

УДК XXX.XX

# Моделирование распространения инфекционных заболеваний на графах с помощью Python и AnyLogic

**А. В. Консманов**

Крымский федеральный университет им. В. И. Вернадского, *e-mail*:

*alexkoshernosi@gmail.com*

**Аннотация.** Рассматривается задача построения двух моделей, описывающих распространения лихорадки Эбола. Для построения моделей использовались два различных подхода, реализующие схожие идеи. Модели обучались на одной популяции и тестировались на другой. Анализ точности аппроксимации проводился по графикам и показывает, что модели подобрано верно, работают с разной точностью и могут быть использованы для предсказания характера вспышек лихорадки Эбола.

**Ключевые слова:** Математическое моделирование, лихорадка Эбола, инфекционные заболевания, Python, AnyLogic.

## Modeling the spread of infectious diseases on graphs using Python and AnyLogic

**A. V. Konsmanov**

V. I. Vernadsky Crimean Federal University.

**Abstract.** . . . .

**Keywords:** . . . .

**MSC 2010:** XXAXX

## Введение

Достигнутые на сегодня успехи в области борьбы с инфекционными заболеваниями не дают эффективного метода полностью предотвратить эпидемию, однако, используя прогностические математические модели, описывающие качественные и количественные характеристики очагов заражения, можно значительно снизить количество заболевших в первые дни вспышки и, тем самым, избежать превращения вспышки в эпидемию. [?, с. 202]

Основные задачи, возникающие при исследовании и прогнозировании распространения инфекционных заболеваний: исследование скорости и масштаба распространения заболевания, последствий инфекционной вспышки, пространственные масштабы заражения, затраты средств и ресурсов, необходимых для изоляции и лечения больных. [?, с. 203] Среди моделей была выбрана мультиагентная модель на графах. Данный класс моделей позволяет учитывать характеристики пространства имитируемой области и отдельных представителей популяции.

## 1. Стохастическая модель

За основу взят мультиагентный подход, где агенты распределены по локациям, которые в случайном порядке посещаются зараженными агентами. Между здоровыми и больными агентами возможно событие инфицирования, вероятность которого описывается эмпирической формулой

$$P_i = 1 - e^{\tau N_r \ln(1 - S_i \rho)} \quad (1.1)$$

, где  $P_i$  – вероятность инфицирования  $i$ -го агента,  $\tau$  – коэффициент расстояния, важный при передаче заболевания на большие дистанции, например, при межконтинентальных перевозках, (в данной модели предполагается нейтральным по умножению, т.е.  $= 1$ ),  $N_r$  – кол-во зараженных данным заболеванием (в данной локации),  $S_i$  – восприимчивость индивида  $i$  (индивидуальный параметр, требующий итеративной калибровки, в данной модели индивиды популяции имеют распределение, близкое к гамма-распределению),  $\rho$  – допустимость, т.е. вероятность заражения одного индивида другим за фиксированное время (обычно, 1 минута, в данной модели – одни сутки, т.е. время пребывания агента в одной локации).

Переходы по “больным” состояниям, описанным ниже в SEIFDR-модели, представляет собой пуассоновский процесс.

Реализация: Python3 + SciPy + GraphX(библиотека для графов) + Matplotlib(визуализация) + Yandex Static API (визуализация на карте).

Преимущества: простота понимания модели, наличие большого количества готовых решений во всей предметной области, алгоритм допускает параллельную реализацию, скорость работы модели в целом выше, чем у следующей, т.к. отсутствует задача численного решения СДУ, исходный код модели, за исключением ресурсных файлов и выходных файлов визуализации, «легче», чем у следующей модели.

Недостатки: отдельные фрагменты требуют долгой реализации, готовая модель требует длительной итеративной калибровки для получения результатов, приближенных к реальным, невозможно заранее установить, будет ли модель выполнять прогностическую функцию или является лишь «подогнанной» к уже произошедшим вспышкам, что требует тестирования на нескольких населенных пунктах; стохастическая природа модели требует мощных и надежных генераторов случайных чисел.

Данная модель уступает в точности следующей, что заметно на графиках.

## 2. Дифференциальная мультиагентная модель на графах

За основу взята модель Кермака-МакКендрика (SIR-модель), модифицированная до SEIFDR-модели. В основе данного подхода лежит формула, разработанная Кермаком и МакКендриком в 193 г. для SIR-модели. Как и в прошлой модели, агенты расположены в локациях, однако в этой модели, агенты могут перемещаться с некоторой динамикой и активно контактировать между собой. Изначальная SIR-модель предполагает, что агенты могут быть в трех состояниях: Suspected –

предполагаемое инфицирование или инкубационный период, *Infected* – агент заражен, *Recovered* – агент выздоровел и приобрел иммунитет.

