

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ
ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение
высшего образования
**«Крымский федеральный университет имени В. И.
Вернадского»**
(ФГАОУ ВО «КФУ им. В. И. Вернадского»)
Таврическая академия (структурное подразделение)
Факультет математики и информатики
Кафедра прикладной математики

Консманов Алексей Витальевич

**Моделирование распространения
инфекционных заболеваний**

Выпускная квалификационная работа

Обучающегося	<u>4</u> курса
Направления подготовки	<u>01.03.04. Прикладная математика</u>
Форма обучения	<u>очная</u>

Научный руководитель
доцент кафедры прикладной
математики, к. ф.-м. н.

Ю.Ю. Дюличева

К ЗАЩИТЕ ДОПУСКАЮ:

Заведующий кафедры прикладной
математики, профессор, д. ф.-м. н.

В.Н. Чехов

Симферополь 2019

Оглавление

Введение	3
1 Подходы к моделированию процесса распространения инфекционных заболеваний	5
1.1 Агентный подход к имитационному моделированию процесса распространения заболевания	7
1.2 Применение случайных графов для распознавания и анализа инфекционных заболеваний	9
1.3 Модели инфекционной динамики на основе предфрактальных графов	13
1.4 SIR - SEIFDR модель	16
1.5 Выводы о существующих моделях	18
2 Реализация компьютерных моделей распространения инфекционных заболеваний	19
2.1 Выбор заболевания, его краткое описание	19
2.2 Сбор и анализ статистики по местности	20
2.3 Общие допущения для моделей	20
2.4 Выбор средств моделирования	21
2.5 Построение SEIFDR модели и ее оптимизация	22
2.6 Построение мультиагентной модели и её оптимизация	24
2.7 Тестирование и анализ разработанных моделей	26
Заключение	28
Список литературы	30

Введение

Несмотря на значительные достижения в области борьбы с инфекционными заболеваниями и порождаемыми ими эпидемиями, данная проблема и связанные с ней задачи все ещё актуальны. Одной из задач, возникающих в связи с борьбой с инфекционными заболеваниями, является исследование путей и способов передачи инфекций. В данной области возникает задача прогнозирования распространения конкретного инфекционного заболевания в некоторой заданной среде, то есть задача построения математической модели, описывающей скорость и масштабы распространения заболевания и его последствия: количество инфицированных и умерших, пространственные масштабы заражения, затраты средств и ресурсов, необходимые для изоляции больных и последующего излечения.

Математическое моделирование является мощным и гибким инструментом для исследования реальных процессов и объектов, а также связей между ними. Математическое моделирование удобно применять в тех ситуациях, когда проведение эксперимента с реальными субъектами затруднено по любой причине, что делает его подходящим способом исследования распространения инфекционных заболеваний. Заметим, что учитывая специфическую природу заболеваний и путей их передачи, которые будут изучены и исследованы далее, важным фактором для проверки адекватности построенной модели является наличие статистической информации, позволяющей тестировать соответствие модели реальным прецедентам.

Важно заметить, что модели, разработанные в данной области, не являются идеальными и универсальными. Основные допущения, принимаемые при построении модели, и недостатки этих моделей будут подробно рассмотрены далее.

Целью данной работы является анализ существующих моделей распространения инфекционных заболеваний и создание нескольких различных собственных моделей, позволяющих прогнозировать распространение конкретных заболеваний на основе существующих статистических данных, реализовать полноценное приложение для использования специалистами-эпидемиологами. Для этого необходимо решить нижеизложенный ком-

плекс задач:

- Проанализировать существующие подходы к моделированию, изучить сильные и слабые стороны этих подходов, принимаемые допущения. На основании этого анализа выбрать подход, который будет использоваться при построении собственной модели.
- выбрать заболевание для тестирования, подробно изучить его и собрать статистические данные о нем, найти статистику, позволяющую сравнить результаты моделирования с реальными;
- выбрать среду моделирования;
- разработать и реализовать базовую модель;
- на основе базовой модели создать модель повышенной точности и детализации;
- итеративно тестировать и улучшать результаты модели повышенной точности;
- разработать отдельное приложение для ввода данных в модель и визуализации полученных моделированием результатов для конечного пользователя – специалиста-медика или эпидемиолога.

Для решения поставленного комплекса задач использовались методы математической статистики и теории вероятности, дискретной математики, математического анализа. Разработанные модели основываются на методах агентного / мультиагентного моделирования, их реализации опираются на объектно-ориентированное программирование и современные средства программного анализа и визуализации данных.

Практическая ценность результатов работы: проведено исследование существующих подходов к моделированию, разработана модель и программный продукт, реализующий эту модель, также разработан программный продукт для ввода данных в модель специалистом и динамической визуализации результатов моделирования. Разработанные приложения позволяют дать количественный и качественный прогноз распространения инфекционного заболевания для любых достаточно точно и полно описанных популяции и заболевания.

1 Подходы к моделированию процесса распространения инфекционных заболеваний

Эпидемии всегда являлись одной из наиболее острых проблем медицины. Внезапно возникающие эпидемии формируют сложную ситуацию, динамически изменяющуюся во времени. В таких условиях медики и связанные чрезвычайные службы зачастую не могут принять адекватного решения и только порождают дополнительный хаос, неэффективно расходуют ресурсы, не способны в целом положительно повлиять на ход эпидемии или реализовать меры по борьбе с заболеванием.

Логично предположить, что одним из способов снижения угрозы для населения и уменьшения хаоса среди медиков и прочих ответственных служб являются меры предупредительного характера. Предупредительные меры могут иметь двойственный характер:

- 1 На основании регулярно обновляемых статистических данных, описывающих эпидемиологическую обстановку, возможно создавать кратковременную прогнозирующую модель, что позволит обнаружить эпидемию ещё до её начала;
- 2 Учет данных о прошлых подобных вспышках заболеваемости позволит составить прогноз для ответственных служб, описывающий характер необходимых ресурсов, их количество и область применения, что в свою очередь позволит уменьшить порождаемый вспышкой хаос.

