

Федеральное государственное автономное образовательное учреждение
высшего образования
«Московский физико-технический институт
(национальный исследовательский университет)»
Физтех-школа Прикладной Математики и Информатики
Кафедра интеллектуальных систем

Направление подготовки / специальность: 03.04.01 Прикладные математика и физика

Направленность (профиль) подготовки: Математическая физика, компьютерные технологии и математическое моделирование в экономике

ОБУЧЕНИЕ МОДЕЛЕЙ РАСПРОСТРАНЕНИЯ РИСКОВ НА ГРАФАХ

(магистерская диссертация)

Студент:

Рогозина Анна Андреевна

(подпись студента)

Научный руководитель:

Воронцов Константин Вячеславович,
д-р физ.-мат. наук

(подпись научного руководителя)

Консультант (при наличии):

(подпись консультанта)

Москва 2021

Аннотация

Рассматривается моделирование распространения инфекции в популяции. В работе предлагаются модели, оценивающие риск заражения для каждого индивида популяции. Проведены эксперименты, исследующие качество получившихся моделей, а так же эффективность противоэпидемиологических мер на основе предсказаний моделей.

Ключевые слова: модель распространения инфекции; имитационная модель; индивидуальное оценивание рисков; логистическая регрессия.

Содержание

1	Введение	4
2	Постановка задачи	5
2.1	Задача оценивания рисков на графах	5
3	Имитационные модели распространения инфекций	6
3.1	Модель SI	6
3.2	Модель SIS	6
3.3	Модель SIR	7
3.4	Модель SEIRS	7
4	Модели оценивания рисков	8
4.1	Частотная модель	8
4.2	Учет вероятности передачи инфекции	9
4.3	Рекурсивное оценивание рисков	9
4.4	Модель распространения рисков по сети	10
4.5	Алгоритм распространения рисков по сети	10
5	Эксперименты	12
5.1	Данные	12
5.2	Сравнение моделей на полусинтетических данных	12
5.2.1	Имитационная модель	12
5.2.2	Детали процесса обучения	12
5.2.3	Результаты	13
5.3	Противоэпидемиологические меры	13
5.3.1	Стратегии вакцинации	15
5.3.2	Вакцинация на полусинтетических данных	15
5.3.3	Генерирование графа	15
5.3.4	Имитационная модель	17
5.3.5	Обучение моделей	17
5.3.6	Результаты	18
6	Заключение	20

1 Введение

Задача моделирования распространения рисков имеет довольно широкое применение: в широком смысле это, например, моделирование распространения инфекций в популяциях[1], моделирование влияния новостей на общество[3], распространение паники в группе[2], и распространения компьютерных вирусов[4].

В работе[1] задача моделирования распространения рисков решается для популяции в целом, отслеживается, например, динамика доли заболевших людей во времени. Модели основаны на системах дифференциальных уравнений или их разностных аналогах. Такой подход позволяет генерировать распространение эпидемии в среднем на большом временном промежутке, но не позволяет понять риск для конкретного индивида в конкретный момент времени. Знание об индивидуальных рисках в популяции позволяет более эффективно препятствовать распространению эпидемии. С появлением данных о реальных графах контактов множества индивидов появилась возможность обучать модели, предсказывающие индивидуальный риск во времени.

В данной работе предложены модели индивидуальных рисков, анализируется их качество и оценивается применимость моделей для повышения эффективности противоэпидемиологических мер.

2 Постановка задачи

2.1 Задача оценивания рисков на графах

Пусть $\mathbf{X} = \{x_1, \dots, x_n\}$ - множество индивидов.

Пусть $T = \{1, 2, \dots, t, \dots, T\}$ - дискретные отрезки времени. Пусть в каждый момент t у каждого индивида x есть состояние $y_t(x)$: $y_t(x) = \mathbf{I}$ - инфицированный или $y_t(x) = \mathbf{S}$ - здоровый. Необходимо для каждого индивида x и момента t оценивать вероятность, что индивид инфицирован, $p_t(x) = \mathbf{P}(y_t(x) = \mathbf{I})$.