Данная модель допускает модификацию до SEIFDR, где *Exposed* – люди, находящиеся в инкубационном периоде, *Funeral* – погибшие, еще захороненные, все еще источник инфекции, *Died* – погребенные, и описывается СДУ:

$$\begin{cases} \frac{dS}{dt} = -\beta_I SI + \beta_F SF \\ \frac{dE}{dt} = \beta_I SI + \beta_F SF - \alpha E \\ \frac{dI}{dt} = \alpha E - I\gamma_D((1 - \delta_D) + \gamma_D\delta_D) \\ \frac{dF}{dt} = \gamma_D\delta_D I - \gamma_F F \\ \frac{dR}{dt} = \gamma_I(1 - \delta_D)I + \gamma_F F \end{cases} \quad (2.1)$$

, где  $\beta_I$  и  $\beta_F$  – интенсивность контактов между людьми и на похоронах,  $\alpha$  – интенсивность инкубационного периода,  $\gamma_I$  интенсивность выздоровления,  $\gamma_D$  интенсивность умирания агентов,  $\gamma_F$  – интенсивность похорон,  $\delta_D$  – вероятность смертности.

Реализация: в качестве имитационной среды выбран AnyLogic, т.к. данное ПО совмещает простоту и вариативность при построении модели, предоставляет широкие и мощные инструменты для численного решения СДУ.

Преимущества: высокая точность дифф.моделей, возможность шире представить воздействующие факторы, более гибкая и настраиваемая модель, наличие готового имитационного ПО, наличие в имитационном ПО встроенных численных методов решения ДУ и СДУ, имеющих гарантированную сходимость, аппроксимацию, устойчивость.

Недостатки: высокая вычислительная трудность в терминах времени и памяти, необходимость многократного повторения эксперимента для подбора точного значения параметров, имеются проблемы, как и в стохастической модели, есть зависимость от генераторов случайных чисел, что порождает аналогичные проблемы.

Данная модель намного лучше аппроксимирует динамику заболеваемости и смертности, что заметно на графиках, представленных ниже.

### 3. «Обучение» и калибровка моделей

После введения параметров и их примерной настройки по предварительным экспериментам, проводился процесс калибровки. Для этого на каждое изменение параметра приходится 20 итераций. [2, с. 190] [3, с. 873] Для выбора оптимального множества параметров используется минимизация целевой функции, которая рассчитывается следующим образом: пусть каждая точка графика, описывающего реальные показатели заболеваемости и / или смертности, обозначается парой координат  $X_i^{\text{целевое}}, Y_i^{\text{целевое}}$ , а каждая точка графика, описывающего работу модели,

–  $X_i^{\text{моделируемое}}, Y_i^{\text{моделируемое}}, i = \overline{1, n}$ . Тогда целевая функция имеет вид

$$Dev(x) = \sum_{i=0}^n \sqrt{(X_i^{\text{целевое}} - X_i^{\text{моделируемое}})^2 + (Y_i^{\text{целевое}} - Y_i^{\text{моделируемое}})^2} \quad (3.1)$$

, что с учетом равенства  $X_i^{\text{целевое}}$  и  $X_i^{\text{моделируемое}}$  дает

$$Dev(x) = \sum_{i=0}^n \sqrt{(Y_i^{\text{целевое}} - Y_i^{\text{моделируемое}})^2} \quad (3.2)$$

и после упрощения

$$Dev(x) = \sum_{i=0}^n (Y_i^{\text{целевое}} - Y_i^{\text{моделируемое}}) \quad (3.3)$$

Для анализа правильности калибровки параметров, модели настраивалась на городе Кенема, затем сравнивались с результатами по Кайлхуну. Анализ проводился по графикам и аналитически по значению функции .



Рис. 1. Вставка рисунка в формате PDF.

Рассмотрим уравнение

$$\int_{-\infty}^{\infty} \dots \quad (3.4)$$

Согласно (3.4) введём

### Определение 1.

Учитывая определение 1, получим следующую

### Теорема 1. ...

*Доказательство.* ...

□

**Следствие 1.** ...

*Доказательство.* ...

□

## Заключение

Основным результатом данной работы является теорема 1. Из следствия 1 ...

### Список цитируемых источников

1. *Бейли Н.* Математика в биологии и медицине // М.: Мир. – 1970. – С.202-205.
2. *Кондратьев М. А.* Применение агентного подхода к имитационному моделированию процесса распространения заболевания // Научно-технические ведомости СПбГПУ. Серия «Наука и образование». — 2010. – том 2, вып. 2. – С. 189–195.
3. *Кондратьев М. А.* Методы прогнозирования и модели распространения заболеваний // Компьютерные исследования и моделирование. – 2013. – том 5, вып. 5 – С. 863–882

*Получена 01.03.2018*