatrix}
0 \\
0 \\
0 \\
1 \\
1
\end{bmatrix}, \quad \lambda_{WV} = \begin{bmatrix}
1 \\
1 \\
1 \\
1 \\
0
\end{bmatrix}
\]

Therefore, the product of \(\mu\) and \(\lambda\) is an integer which shows the total case number happening in the bordering states of state \(i\) in year \(j-1\). After that, we divide it by the number of the states who are the bordering states of state \(i\) expressing as \(\text{Sum}(\lambda_i)\), the sum of matrix \(\lambda_i\).

Totally, we will get \(5 \times 7\) raw data. In order to consider the interaction between states, we decide to put them together to go on Logistic Regression instead of going on regression by state.

3.1.2. CA Model Outcome. Having such a predict model, we need to solve the coefficients \(A\) and \(B\). We will employ Logistic Regression method. Usually, Logistic Regression method analysis is used to obtain the weight of independent variables.

The result is shown below:

**Table 1. Regression Statistics**

| Regression |  |
|------------|---|
| Multiple R | 0.9847 |
| R Square   | 0.9697 |

**Table 2. Regression parameters**

| Coefficient |  |
|-------------|---|
| \(\theta\)  | 904.5441 | 0.8043 |
| \(A\)       | 1.0327   | 2.5958*10^{-24} |
| \(B\)       | -0.0415  | 0.5448 |

3.2. ABM Model

The operating mechanism of the multi-agent system is to realize the global behavior or state change through the internal decision-making communication and behavior interaction of the observant subject. Although each micro subject is relatively simple, it can cause complex changes in the whole world through the decision-making communication and behavior interaction between micro subjects.
3.2.1. **ABM Model Establishment.** As a spatial entity with autonomous ability and capable of making relevant decisions, agents are used to represent each micro-decision-making subject in the model. An Agent can represent an individual or a group of individuals. In this model, we first set 3 agents and their decisions interact with each other:

\[ At = \{\text{governor agents, environment agents, intrinsic factor agents}\} \]

This model takes the changing rate of the case number as research object. We don’t use the raw data directly because the absolute value makes no sense for different states have different population. Thus we switch to adopt changing rate as our variable. Here, the governor and environment both can have an independent effect on the changing rate. However, at the same time, they will also interact with each other to make a difference together which can be expressed as:

\[ At = \alpha \cdot E_{\text{Governor}} + \beta \cdot E_{\text{Environment}} + \gamma \cdot E_{\text{Intrinsic Factors}} + \theta \]

In this model, stochastic utility model and discrete selection model are adopted to study the internal mechanism of opioid development unit selection decision.

The working mechanism is shown below:

![Figure 4. Working mechanism of ABM model](image)

3.2.2. **Governor Agents’ Decision Behavior.** Government, as an agent doing macro-control, has to focus on which kind of opioid deserves special care. That is to say, government needs to study the influence of each kind of opioid. Here, we divide all opioid into two parts: Heroin and Synthetic Opioid.

\[ E_{\text{governor}} = f ( P_{\text{Heroin}}, P_{\text{Synthetic Opioid}} ) \]

Both these two parts are in form of changing rate. So, first we need to calculate the rate of each kind of opioid change based in the given data. Due to the facts that each state’s government just need to care the change of its own state, we need to calculate the weight of the two parts state by state.

