

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г.ЧЕРНЫШЕВСКОГО»**

Кафедра математического и компьютерного моделирования

Эволюционное моделирование постоянно мутировавших

в течение пандемии вирусов

АВТОРЕФЕРАТ МАГИСТЕРСКОЙ РАБОТЫ

Студента(ки) 2 курса 247 группы
направления (специальности) 09.04.03 — Прикладная информатика

механико-математического факультета

Хрипунова Константина Сергеевича

Научный руководитель
доцент, к.ф.-м.н., доцент

С.П. Шевырев

Заведующий кафедрой
зав.каф., д.ф.-м.н., доцент

Ю.А. Блинков

Саратов 2023

Введение. Вирусы являются внутриклеточными паразитами человека, животных, растений, насекомых, бактерий, грибов, простейших и других живых существ и могут менять свои свойства как в естественных среде, так и в экспериментах. У животных вирусные инфекции вызывают иммунный ответ, который чаще всего приводит к избавлению от болезнетворного вируса. Иммунный ответ также можно вызвать вакцинами, дающими активный приобретённый иммунитет против конкретной вирусной инфекции. Иногда вирусам удаётся избежать иммунного ответа, вызывая хроническую болезнь. Проблема заболеваемости вирусными инфекциями в период эпидемии требует оперативного решения и активно исследуется, в том числе и с использованием математического аппарата.

С точки зрения математики, динамика эпидемии может быть изложена с помощью системы дифференциальных уравнений, анализа временных рядов или методов теории игр, решения которых могут характеризовать динамику изменения численности в различных подгруппах популяции. При этом важно отметить, что вирус постоянно меняется, что влечет за собой неизбежность изменения самой вакцины. На примере COVID-19 генетическая структура вируса изменяется с такой скоростью, что ко времени окончания разработки новой вакцины она уже может оказаться не нужной. Чтобы избежать этого, необходимо построить модель распространения вирусной инфекции.

Целью дипломной работы является ознакомиться с теорией, разработать web- приложение и провести численный анализ.

Объект исследования — статистические данные регионов РФ.

Предмет исследования — алгоритмы моделирования и прогнозирования статистических данных заболеваемости COVID-19.

Для достижения поставленной цели в работе необходимо решить следующие **задачи**:

- рассмотреть методы теории игр в моделировании эпидемии;
- рассмотреть прогнозирования временных рядов;
- исследовать классическую модель эпидемии, а также ее модификации;
- провести численное моделирование модели методом Эйлера, Рунге-Кутты Адамса;
- рассмотреть теорию парсинга;

- написать парсер сайта стопкоронавирус.рф;
- исследовать создание картинок в формате SVG;
- разработать карту регионов РФ;
- применить модель к полученным данным;
- провести анализ полученных результатов, с помощью разработанного сервиса.

Практическая значимость. Исследован способ моделирования статистических данных заболеваемости каждого региона РФ и написан код, позволяющий смоделировать и идентифицировать число неинфицированных, зараженных, вылечившихся и умерших индивидуумов, что позволит адекватно прогнозировать дальнейшее состояние эпидемии.

Структура и содержание дипломной работы. Работа состоит из введения, десяти разделов, заключения, списка использованных источников, содержащего 20 наименований, и семи приложений. Общий объем работы составляет 63 страницы.

Моделирование распространения эпидемии

Пусть каждый агент популяции может находиться в одной из трех состояний:

1. восприимчивый;
2. инфицированный;
3. иммунный.

Применимо к терминам теории игр, представителей популяции будем называть игроками, а состояния, в которых они находятся – стратегиями. Подмножества агентов, находящихся в одной и том же состоянии, образуют подгруппы. Считается, что взаимодействие игроков каждой из подгрупп носит случайный характер и может быть описано симметричной игрой двух лиц.

Пусть $K = \{S, I, R\}$ – это множество чистых стратегий, где:

- S – чистая стратегия подверженных заболеванию лиц;
- I – чистая стратегия инфицированных;
- R – чистая стратегия иммунных.

Также введем следующие обозначения:

- n_i – число агентов популяции, использующих i -ую стратегию;
- N – число агентов популяции;

- $x_i = \frac{n_i}{N}, i = 1, 2, 3$ — доля агентов популяции, использующих i -ую стратегию;
- $x = (x_s, x_i, x_r)$ — вектор состояния популяции;
- $\Delta = \{(x_1, \dots, x_n) : \sum_{i=1}^n x_i = 1, x_i \geq 0\}$ — множество состояний популяции.