Подобная идея не является революционно новой и к моделированию распространения заболеваний уже существует достаточно много подходов, основанных на разных принципах: цепи Маркова, агентный подход, стохастические эксперименты, дифференциальные уравнения на графах и другие различные алгоритмы на графах, полигамные модели. Далее будут подробно рассмотрены некоторые из этих подходов.

Некоторые общие допущения для всех моделей:

- Равномерность и стационарность в смысле неизменяемости во времени распределения возрастов, то есть все в популяции живут до

некоторого возраста L и затем умирают и для каждого возраста (включительно до L) количество людей в этой возрастной группе примерно или строго равно. Этот подход хорошо подходит для экономически развитых стран и стран, находящихся на постиндустриальном этапе экономического развития, где детская смертность мала и большинство людей доживают до ожидаемого возраста. Данное предположение может быть отвергнуто для стран, не подходящих под условия выше, например, при моделировании распространения лихорадки Эбола в странах Центральной и Южной Африки;

- Гомогенность или однородность перемещений внутри популяции, то есть индивидуумы в популяции *under scrutiny assort* и контактируют случайно, не замыкаясь в более мелких подгруппах. Данное допущение является спорным, т.к. социальная структура широко-масштабная и сложная, то есть индивидуумы внутри одной большой группы могут находиться в таких социальных отношениях, что большинство контактов будет приходиться на их собственную подгруппу, в то время как количество контактов вне группы будет предельно малым. Однако, такое допущение имеет место быть для упрощения построения модели и понимания результатов моделирования.

Также опишем один общий принцип разделения моделей по признаку случайности. По такому признаку модели можно разделить на стохастические и детерминистические. Стохастические модели предполагают наличие случайных величин и являются инструментом для оценки распределения вероятности потенциальных исходов посредством допущения случайных изменений в одной или нескольких переменных с ходом времени. Такие модели зависят от случайных изменений рисков, связанных с продолжительностью экспозиции, вероятности заболевания и прочих динамических изменяющихся параметров заболевания. Для описания различных этапов заболевания часто используют буквы M(maternally-derived immunity), S(suspicious), E(exposed), I(infected), R(recovered).

1.1 Агентный подход к имитационному моделированию процесса распространения заболевания

Существует несколько понятий определения "агент" и "агентный подход" в целом. Обобщая их, можно выделить, что *агент* – это некоторая сущность, имеющая активность, автономное поведение, способная самостоятельно принимать решения в соответствии с некоторой совокупностью правил, взаимодействовать с окружающей средой, если такая предусмотрена, и другими агентами, если такие существуют. Основное поле применения таких моделей – децентрализованные системы, динамика функционирования которых не является следствием некоторых, зачастую внешних, правил и законов, но наоборот, такие правила и законы являются внутренним неотъемлемым результатом работы множества агентов. Такие модели обычно являются дискретно-событийными или дискретными с непрерывными элементами, то есть гибридными.

Данный подход удобно иллюстрировать на примере моделирования распространения гриппа (обычно, гриппа А), т.к. грипп вносит весомый вклад в смертность при инфекционных заболеваниях и при этом хорошо известны его способы передачи и формы течения болезни. Также по данному заболеванию собрана достаточно обширная статистика, который можно использовать для проверки точности модели и последующей калибровки. Такая модель достаточно подробно рассмотрена в описываемой модели, способную определять число больных гриппом в каждый день определенного краткосрочного периода (2-3 недели) во время сезонных вспышек, то есть построение краткосрочного прогноза. Для этого популяция разбивается на возрастные группы, позволяющие характеризовать контактов с другими людьми в сутки и, что важно, потенциально посещаемые места. В работе вводятся два основных типа объектов: локации и агенты и связанные с ними события перемещения агентов между локациями, протекания заболевания у агентов и событие «контакта». Также в данной работе вводятся правила, описывающие поведение агентов с помощью UML Statecharts (далее «стейтчарты»). С помощью таких стейтчартов описывается переход между различными состояниями индивидуума как вне течения болезни, так и после, и может быть обозначен с помощью графической нотации как:



Данная модель обладает рядом преимуществ: высокая скорость разработки в силу выбора готового имитационного ПО (специфично для данной рассматриваемой работы, хотя в целом верно для разработки любой агентной модели в силу простоты инкапсуляции агентов и построения связей между объектами); высокая скорость работы, так как выбранная модель допускает высокий параллелизм в рамках одной итерации, когда вычисления для каждого объекта-места допускают распараллеливание; свобода от дифференциальных уравнений значительно упрощает расчеты, в том числе исчезает необходимость использовать численные методы и связанные с ними понятия «сходимости» и «устойчивости», при этом последнее особенно важно, так как входные статистические данные зачастую имеют шумы; сам агентный подход допускает быструю модификацию модели и возможность учета применяемых административных мер и их влияния на ход распространения инфекции.

Недостатки: в силу собственной стохастичности, модель требует многократного запуска и оценкой экспертом полученных результатов для получения некоторого «среднего результата»; достаточно сложная модель может потребовать много времени на разработку даже при использовании готового имитационного ПО, при этом такая модель может оказаться ресурсозатратной в терминах машинных мощностей (время, память); невозможно точно определить, насколько точна данная модель по отно-

шению к реальным, кроме как посредством многократного сравнения результатов работы модели и собранных существующих данных, при этом невозможно убедиться, что данная модель и текущие параметры могут быть применимы для прогнозирования, а не только откалиброваны для соответствия предыдущим результатам; стохастическая природа модели при компьютерной реализации опирается на генерацию случайных чисел, что требует использования мощных и проверенных генераторов.