Обучающая выборка D состоит из потока монотонных по времени записей. Запись содержит информацию о контакте индивида u с индивидом v в момент времени t , и признаках этого контакта. Предполагается, что в каждый момент времени известны состояния всех индивидов в обучающей выборке.

Оптимизационным критерием является максимизация логарифма правдоподобия модели:

$$\mathbb{L}(w) = \frac{1}{D} \sum_{(t,x) \in D} [y_t(x) = \mathbf{I}] \log(p_t(x, w)) - [y_t(x) \neq \mathbf{I}] \log(1 - p_t(x, w)) \rightarrow \max_w \quad (2.1.1)$$

В данной работе рассматриваются несколько последовательно усложняющихся моделей, сравнивается их качество, применимость для противоэпидемиологических мер и устойчивость.

3 Имитационные модели распространения инфекций

Чтобы смоделировать распространение инфекции по известному графу контактов, используются имитационные модели. В данной работе рассматриваются разностные имитационные модели, разделяющие популяцию на группы восприимчивости.

3.1 Модель SI

Дана популяция в N индивидов. Вся популяция делится на три группы: S - (*susceptible*) – восприимчивые, то есть здоровые люди, которые могут заразиться, I - (*infected*) - инфицированные, люди, которые являются носителями инфекции и заражают окружающих. Предполагается, что, заболев однажды, человек уже не может выздороветь. Таким образом, множество заболевших только расширяется во времени. Так моделируются не только эпидемии, но и, например, распространение новостей в информационном пространстве[3].

Вводится параметр α - вероятность заразиться при контакте больного и здорового. Вероятность контакта между индивидами полагается равной $\frac{1}{N}$. Так мы получаем систему дифференциальных уравнений:

$$\begin{cases} \frac{dS}{dt} = -\frac{\alpha IS}{N} \\ \frac{dI}{dt} = \frac{\alpha IS}{N} \end{cases} \quad (3.1.1)$$

Или ее разностный аналог

$$\begin{cases} S_{n+1} = S_n - \frac{\alpha S_n I_n}{N} \\ I_{n+1} = I_n + \frac{\alpha S_n I_n}{N} \end{cases} \quad (3.1.2)$$

3.2 Модель SIS

Группы такие же, как в модели SI , но теперь зараженный (I) по прошествии какого-то времени может стать восприимчивым (S). Кроме параметра α добавляется еще параметр λ - вероятность, что за промежуток времени зараженный выздоровеет. Скорректированная система дифференциальных уравнений:

$$\begin{cases} \frac{dS}{dt} = -\frac{\alpha IS}{N} + \lambda I \\ \frac{dI}{dt} = \frac{\alpha IS}{N} - \lambda I \end{cases} \quad (3.2.1)$$

Или ее разностный аналог

$$\begin{cases} S_{n+1} = S_n - \frac{\alpha S_n I_n}{N} + \lambda I_n \\ I_{n+1} = I_n + \frac{\alpha S_n I_n}{N} - \lambda I_n \end{cases} \quad (3.2.2)$$

3.3 Модель SIR

Вся популяция делится на три группы: кроме восприимчивых (S) и инфицированных (I), появляются еще выздоровевшие R - (*recovered*) R - (*recovered*), то есть люди, у которых есть иммунитет и которые больше не заболеют. Вводятся три параметра: α - вероятность заразиться при контакте инфицированного и здорового, вероятность контакта между индивидами (полагается равной $\frac{1}{N}$) и γ - скорость перехода из состояния I , «заболевший» в состояние R , выздоровевший. Таким образом, изменение количества заболевших, восприимчивых и выздоровевших, можно описать системой дифференциальных уравнений:

$$\begin{cases} \frac{dS}{dt} = -\frac{\alpha IS}{N} \\ \frac{dI}{dt} = \frac{\alpha IS}{N} - \gamma I \\ \frac{dR}{dt} = \gamma I \end{cases} \quad (3.3.1)$$