Попарное взаимодействие агентов, вызывающее изменение их состояния, напрямую влияет на их доход. Допустим, что в подгруппе подверженных заболеванию проведена вакцинация, стоимость которой составила q . При этом доход каждого агента равен β . При взаимодействии восприимчивого агента с представителем своей же подгруппы или с представителем подгруппы иммунных его доход уменьшится на затраты на вакцинацию. Как уже было описано в предположениях модели, взаимодействие агентов носит случайный характер, в результате которого каждый из агентов может либо сохранить свой доход на прежнем уровне, либо потерять часть дохода за счет издержек на лечение. При взаимодействии восприимчивого агента с представителем подгруппы инфицированных с вероятностью θ вакцина не даст должного эффекта. В таком случае данный игрок перейдет в подгруппу инфицированных с интенсивностью δ , при этом он потеряет часть своего дохода вследствие появления затрат на лечение. Представитель подгруппы инфицированных может перейти в подгруппу иммунных с вероятностью ψ , при этом уровень его дохода может увеличиться на γ единиц, поскольку он не будет нести затрат на лечение. При взаимодействии двух представителей подгруппы инфицированных состояние каждого из них может стать хуже с вероятностью ϕ . Также при взаимодействии представителя подгруппы иммунных с представителями любой другой подгруппы его состояние, а, следовательно, и доход останутся прежними. В конечном итоге можно сформировать матрицу выигрышей в соответствии с таблицей 1, которая описывает всевозможные варианты взаимодействий:

В данной таблице присутствуют следующие обозначения:

- β — доход игрока;
- q — затраты на вакцинацию;
- c — затраты на лечение;

Таблица 1 — Матрица всевозможных взаимодействий

h	S	I	R
S	$(\beta - q, \beta - q)$	$(\beta - q - c\theta\delta, a + \gamma)$	$(\beta - q, \beta)$
I	$(a + \gamma, \beta - q - c\theta\delta)$	$(a + \gamma - \sigma, a + \gamma - \sigma)$	$(a + \gamma, \beta)$
R	$(\beta, \beta - q)$	$(\beta, a + \gamma)$	(β, β)

- θ — вероятность неэффективности вакцины;
- δ — интенсивность перехода из подгруппы восприимчивых в подгруппу инфицированных;
- $a = (\beta - c)$ — доход агента, сократившийся за счет затрат на лечение;
- $\gamma = c\phi$ — рост дохода при выздоровлении игрока;
- ϕ — интенсивность перехода из подгруппы инфицированных в подгруппу иммунных;
- $\sigma = c\psi$ — сокращение дохода при ухудшении состояния;
- ψ — вероятность подобного ухудшения при взаимодействии двух агентов из подгруппы инфицированных.

Анализ временных рядов Чтобы спрогнозировать дальнейший процесс эпидемии, можно использовать анализ временных рядов. Для того, чтобы применить данный алгоритм, достаточно всего лишь небольшого объема выборки за определенный период.

В отличие от стандартной модели Хольта, модифицированная модель учитывает экспоненциальный тренд и аддитивную сезонность.

Пусть дан временной ряд $y_i, \dots, y_t, \quad y_i \in R$. Необходимо решить задачу прогнозирования временного ряда.

$$y_{t+d}^{\hat{}} = a_t(r_t)^d\theta_t + (d \bmod s) - s, \quad (1)$$

$$a_t = \alpha_1 \cdot \frac{y_i}{\theta_{t-s}} + (1 - \alpha_1)a_{t-1}r_{t-1}, \quad (2)$$

$$r_t = \alpha_3 \cdot \frac{a_i}{a_{i-1}} + (1 - \alpha_3)r_{t-1}, \quad (3)$$

$$\theta_t = \alpha_2 \cdot \frac{y_i}{a_i} + (1 - \alpha_2)\theta_{t-s}, \quad (4)$$

где s — период сезонности, θ_i , $i \in 0, \dots, s - 1$ — сезонный профиль, r_t — параметр тренда, a_t — параметр прогноза, очищенный от влияния тренда и сезонности.

Оптимальные параметры α_1 , α_2 , $\alpha_3 \in (0, 1)$ предлагается находить экспериментальным путем. Первая часть данных использовалась для построения модели, а на основе остальных данных проверялась точность прогнозирования.