1.2 Применение случайных графов для распознавания и анализа инфекционных заболеваний

В качестве другого подхода к моделированию распространения инфекционных заболеваний можно рассмотреть переносчиков заболеваний как саморазмножающиеся сущности, т.е. такие сущности, которые способны самореплицироваться, при этом свойство саморепликации передается не только между «родителями» и «потомками» (вертикальный перенос), но возможен и горизонтальный перенос, т.е. передача свойства саморепликации и, возможно, но не обязательно, других свойств, объектам, которые так или иначе соседние или контактируют с данным. Примерами являются размножение компьютерных вирусов и сетевых червей или инфицирование организмов вирусами и инфекциями. В данных моделях обычно применяются графы, используемые для моделирования эпидемий «мобильных червей» (т.е. вредоносных программ, распространяющихся между устройствами при помощи беспроводных протоколов типа Bluetooth или Wi-Fi, широко применяемых в мобильных устройствах, отсюда и происходит название) или воздушно-капельных инфекций среди высших животных. Согласно К.Е. Климентьеву, для подобных эпидемий характерны следующие черты:

- Постоянные изменения топологии среды моделирования в связи с высокой мобильностью объектов;
- ограниченный радиус «заражения», как для биологических агентов, так и для компьютерных вирусов, обусловленный физическими свойствами оных.

Для моделирования среды существования таких объектов используется класс сетей, называемый «специальным» («Ad hoc»), представляющий собой множество случайных графов с разнообразной топологией. При этом нетрудно выделить общие для таких графов характеристики:

- Вероятностное распределение степеней вершин, обозначаемое k_i ;
- также вероятностное распределение для локальных коэффициентов кластеризации вершин, обозначаемое c_i .

При этом под «степенью вершины» x подразумевается количество вершин графа G , инцидентных вершине x , т.е. степень вершины указывает количество вершин, непосредственно соединенных только одним ребром с данным. «Локальный коэффициент кластеризации вершины» понимается в том же смысле, что и у К.Е. Климентьева и Б. Хогана, т.е. «коэффициент кластеризации» рассматривается как вероятность того, что два ближайших соседа этого узла сами являются ближайшими соседями, а ЛККВ является мерой того, насколько хорошо связаны между собой соседи данного узла и рассчитывается как отношение числа связей между соседями данного узла к возможному числу связей между соседями (по Хогану) или, менее формально, по К.Е. Климентьеву: «доля «треугольников» данной вершины, образованных из «соседей» данной вершины и являющихся «соседями» друг для друга в общем количестве потенциально возможных треугольников, где $c_{max} = k_i \frac{k_i - 1}{2}$ ».

В качестве модели пространства, используемых для имитации развития и распространения эпидемии, применяются неориентированные, маркированные графы (в смысле частного случая сети Петри, в которой каждая позиция входом и выходом точно для одного перехода), являющиеся частным случаем дистанционных графов, при этом дистанционный граф понимается как $G = (V, E)$ – n -мерный дистанционный граф (граф расстояний), если $V \subseteq \mathbb{R}^n, E \subseteq \{\{\bar{x}, \bar{y}\} : \bar{x}, \bar{y} \in V, |\bar{x} - \bar{y}| = a, a \in \mathbb{R} > 0\}$, то есть множество вершин M является подмножеством или совпадает с n -мерным пространством, а множество ребер E является подмножеством или совпадает с множеством всевозможных пар вершин \bar{x}, \bar{y} , таких, что евклидово расстояние между \bar{x}, \bar{y} равно некоторому фиксированному вещественному положительному заранее заданному a .

Топология таких графов связана с принципом их построения, т.е. каждая вершина графа имеет некоторые пространственные координаты (обычно в \mathbb{R}^2 или \mathbb{R}^3) и «соседними вершинами» считаются только те, расстояние до которых меньше заданного или равно заданного r_0 (расстояние передачи вирусного воздействия).

Допущения:

- Инфекция передается только между инцидентными вершинами за некоторое время;
- вершины графа активно перемещаются, постоянно изменяя его конфигурацию.

Входные параметры модели:

- Начальная конфигурация графа, т.е. количество и координаты вершин вместе с их начальным состоянием, обычно в виде перечисления «здоров», «вакцинирован», «болен», однако допускается и большее число состояний, подобное применяемому в агентных подходах;
- максимальное расстояние инфицирующего воздействия r_0 ;
- правила или порядок перемещения вершин в процессе моделирования.

Способы построения исходного графа, исходя из «геометрических» и «географических» по К.Е Климентьеву соображений подробно рассмотрено в [1]. При этом для построения геометрического графа по заранее заданным параметрам типа k_i и c_i используется случайный граф Радо (RRG) или случайный геометрический граф (RGG) и в этой же работе выводится функция распределения между двумя случайными точками в единичном квадрате и математическое ожидание расстояния.