The results are:

**Table 3. Regression parameters of governor agents**

|        | OH    | VA    | PA    | KY    | WV    |
|--------|-------|-------|-------|-------|-------|
| Intercept | 0.034 | -0.033| -0.075| 0.023 | -0.033|
| \( W_{\text{Heroin}} \) | 0.272 | 0.167 | 0.434 | -0.010| 0.080 |
| \( W_{\text{Synthetic Opioid}} \) | 1.300 | 0.025 | 0.218 | 0.365 | 0.737 |

*W represents the weight of each kind of opioid
3.2.3. Environment Agents’ Decision Behavior. A certain state is usually bordered with several states and its neighbors often play an important role in the opioid diffusion process, which largely affects the case number of neighboring states. However, in this agent, we have to stress that if we just take neighboring changing rate as the only variable, the result will be unimaginably bad. The reason is the case number change of a certain state can’t be separated with itself, thus, the final expression should be:

\[
\text{Environment} = f (P_{\text{itself}}, P_{\text{neighborhood}}, \gamma)
\]

In this agent, we take changing rate as research objects. How case number change in its own state and how case number change in its neighboring states together contribute to the case number change next year. We need to employ the matrix \( \lambda \) in CA model again.

To calculate \( P_{\text{neighborhood}} \), we need to employ a similar method to CA model:

\[
P_{\text{neighborhood}} = CR \cdot \lambda
\]

In the expression, \( CR \) is the case number changing rate matrix like:

\[
CR_i = [CR_i(KY) CR_i(OH) CR_i(PA) CR_i(VA) CR_i(WV)] (i \in [1, 8])
\]

Still, we need to solve the weight of the two elements state by state and the result is:

**Table 4.** Regression parameters of environment agents

|          | OH   | VA   | PA   | KY   | WV   |
|----------|------|------|------|------|------|
| Intercept| 0.034| -0.032| -0.075| 0.022| -0.033|
| *WHeroin*| 0.271| 0.167| 0.434| -0.009| 0.080|
| *WSynthetic Opioid*| 1.299| 0.025| 0.218| 0.365| 0.737|

*W represents the weight of each kind of opioid

3.2.4. Intrinsic Factor Agents’ Decision Behavior. The number of attributes listed in the census data is too large to analyze one by one. Therefore, we have to extract and summarize main factors.

We first go through all attributes and comprehend their meaning. During the process, we delete the irrelevant attributes. Afterwards, we summarize relevant attributes into 9 main factors.

The factors are listed below:

**Table 5.** Main factors collected from census form

| Symbol | Factor                                |
|--------|---------------------------------------|
| f1     | Family—child                         |
| f2     | Family—single parent                 |
| f3     | Man—single                           |
| f4     | Person—divorced                      |
| f5     | Woman—premarital pregnancy           |
| f6     | Person—undereducated                 |
| f7     | Person—elderly                       |
| f8     | Person—disabled                      |
| f9     | Person—Migration                     |

Totally, we have 9 attributes. In order to decide their contribution, we decide to employ Logistic Regression method to analyze their weight.
To avoid the effect of the population base, we turn our aim at the changing rate of each factor through years. After deleting the deficient data which missing some data in some year, we are able to reach the relevant rate of each factor:

| Factor | Intercept | factor1 | factor2 | factor3 | factor4 |
|--------|-----------|---------|---------|---------|---------|
| P-vale | 0.1335    | 0.0968  | 0.2774  | 0.8988  | 0.1271  |
| Factor | factor5   | factor6 | factor7 | factor8 | factor9 |
| P-vale | 0.1084    | 0.1108  | 0.9330  | 0.1572  | 0.0337  |

According to P-value, we select 6 out of 9 factors as main factors. Combining with coefficient, we could give the following relationship:

• The more families with children and more migration people, the lower increasing rate of case number;
• The higher divorce rate, higher premarital pregnancy rate, higher undereducated rate and higher disability, the lower increasing rate of case number.

3.2.5. ABM Model Outcome. Based on the two agents, in this part we need to focus on the interact of these two parts. Now we can obtain the result of each agent’s effect on the case number change in a certain state. Thus, we can study the interact by analyzing the accuracy of the two agents. What’s more, we need to combine all states’ data together as a big data set in order to better consider their interactive effects.

We adopt Multiple Linear Regression and the result is:

| Statistics |  |
|------------|---|
| Multiple R | 0.99941 |
| R Square   | 0.99881 |

From the p-value, we can say environment plays a relatively significant role.