Классическая модель эпидемии и ее модификации

Традиционной моделью, описывающей процесс развития эпидемии, является модель SIR (Susceptible-Infected-Recovered), предложенная У. Кермаком и А. Маккендриком в 1927 г. Авторы разделили всю популяцию на три подгруппы:

1. $S(t)$ — восприимчивые к инфекции, количество неинфицированных к моменту времени t ;
2. $I(t)$ — инфицированные;
3. $R(t)$ — излечившиеся от болезни.

Модель эпидемии описывается с помощью следующих дифференциальных уравнений:

$$\frac{dS(t)}{dt} = -\beta SI, \quad (5)$$

$$\frac{dI(t)}{dt} = \beta SI - \gamma I, \quad (6)$$

$$\frac{dR(t)}{dt} = \gamma I, \quad (7)$$

где β — частота контактов между представителями, $1/\gamma$ — средняя продолжительность периода заболевания.

Данная модель основана на предположении, что скорости заражения инфекцией и излечения от нее значительно превосходят темпы рождаемости и смертности, вследствие чего упомянутые факторы в модели игнорируются.

В оригинальной модели присутствует лишь три состояния. Это число восприимчивых, зараженных, излечившихся, однако будем рассматривать еще

одно состояние — число умерших, чтобы модель была более реалистичной при моделировании COVID-19.

Популяцию $N(t)$ в данной модели можно разделить на несколько групп:

1. $S = S(t)$ — восприимчивые, количество неинфицированных к моменту времени t ;
2. $I = I(t)$ — количество инфицированных вирусом;
3. $R = R(t)$ — количество излечившихся от болезни;
4. $D = D(t)$ — количество умерших от вируса.

При этом в каждый момент времени $N(t) = S(t) + I(t) + R(t) + D(t)$.

В течение некоторого рассматриваемого периода индивиды меняют свой статус, перемещаясь из одной группы в другую. Это динамический процесс, сопровождающийся переходом индивида из класса «уязвимых» (восприимчивых) в класс инфицированных, а позже — в класс излечившихся. Вирусная эпидемия — это стремительно развивающийся процесс, поэтому важным вопросом здесь является поиск методов защиты населения во время ежегодно повторяющейся эпидемии. К ним, например, относится введение карантина, а также фармакологические методы (профилактические прививки или противовирусные препараты).

В модели существует ряд предположений:

1. Во внимание не принимаются демографические изменения: рождение и смерть, миграции и т.д. Таким образом, размер популяции считается постоянным на временном отрезке моделирования:

$$S(t) + I(t) + R(t) + D(t) = N(t) = \text{const.}$$

2. Эпидемия начнется, если в начальный момент времени имеются индивиды, которые уже заражены, то есть:

$$I(0) > 0.$$

3. Все индивиды популяции имеют одинаковую вероятность заражения:
- δ — интенсивность заражения при контакте с инфицированными вирусом V , $0 \leq \delta \leq 1$;
4. Болезнь передаётся только при общении с инфицированным.

5. Контакт с заболевшим вызывает заражение, если контактирующий не имеет приобретенного иммунитета к болезни.
6. Врождённый иммунитет отсутствует.
7. Инкубационный период заболевания пренебрежимо мал, т.е. заболевание в данной модели происходит мгновенно.
8. Доля выздоровевших постоянна:
 - σ – интенсивность перехода из подгруппы инфицированных вирусом V в подгруппу излечившихся, $0 \leq \sigma \leq 1$;
9. В результате болезни всегда приобретается иммунитет.
10. Заболевание никогда не приводит к смерти.

Для упрощения вычислений будем рассматривать доли восприимчивых, зараженных, излечившихся и скончавшихся. Обозначим:

1. $s(t) = \frac{S}{N}$ — доля восприимчивых (неинфицированных),
2. $i(t) = \frac{I}{N}$ — доля зараженных вирусом,
3. $r(t) = \frac{R}{N}$ — доля излечившихся,
4. $d(t) = \frac{D}{N} = 1 - s(t) - i_1(t) - r(t)$ — доля скончавшихся.

Для защиты населения можно использовать различные методы, такие как введение карантина или вакцинирование населения.