Намного более интересным свойством данных моделей являются правила перемещения вершин графа: если в описанных ранее подходах перемещения вершин присутствуют скорее в силу необходимости для динамичности во времени и модель учитывает только их положение по отношению к другим вершинам и, иногда, скорость движения этих, а топология графа не является предметом исследования, то в данном подходе

топологию графа и правила перемещения являются ключевым элементом при построении модели. В результате построения множества моделей, использующих данный подход, можно сделать следующие выводы, описывающие правила перемещения вершин в графе:

- Классические модели блуждания, описывающие, например, броуновское движение, мало пригодны для описания передвижения людей, особенно движущихся вне рутинного маршрута, а высокоподвижных агентов, в силу своей слабой стационарности. Следовательно, при анализе движений по плоскости, направление движения распределено равномерно на интервале $[0; 2\pi]$, что показано в ;
- скорость и продолжительность движения распределены по Леви. При этом «перемещение» или «движение» понимается как самый длинный прямолинейный переход объекта с одного места в другое без паузы или изменения направления ;
- тем не менее, перечисленные правила перемещения актуальны только для перемещений на ограниченном участке плоской (или приближенной к плоской) поверхности без учета препятствий, конфигураций помещения, уставившихся маршрутов, столкновений с другими телами и необходимости их избегать. Для внесения корректировок, учитывающих такие особенности, необходим сбор статистики на множестве реальных наблюдений с последующим внесением в некоторую ГИС.

Базируясь на работах , , , можно сделать следующие выводы, касающиеся основных свойств данного подхода к моделированию:

- 1 Данный подход хорошо подходит для моделирования распространения инфекционных заболеваний среди групп людей и животных, особенно блуждающих, в т.ч. по Леви, в некотором постоянном данном ареале;
- 2 использование данных моделей улучшает понимание того, какие графы лучше описывают топологию быстро движущихся источников инфекции. Так, в работах и четко прослеживается зависимость

между «Ad hoc» графами и графами Радо и той картиной, которую дают реальные перемещения мобильных устройств при вспышке компьютерных инфекций.

Достоинства данного подхода: высокая скорость работы моделей при правильном выборе инструментария (готовые фреймворки для работы с графами); высокая точность предсказания протекания вспышки инфекции на начальном этапе этой вспышки; наличие уже разработанных алгоритмов, приборов и готовых собранных данных для занесения в многослойные ГИС; возможность модификации алгоритмов для учета рельефа местности и возможных столкновений инфицируемых существ; математический аппарат блужданий Леви допускает модификации, позволяющие легко настраивать и изменять принципы перемещения существ; сопоставление результатов моделирования с реальными процессами позволяет изучать зависимости между некоторыми специальными графами, описанными выше, и путями распространения инфекций.

Недостатки: в силу специфики требований, предъявляемых к моделям данного типа, точное моделирование возможно только в течении небольшого количества итераций модели, т.к. предполагается характер «вспышки», то есть быстрого бурного увеличения числа зараженных индивидуумов с достаточно быстрым падением числа больных при изобретении лекарства или разработке антивируса; для построения точных моделей требуется не только точно знать координаты физического пространства, на котором будет происходить распространения заболевания, но и иметь детальные модели самих поверхностей или рельефов, что необходимо для точного построения блужданий Леви; используемые графы высокодинамичные и могут требовать значительных затрат как по памяти, так и по процессорному времени.

1.3 Модели инфекционной динамики на основе предфрактальных графов

Для моделирования распространения инфекционных заболеваний можно также использовать предфрактальные графы. В основу данного подхода к моделированию положена идея о том, что графы, использующие опера-

цию "замены вершины затравкой" (ЗВЗ), которая будет рассмотрена далее, «прирастают» не отдельными вершинами, а самоподобным или приближенно самоподобным графом, т.е. фрактальным, хорошо описывают системно «прирастающие» структуры: телекоммуникационные, инфраструктурные, социальные и технические сети, некоторые популяции, в которых существует разбиение на группы, в которых связей внутри группы больше, чем между отдельными группами. При этом показано, что такие модели могут быть использованы для моделирования распространения инфекционных заболеваний, а именно, что процессы разрастания графов релевантны динамике распространения инфекционных заболеваний и распад таких графов соответственно релевантен спаду эпидемий. Математико-алгоритмическая идея построения фрактальных графов – операция замены вершины затравкой (ЗВЗ). Для этого вводится несколько понятий, рассмотренных ниже. По определению, динамический граф G_D – последовательность обычных графов G_l , не имеющих параллельных ребер и петель. Затравка – какой-либо произвольный связный граф, т.е. граф, в котором между любой парой вершин существует хотя бы один путь, сама затравка обозначается как $H = (W, Q)$. Предфрактальный граф обозначается как $G_L = (V_L, E_L)$, определяется рекуррентно, заменяя каждый раз в построенном на предыдущем этапе $l = \overline{1, L-1}$ графе $G_l = (V_l, E_l)$ каждую вершину затравкой H . Вершины могут соединяться случайно или по заданному правилу или в некоторой заранее заданной последовательности. На этапе $l = 1$ предфрактальному графу соответствует затравка $G_1 = H$. Говорят, что предфрактальный граф G_L *порожден* затравкой H , при этом процесс построения предфрактального графа G_L по сути является процессом построения последовательности других предфрактальных графов $\overline{G_1, G_L}$, называемый «траекторией».

Суть операции замены вершины затравкой (ЗВЗ): в графе $G = (V, E)$ из выбранной для замещения затравкой вершины $\tilde{v} \in V$ выделяется множество $\tilde{V} = \{\tilde{v}_j\} \subseteq V$, где $j = \overline{1, |\tilde{V}|}$, смежных ей, вершине $\tilde{v} \in V$, вершин; затем из графа G удаляется вершина \tilde{v} и все инцидентные ребра, т.е. ребра, для которых эта вершина общая; после $\forall \tilde{v}_j \in \tilde{V}, j = \overline{1, |\tilde{V}|}$ соединяется ребром с одной из вершин затравки H , при этом соединение может происходить случайно или по заранее заданному правилу или в заданной

последовательности.

Под «распознаванием предфрактального графа» в таких подходах к моделированию обычно понимается определение траектории построения предфрактального графа при условии, что заданы виды и типы затраток. При этом задача распознавания различных типов предфрактальных графов хорошо изучена и отработана в практических алгоритмах, что дает широкую теоретико-практическую базу для решения данной задачи.