3.3. CA-ABM Integration

The cellular automata layer and the multi-agent layer respectively obtained a set of transfer rules. Finally, the combination of the two rules was used to construct the CA-ABM model. There are two ways of combining them:

• Using CA independent first to predict the possible case number, and then use ABM model to test whether it’s of highly possible;
• Take CA and ABM as a combined body to predict the future value.

In the real world, opioid jointly promotes its expansion due to the decision making interaction between self-organization and micro subjects, so the second method is adopted in this study and expressed by the following formula:
\[ CN_{\text{next year}} = \gamma \cdot (\xi \cdot CN_{CA} + \lambda \cdot CN_{ABM} + \varphi) \]

In the formula, \( CN \) refers to the case number. \( \xi \) and \( \psi \) are weights of the certain part while \( \varphi \) is the random control coefficient, which can be employed to describe the randomness of case number change. The final result is:

**Table 9. CA-ABM model parameters**

| \( \xi \)  | 0.0114 |
|-----|--------|
| \( \psi \) | 1.01113 |
| \( \gamma \) | 1.1658 |
| \( \varphi \) | 16.0303 |

### 4. Analysis of Opioid Future Spread

#### 4.1 Threshold Plan

The paper also points out the future tendency of opioid and gives a practical plan to solve the problem when a state should be paid attention to.

Clearly from the case number of each state by year, we can see case number of some states is in upward tendency while others’ is in downward tendency. We believe it’s impossible for case number in a state to switch from upward trend to downward trend in a while. Thus, we put forward several rules according to different situations:

- For upward trend state, we can’t require it directly turn to downward trend state, but its increasing rate should show a downward trend;
- For downward trend state, firstly we require it can’t change to a upward state; what’s more, it’s decreasing rate shouldn’t show a downward trend;

Anchored on the rules, we take Derivative of Case Number Changing Rate as our indicator; also it’s the same meaning with Second Derivative of Case Number (\( CN'' \)). The indicator works like follows:

**Table 10. CA-ABM model parameters**

| \( CN' > 0 \) | \( CN'' \geq 0 \) | Concerned |
|--------|----------------|----------|
| \( CN'' < 0 \) | Not Concerned |
| \( CN' < 0 \) | \( CN'' \geq 0 \) | Not Concerned |
| \( CN'' < 0 \) | Concerned |

#### 4.2 Future Data

In our CA-ABM model, if given specific data of this year, the model can predict the case number of next year with highly accuracy. However, what we need is a long term opioid spread trend, therefore, the data we need is in stringent specification. Simply to say, we need to simulate the case number of heroin and synthetic opioid respectively. With it, we can calculate the governor agent in ABM model, and the rest part, CA model and neighbor agent in ABM model, can be carried out easily.

To start with, we use ARIMA model to simulate the case number of heroin and synthetic opioid respectively.

Particular in time series analysis, fitting autoregressive integrated moving average (ARIMA) model to time series data can either better understand the data or better predict future points in the series. In ARIMA, we employ BIC model which is a criterion for model selection among a finite set of models to help it set the order.

#### 4.3 Concerns

Based on our threshold plan, we can divide these states to several types. Compare the data in column \( CN'' \) with 0, we can reach to the conclusion that which state deserves concerns.
5. Conclusion
In the paper, aiming at easing the opioid crisis in the east US through data mining method, we build a practical model to simulate the opioid spread situation based on the data from 2010 to 2017. In order to solve the problem, we will employ CA-ABM model, ARIMA and Logistic regression method. Through it, we have a chance to offer suggestions for local government.

References
[1] Wang Ruiliang, Simulation study of urban expansion, cultivated land protection and wetland loss based on ABM[D], Zhejiang University, 2013
[2] https://en.wikipedia.org/wiki/Autoregressive_integrated_moving_average
[3] https://en.wikipedia.org/wiki/Bayesian_information_criterion