Во время эпидемии состояние системы изменяется с течением времени, что можно описать следующей системой обыкновенных дифференциальных уравнений:

$$\frac{ds}{dt} = -\delta si, \quad (8)$$

$$\frac{di}{dt} = (\delta s - \sigma) \cdot i, \quad (9)$$

$$\frac{dr}{dt} = \sigma \cdot i, \quad (10)$$

$$\frac{dd}{dt} = \mu \cdot i, \quad (11)$$

При этом в начале процесса имеем начальные состояния

$$s(0) = s_0, i(0) = i_0, r(0) = r_0, d(0) = d_0 = 1 - s(0) - i(0) - r(0).$$

Итоговый сервис

Стек проекта в соответствии с рисунком 1 выглядит следующим образом:

1. Производится сборка данных с сайта стопкоронавирус с помощью Python
2. Далее данные закидываются в базу данных с помощью PostgreSQL
3. С базы данных Nodejs с помощью get запроса закидывает их на клиент React
4. Внутри React производится построение модели с помощью JSX

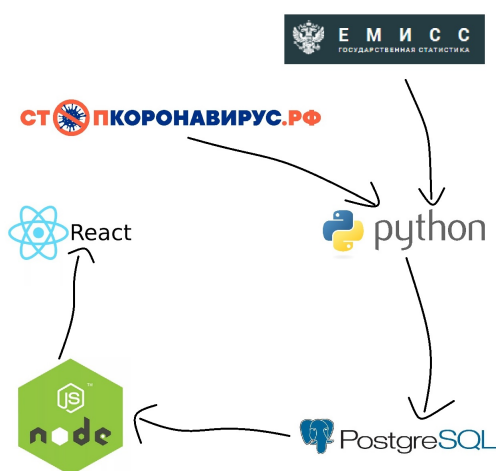


Рисунок 1 — Стек сервиса

Также, в основу сборки данных был положен парсинг с стопкоронавирус.рф, а также список населения каждого региона с сайта ЕМИСС.

Так как в дальнейшем данные перекидываются на сервер PostgreSQL, то пришлось его немного видоизменить, а именно:

1. Добавить функцию, которая будет перекидывать число населения каждого региона из Excel таблицы сайта ЕМИСС в базу данных PostgreSQL.
2. Добавить функцию, которая будет перекидывать с парсера число инфицированных, умерших или находящихся в зоне риска в PostgreSQL

В соответствии с рисунком 2, получилось множество таблиц, каждая из которых описывает число случаев за день для любого региона РФ.

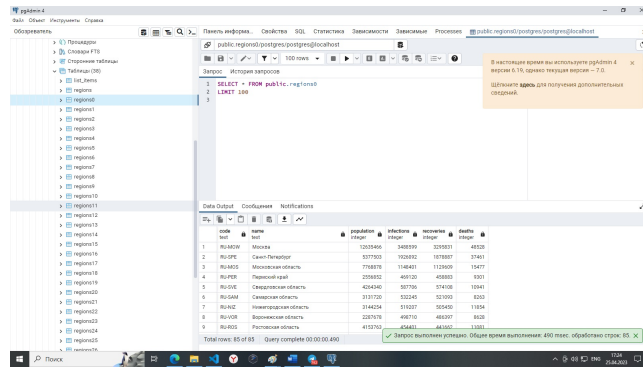


Рисунок 2 — Стек сервиса

Посредством get - запроса с помощью Nodejs забираются данные из бд в формате JSON. Причем сервер будет работать на 3001 порту, а клиентская часть сервиса на 3000.

```

1 app.get('/days', (req, res) => {
2   res.json({
3     array: data
4   })
5 })

```

Далее, на стороне клиента, принимаем данные, и реализуем их в виде таблицы и графиков в соответствии с рисунком 3. Причем, в процессе разработки веб-сервиса пришлось отказаться от построения SIRD модели в пользу модели Хольта-Уинтерса, поскольку модель SIRD предназначена для построения эпидемии с начальными данными, которые получить практически невозможно. Коды этих алгоритмов представлены в приложении. Ниже представлено применение данного метода для группы зараженных индивидуумов.

```

1 let [alpha, beta, gamma, err] =adjustParams(arrInfTh, period, m)
2 var predictionsInf = [];
3 predictionsInf = predict(arrInfTh, alpha, beta, gamma, period, m);

```

Причем подбор коэффициентов в экспоненциальной модели производится путем полного перебора с использованием среднеквадратичной ошибки:

```

1 function computeMeanSquaredError(data, pred) {
2   var cmse = 0
3   for (let i =0; i<28;i++){

```

```

4     cmse += Math.pow(data[i]-Math.round(pred[i],2), 2)
5   }
6   cmse = cmse/cmse.length
7   return cmse
8 }