В предфрактальном графе G рёбра, порожденные на этапе l , где $l \in 1, 2, 3, \dots, L$ называются ребрами ранга l , при этом «новыми» ребрами называют ребра ранга L в предфрактальном графе G_L , все остальные называются «старыми». При этом, под рангом вершины графа понимается наименьший ранг l , выбираемый среди всех рёбер, инцидентных для данной вершины. Вершина ранга l обозначается как $v^{(l)}$, $l \in \{1, 2, \dots, L\}$. Также вводится понятие очага заражения ранга l , который обозначается как $\tilde{v}^{(l)}$ и является, по сути, вершиной ранга l .

Одной из самых важных задач данного подхода к моделированию является задача распознавания фрактального графа, при этом различают два вида распознавания: неявное – определение фрактальности графа и установление некой n -вершинной затравки, на которой он базируется; явное – представление множества рёбер для всякого ранга в явном виде или же представление порождающей траектории в явном виде для некоторого заданного графа G . Для этого используется несколько алгоритмов, показанных в и и обозначаемых $\alpha_1 \dots \alpha_4$.

Как показано в и , данные модели могут быть использованы для описания распространения и протекания как инфекционных заболеваний, так и для эпидемий «компьютерных вирусов», особенно в локальных сетях, т.к. их топография зачастую подобна предфрактальным или фрактальным графам; модели, которые можно построить с помощью данного подхода, могут быть использованы для выявления кластеров заражения или очагов заражения (для эпидемий, не предполагающих иммунитета). Также, значительным преимуществом данных моделей является то, что для распознавания фрактального графа имеется набор готовых алгоритмов $\alpha_1 \dots \alpha_4$, являющихся полиномиальными; модели позволяют учитывать

уровень иммунитета каждого человека или каждого устройства в сети; уже существуют готовые реализации данных моделей в виде программных реализаций. Также, для прикладной реализации важной является вычислительная сложность, для $\alpha_2 = O(|E|L)$ (если «старые» ребра не пересекаются) и $\alpha_{1,3,4} \leq O(|E|L)$, где $|E|$ понимается как количество ребер.

Недостатком данного подхода можно назвать то, что в силу конечного числа шагов трудно добиться полной фрактальности графа и зачастую используются приближенно фрактальные модели.

1.4 SIR - SEIFDR модель

SIR модель впервые была предложена Кермаком и МакКендриком в 1927 г. для описания заразных заболеваний в закрытых популяциях с течением времени. Модель предполагает фиксированный размер популяции (без смертей от болезни и естественных причин, без рождаемости), мгновенность инкубационного периода возбудителя болезни, равную продолжительность заразности и самой болезни, полную однородность популяции в терминах возраста, пола, пространства или социальной структуры, т.е. данные параметры не учитываются. Модель состоит из системы нелинейных дифференциальных уравнений:

$$\begin{cases} \frac{dS}{dt} = -\beta SI \\ \frac{dI}{dt} = \beta SI - \gamma I \\ \frac{dR}{dt} = \gamma I \end{cases} \quad (1.4.1)$$

, где t – время, $S(t)$ – количество подозреваемых (Suspected), $I(t)$ – инфицированных (Infected), $R(t)$ – выздоровевшие с иммунитетом (Recovered), β – интенсивность инфицирования, γ – интенсивность выздоровления. Важным параметром, управляющим развитием модели, является эпидемиологический порог $R_0 = \frac{\beta S}{\gamma}$. Если $R_0 < 1$, то один больной человек заразит менее одного человека перед своим излечением, т.е. вспышка исчезнет; если $R_0 > 1$, то инфекция будет распространяться, т.е. $\frac{dI}{dt} > 0$. Данная модель и ее частная версия Кермака-МакКендрика настолько

интересны, насколько и просты, что делает их подходящими только для описания коротких вспышек не летальных заболеваний, например, холеры в Лондоне в 1865 г. Для описания других болезней необходимо исследовать их особенности и, в первую очередь, состояния, которые свойственны больным. Так возможно модифицировать базовую SIR модель до модели, учитывающей смертности и заразность трупов до их захоронения. Такая модель называется SEIFDR модель и подробно исследована в [1] и является глубокой модификацией SIR, где вводятся три новых состояний и ряд переменных, от которых динамика этих состояний зависит. Так, SEIFDR модель описывается системой:

$$\begin{cases} \frac{dS}{dt} = -\beta_I SI + \beta_F SF \\ \frac{dE}{dt} = \beta_I SI + \beta_F SF - \alpha E \\ \frac{dI}{dt} = \alpha E - I\gamma_D((1 - \delta_D) + \gamma_D \delta_D) \\ \frac{dF}{dt} = \gamma_D \delta_D I - \gamma_F F \\ \frac{dR}{dt} = \gamma_I(1 - \delta_D)I + \gamma_F F \end{cases} \quad (1.4.2)$$

, где β_I и β_F – интенсивность контактов между людьми и на похоронах, α – интенсивность инкубационного периода, γ_I интенсивность выздоровления, γ_D интенсивность умирания агентов, γ_F – интенсивность похорон, δ_D – вероятность смертности. Преимущества: чистый математический подход, выраженный СДУ; наличие большого количества готового имитационного ПО, способного быстро и достаточно точно решать (1.4.2); высокая прогностическая точность при правильном подборе параметров для заданной болезни; возможность динамически расширять возможные состояния в будущем.

Недостатки: система требует длительной калибровки параметров для максимально точного описания болезни; требуется минимально зашумленная статистика хотя бы инфицированным и умершим для сравнения с работой модели; после калибровки модели для данной болезни на одном городе необходимо тестировать откалиброванную модель на другом городе для избежания создания модели, которая будет только «подогнанной» под конкретный случай, но не имеющей прогностических

свойств; каждая новая болезнь требует новой калибровки с нуля; модель не учитывает социальную структуру популяции.