```



		Всего За сутки	
Регион	Красноярский край	Заражений	438034 79
Население	2849169	Выздоровлений	425494 36
		Смертей	11328 0

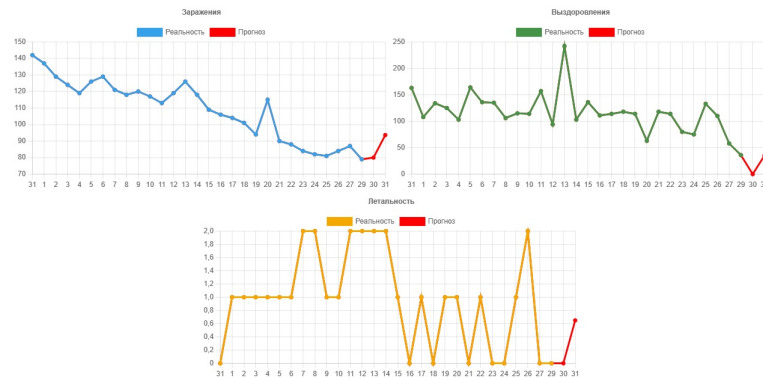


Рисунок 3 — Веб-сервис

Анализ результатов.

Рассмотрим построение на примере Саратовской области в соответствии с рисунком 4. Выясним значения коэффициентов модели Хольта-Уинтерса. Изначально полагаем, что модель имеет четыре периода по 7 дней. Это означает что можно спрогнозировать значения на два дня вперед.

В соответствии с рисунком 5 найдем коэффициенты модели для графика заражений с помощью console.log.

		Всего За сутки	
Регион	Саратовская область	Заражений	321778 101
Население	2360959	Выздоровлений	314809 183
		Смертей	6001 0

Рисунок 4 — Статистика Саратовской области

```

1 console.log('Заражение')
2 console.log(alpha)
3 console.log(beta)
4 console.log(gamma)

```



Рисунок 5 — Статистика Заражений Саратовской области

Видно, что график имеет нисходящий тренд

Часть графика синего цвета — это реальная статистика по Саратовской области. Та часть, которая выделена красным - прогнозируемые значения на два дня вперед.

Таким образом на сегодня — 30 апреля, количество заражений равно 101 случаю. На завтра — 1 мая, прогнозируемое количество заражений равно 99. И на 2 мая прогнозируемое количество заражений равно 122. Можно дождаться 1 и 2 мая, и уже на реальных данных оценить правильность предсказания модели, но также можно сравнить с имеющейся выборкой, что и было сделано. Таким образом коэффициенты самой модели α , β , γ равны 0.1, 0.3, 0.1 соответственно.

Аналогичным образом в случае выздоровления в соответствии с рисунком, коэффициент модели равны 0.1, 0.7, 0.1.

Аналогичным образом в случае смертности в соответствии с рисунком, коэффициент модели равны 0, 0, 0, поскольку сама смертность равна нулю за весь период.



Рисунок 6 — Статистика Выздоровлений Саратовской области

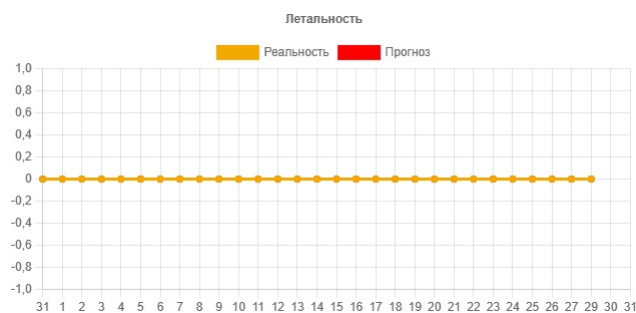


Рисунок 7 — Статистика смертности Саратовской области

Заключение. В заключении приведены результаты дипломной работы:

- рассмотрены методы теории игр в моделировании эпидемии;
- рассмотрены методы анализа временных рядов;
- исследована классическая модель эпидемии, а также ее модификации;
- проведено численное моделирование модели методом Эйлера, Рунге-Кутты, Адамса;
- рассмотрена теория парсинга;
- написан парсер сайта стопкоронавирус.рф;
- исследовано создание картинок в формате SVG;
- разработана карта регионов РФ;
- применена модель к полученным данным;
- проведен анализ полученных результатов, с помощью разработанного сервиса.