1.5 Выводы о существующих моделях

Как было показано выше, существует широкое множество различных подходов к имитированию распространения инфекционных заболеваний. При этом каждый из таких подходов обладает своим собственным рядом преимуществ и недостатков. Для реализации и сравнения выбраны две модели: мультиагентная модель, за простоту и скорость, и SEIDFR модель за её строгость и гибкость.

2 Реализация компьютерных моделей распространения инфекционных заболеваний

Как уже было указано в предыдущем разделе, для реализации выбраны мультиагентная и SEIFDR модели. Как указано в целях работы, необходимо выбрать целевое заболевание и среду моделирования.

2.1 Выбор заболевания, его краткое описание

Прежде всего, исследуемое заболевание должно обладать ярко выраженными симптомами, что позволяет получить высокорепрезентативную выборку.

В качестве целевого заболевания для анализа выбрана геморрагическая лихорадка Эбола – острое вирусное высококонтагиозное заболевание, поражающее человека и других приматов, . Данное заболевание характеризуется высокой заразностью (человеческие останки остаются заразными до 50 дней), коротким инкубационным периодом, резкими клиническими проявлениями и очень высокой смертностью (до 90%, в среднем около 50%). Инфицирование происходит при прямом контакте через биологические жидкости, выделения, предметы. Основные симптомы (по времени проявления): усталость, лихорадка, слабость, миалгия, артралгия, боли в горле, чиханье и диарея, обширные внутренние и внешние кровотечения, в том числе коагулопатия, кровохарканье, кровь в глазном белке, гиповолемический шок, приводящие к смерти.

Данное заболевание удобно для изучения тем, что распространяется в странах с низким уровнем гигиены и медицинской поддержки, что позволяет изучить «чистый» вирус на «чистой» популяции, т.е модель, описывающая вирус, может пренебрегать медицинскими мероприятиями; также имеется подробная посуточная статистика с низким уровнем шума по государствам и отдельным городам.

Типичные группы риска: медработники, участники похоронных команд или другие лица с непосредственным доступом к трупам, члены семьи и другие лица, находящиеся в тесном контакте с инфицированным, ведущие промысел во влажных экваториальных лесах охотники,

при любом контакте с трупами инфицированных животных или при использовании зараженных животных в пищевой промышленности. Также в этой группе находятся все жители Центральной и Южной (кроме ЮАР) Африки и Западного и Восточного побережий – в этих регионах наблюдается низкий уровень санитарии, небезопасные промыслы на диких животных, низкий уровень медицины и образования людей в целом и неблагоприятная социально-политическая ситуация, что особенно хорошо видно на примере вспышки лихорадки Эбола в ДРК весной 2019 г.

2.2 Сбор и анализ статистики по местности

Прежде всего, следует отметить, что низкий уровень инфраструктуры и подготовки медицинского персонала на местах вместе с большими расстояниями значительно затрудняет и «зашумляет» сбор статистики, поэтому среди источников для анализа необходимо выбрать наиболее авторитетный и полный, что снимет необходимость интерполяции данных и фильтрации выборки от шумов.

Усилия WHO направлены на детальный ежесуточный сбор статистики по государством и отдельным городам. Основываясь на анализе репозитория WHO, можно утверждать, что самая детальная статистика дана по государству Сьерра-Леоне и наиболее пострадавшие города, также имеющие самую репрезентативную статистику, это Кайлахун и Кенема, что соответствует выводам.

2.3 Общие допущения для моделей

Как показано выше, лихорадка Эбола является высококонтагиозным заболеванием с высокой летальностью. Рассмотрим основные этапы данного заболевания в терминах симптомов и продолжительности, что важно при реализации моделей:

- 1 Заражение. Происходит при контактах с телесными жидкостями зараженного человека или примата.
- 2 Инкубационный период. Занимает от 4 до 10 дней.

- 3 Явное проявление болезни. Сопровождается усталостью, жаром и болями в теле. На данном этапе наиболее важными и опасными симптомами являются гематемезис, гемоптизис, кровотечения в ЖКТ и мелена, петехии. Данные симптомы представляют опасность здоровую инфицированного ввиду гиповолемического шока, но и являются основной причинной передачи заболевания между людьми.
- 4 Выздоровление. Обычно наступает через 1-2 недели после проявления симптомов. Считается, что после этого вырабатывается иммунитет сроком до 10 лет, однако неизвестно взаимодействие данного специфического иммунитета с другими штаммами лихорадки Эбола. На данный момент ведутся исследования в области сохранения вирулентных клеток у пострадавших даже после выздоровления. В силу спорности данного утверждения и отсутствия значимой статистики, описываемые далее модели не будут учитывать этот источник заражения.
- 5 Смерть. Обычно наступает через 6-16 дней после проявления первых симптомов из-за гиповолемического шока. Важно заметить, что умершие тела все еще являются источником заражения, поэтому вероятность контакта между умершими и здоровыми людьми учитывается в описываемых далее моделях.

Важно заметить, что в нулевой момент времени вся популяция не может находиться в состоянии «Suspected», иначе распространение болезни происходить не будет, т.е. в модели вводится параметр, отвечающий за начальное количество людей, находящимся в инкубационном периоде, т.е. в состоянии «Exposed» в терминах SEIFDR модели.

2.4 Выбор средств моделирования

Для построения SEIFDR модели требуется имитационное ПО, способное решать системы дифференциальных уравнений и гибко визуализировать результаты моделирования, а также сопоставлять их с реальными результатами. При этом ПО не должно быть привязано к конкретной популяции, но должна быть возможность точно и гибко калибровать мо-

дель. Для анализа были взяты NetLogo, SOARS и AnyLogic. NetLogo был отброшен, т.к является ПО для обучения и требует модификации для сложных вычислений и экспериментов на внешних языках программирования. SOARS также был отброшен, т.к не имеет возможности решать системы дифференциальных уравнений. Оптимальным выбором был признан AnyLogic, т.к данное ПО является популярным промышленным решением, обладает методами для численного решения систем дифференциальных уравнений, инструментами для проведения оптимизационных экспериментов и инструментами для построения графических интерфейсов пользователя.

Для построения мультиагентной модели был выбран язык программирования Python в связке с библиотеками NetworkX , используемой для построения графа локаций и WolframClient 1.0.2 , используемой для прямого взаимодействия с системой компьютерной алгебры Wolfram Alpha, что позволяет отказаться от изучения и использования NumPy и SciPy .

2.5 Построение SEIFDR модели и ее оптимизация

Данная модель представляет собой систему дифференциальных нелинейных уравнений и имеет ряд переменных, описывающих вероятность контакта и смерти. Очевидно, что каждое заболевание имеет свой характер и, следовательно, свое значение этих параметров. Таким образом, основной задачей при построении данной модели является проведение ряда экспериментов с целью подбора этих параметров с последующим тестированием на другой популяции, что позволит подтвердить правильность подбора данных параметров. Результатом работы модели будет количество больных, умерших и выздоровевших в каждый дискретный момент времени.

При построении данной модели возможно использовать средства AnyLogic из палитры «Системная динамика»: количество людей в каждом стоянии описывается одноименным накопителем, а переменные, отвечающие за интенсивность контактов, вероятность смерти и т.д., представляют собой статические параметры. В силу простоты модели и незначительного количества связей между параметрами и накопителями, вместо использования потоков возможно применить связи – это не оказывает

влияния на решение системы дифференциальных уравнений.

После построения базовой модели с неоткалиброванными параметрами необходимо ввести целевую функцию, минимизация которой и является ключевым этапом калибровки модели. Согласно [1], в качестве целевой функции может быть выбрано среднеквадратичное отклонение результатов работы модели от реально наблюдаемых результатов для больных и умерших по отдельности. Введение целевой функции позволяет численно, т.е. аналитически, сравнивать результаты запусков модели с различными параметрами. После введения целевой функции появляется возможность провести стохастический и оптимизационный эксперименты.

Стохастический эксперимент является базовым на начальном этапе настройки модели, т.к. многократным повторением эксперимента достигается возможность получения достоверных выводов о работе модели в целом. Очевидно, что при наличии стохастического параметра модели, недостаточно будет одного прогона. Количество итераций обосновано в [1] и было использовано при проведении данного эксперимента.

На гистограммах выше каждая линия указывает на результаты одного из 2000 стохастически моделируемых прогноза эпидемии. Более плотный цвет указывает на большую вероятность попадания данных эксперимента в указанное значение в указанный момент времени. Следует отметить, что наличие только одного стохастического параметра («Exposed», т.е. начальное количество людей в инкубационном периоде, нормально распределённое в интервале $[1; 10]$) приводит к незначительному разбросу между прогнозами.

Идея оптимизационного эксперимента заключается в многократном выполнении модели с итеративным изменением параметров, подлежащих уточнению, для улучшения модели. Для эксперимента использовались средства AnyLogic и системы компьютерной алгебры Wolfram Mathematica. По окончании эксперимента возможно наблюдать лучшую итерацию, её оптимизированные параметры, значение целевой функции для каждого интересующего нас состояния.

2.6 Построение мультиагентной модели и её оптимизация

Данная модель представляет собой динамично взаимодействующее множество агентов двух типов: агентов-людей и агентов-локаций.

Множество агентов-локаций описывает территорию, где происходит действие, при этом необязательно учитывать реальный геодезический или географический характер местности, а достаточно нормально распределить агентов-людей по локациям, что будет соответствовать реальному распределению людей в городах между общественными местами (школы, больницы, предприятия) и местами жительства. Агенты-люди описывают людей с их восприимчивостью, возрастом и текущей позицией и являются основным действующим лицом в модели, т.к именно они перемещаются по локациям и участвуют в событии контакта, которое может привести к заражению. Состояния агентов-людей полностью соответствуют таковым у SEIFDR модели, однако, механизм перехода по состояниям значительно отличается и не описывается дифференциальными уравнениями. Как было указано в разделе 2.3, вся популяция изначально полагается в стоянии «Suspected» и, подобно SEIFDR модели, имеется некоторое случайное число агентов с заболеванием в инкубационном периоде, распределенное аналогичным образом. Информация, инкапсулируемая классом «Агент-человек» используется в вычислениях: так, возраст влияет на вероятность попадания в новую локацию (старiki и дети менее подвижны) и восприимчивость к болезни (иммунитет детей и стариков развит меньше).

Принцип взаимодействия элементов модели и некоторые числовые параметры выбраны на основании уже проведенных исследований, связанных с лихорадкой Эбола.

Агенты-люди распределены по локациям и каждый день раз в 8 часов перемещаются между ними случайным образом, что примерно соответствует суточной активности человека. При этом между агентами, находящимися в одной локации, возможно событие контакта и передачи инфекции, что будет описано подробно далее.

Основной событийный цикл модели выглядит следующим образом:

- 1 В каждой локации для каждого агента в состояниях «Exposed», «Infected» или «Funeral» с некоторой вероятностью происходит контакт с другим здоровым агентом, т.е «Suspected» агентом. Вычисляется вероятность заражения жертвы. Если событие заражения произошло, жертва переходит в состояние «Exposed».
- 2 Раз в день проверяется событие перехода по зараженным состояниям. Переход между этими состояниями происходит с заданной для каждого состояния интенсивностью. Так, переход из «Exposed» в «Infected» происходит в среднем раз в 7 дней, «Infected» – «Funeral» – 11 дней, «Infected» – «Recovered» – 12 дней. Переход «Funeral» – «Dead» зависит от интенсивности похорон, т.е является специфичным для каждого государства и даже города в зависимости от мобилизованных ресурсов и общего уровня здравоохранения, но т.к данная модель обучалась на достаточно бедном и неразвитом государстве в период, предшествующий мобилизации ресурсов, предполагается, что похороны могут занимать до 3 дней.
- 3 Происходит перемещение агентов по локациям. В общем случае, нельзя утверждать, что локация, в которую перейдет агент не будет совпадать с той, в которой он находился до этого. Исполнение цикла возвращается в 1.

Основой для вычислений является эмпирическая формула вычисления вероятности заболевания:

$$P_i = 1 - e^{\tau N_r \ln(1 - r S_i \rho)}, \quad (2.6.1)$$

где P_i – вероятность инфицирования i -го агента, τ – коэффициент расстояния, важный при передаче заболевания на большие дистанции, например, при межконтинентальных перевозках, (в данной модели равен 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), ρ – допустимость, т.е вероятность заражения одного индивида другим за фиксированное вре-

мя (в данной модели – одни сутки, т.е время пребывания агента в одной локации).

Как и в SEIFDR модели, целевой функцией выбрано среднеквадратичное отклонение реальных результатов от результатов работы модели по больным и умершим. Аналогично, проводятся стохастический и оптимизационный эксперименты. Для данной модели, где присутствует больше стохастических переменных, хорошо наблюдаем намного больший разброс, чем в аналогичном эксперименте у SEIFDR модели.

Аналогично предыдущей модели проводится оптимизационный эксперимент. Для выполнения расчетов привлекались средства Wolfram Mathematica.

2.7 Тестирование и анализ разработанных моделей

Предполагается, что модели будут использоваться в реальных условиях на современных персональных компьютерах, поэтому обучение и тестирование модели проводились на аналогичной машине:

ОС: Windows 10

Разрядность: x64

CPU: Intel(R) Core(TM) i7-8550U CPU @ 1.80GHz (8 CPUs), 2.0GHz

Оперативная память: 16 Гб

Для анализа и сравнения результатов работы отдельных итераций каждой модели и обеих моделей между собой, необходимо было не только откалибровать модели по параметрам, но и проверить корректность моделей в целом. Для этого модели сначала обучались на городе Кенема, а затем тестировались на городе Кайлахун. Сравнение проводилось по графикам заболеваемости и смертности визуально и аналитически по значению целевой функции для больных и умерших.

Согласно , при обучении моделей необходимо учитывать принятие мер по борьбе с распространением заболевания, следовательно, модели будут обучаться на временном интервале до принятия мер.

Как видно на графиках, с определенного дня приращение числа больных резко снижается, что соответствует времени принятия мер по защите населения. Исходя из этого, модели будут обучаться на временном интервале, соответствующему $[0; 22]$ дня на графиках выше.

В таблицах ниже приведены параметры для обеих моделей при первом построении и после окончательной оптимизации.

На графиках ниже представлены результаты работы откалиброванных моделей для города Кайлхун.

Таблица ниже показывает аналитические характеристики работы моделей. Заметно, что SEIFDR модель показывает результат несколько превосходящий результаты мультиагентной модели.

В ходе калибровки обеих моделей было замечено, что при росте точности прогнозирования значительно возрастает время работы модели, особенно для мультиагентной модели. Таким образом, дальнейшая детализация этой модели приведет к значительному росту затрат по времени при одновременном малом приросте точности моделирования.

Заключение

Выбранные для реализации подходы представляют собой сочетание строгой формализации с одной стороны и гибкость, позволяющую обходить ограничения и трудности, встречающиеся в других моделях, с другой стороны.

Агентный подход имеет большой потенциал, т.к. позволяет динамично менять модель по мере детализации предметной области и получения большего количества данных, а также учитывает стохастическую природу реальных эпидемических процессов.

Обе построенные модели позволяют прогнозировать врачу-специалисту эпидемическую обстановку на основе имеющихся у него статистических данных, а также оценить количество ресурсов, необходимых для борьбы со вспышкой инфекции, если такая будет иметь место. Естественным ограничением является территориальный масштаб данных моделей, ограничиваемый одним городом или крупным районом мегаполиса.

Были решены следующие задачи:

- Проведен анализ существующих подходов к моделированию распространения инфекционных заболеваний. На основе анализа были выбраны две самые перспективные по мнению автора работы модели;
- было выбрано и проанализировано заболевание, которое будет моделироваться, была найдена подробная характеристика по нему;
- были выбраны средства моделирования;
- было реализовано две модели, затем их точность была увеличена, были проведены стохастические и оптимизационные эксперименты с целью оптимизации параметров моделей;
- получены приложения, позволяющие прогнозировать специалисту возможные вспышки заболевания;
- построенные модели после обучения были протестированы на другом городе, что подтверждает правильность выбранных подходов и параметров моделей.

Очевидно, что данные модели не могут описать всех факторов, формирующих эпидемиологическую динамику (социальные отношения, меры по защите здоровья, влияние географических и климатических факторов), но даже этого достаточно для выявления общих закономерностей при вспышке лихорадки Эбола и примерной оценки ущерба и ресурсов, необходимых для борьбы с ней. Отдельно заметим, что модели, особенно мультиагентная, допускают гибкое модифицирование в будущем.

Список литературы