Или ее разностным аналогом

$$\begin{cases} S_{n+1} = S_n - \frac{\alpha S_n I_n}{N} \\ I_{n+1} = I_n + \frac{\alpha S_n I_n}{N} - \gamma I_n \\ R_{n+1} = R_n + \gamma I_n \end{cases} \quad (3.3.2)$$

3.4 Модель SEIRS

Данная модель учитывает наличие инкубационного периода у вирусных инфекций: к группам восприимчивых (S), заболевших (I), выздоровевших (R) добавляется группа зараженных, E (*exposed*) - группа зараженных, но пока не заразных индивидов. Вместо перехода *восприимчивый* (S) \rightarrow *заболевший* (I) \rightarrow *выздоровевший* (R) индивиды из состояния *восприимчивый* (S) переходят в состояние *зараженный* (E) и только затем в состояние *заболевший* (I). Добавляется также параметр δ - вероятность, что за день *зараженный* станет *заболевшим*, то есть произойдет переход ($E \rightarrow I$), параметр ξ вероятность, что *выздоровевший* (R) снова станет *восприимчивый* (S) и параметр μ вероятность

контакта извне популяции. Система дифференциальных уравнений корректируется следующим образом:

$$\begin{cases} \frac{dS}{dt} = -\frac{(\alpha+\mu)IS}{N} + \xi R \\ \frac{dE}{dt} = \frac{(\alpha+\mu)IS}{N} - \sigma E \\ \frac{dI}{dt} = \sigma E - \gamma I \\ \frac{dR}{dt} = \gamma I - \xi R \end{cases} \quad (3.4.1)$$

Разностный аналог корректируется так:

$$\begin{cases} S_{n+1} = S_n - \frac{(\alpha+\mu)S_n I_n}{N} + \xi R_n \\ E_{n+1} = E_n + \frac{(\alpha+\mu)I_n S_n}{N} - \delta E_n \\ I_{n+1} = I_n + \delta E_n - \gamma I_n \\ R_{n+1} = R_n + \gamma I_n - \xi R_n \end{cases} \quad (3.4.2)$$

В работе модель SEIR используется в качестве имитационной модели распространения инфекции, предположение о моделировании распространения инфекции берется из модели SIS.

4 Модели оценивания рисков

4.1 Частотная модель

В качестве основного предположения о распространении инфекции возьмем разностное уравнение протекания заражения из модели SIS[1]. В нем вероятность инфицирования индивида x в момент времени t через две составляющие: $q_t(x)$ - вероятность получить инфекцию в интервал $[t-1, t]$, и $p_{t-1}(x)$ - вероятность инфицирования ранее:

$$p_t(x) = (1 - \mu)p_{t-1}(x) + \beta(1 - p_{t-1}(x))q_t(x) \quad (3.1)$$

Вероятность инфицирования в интервал $[t-1, t]$ можно оценить моделью логистической регрессии, основанной на количестве контактов в интервал времени:

$$q_t(x) = \sigma(w_1 k_t(x) - w_0) \quad (3.2)$$

$$k_t(x) = \sum_{\langle t:(x,v) \rangle} [t-1 \leq t' \leq t] \quad (3.3)$$

Данная модель не учитывает что у разных контактов x может быть как разная вероятность передачи инфекции x , так и разная вероятность инфицирования от x . Кроме того, модель не учитывает, что при получении

информации о смене состояния x , вероятности всех недавно контактировавших с ним индивидов u так же должны измениться.

4.2 Учет вероятности передачи инфекции

Добавим в оценку количества контактов (формула 3.3) оценку вероятности передачи инфекции при данном контакте:

$$k_t(x) = \sum_{\langle t:(x,v) \rangle} [t - 1 \leq t' \leq t] a_{t'}(x, v) \quad (3.4)$$

Вероятность передачи инфекции $a_{t'}(x, v)$ оцениваем так же с помощью модели логистической регрессии:

$$a_{t'}(x, v) = \sigma(-\alpha_0 + \sum_{j=1}^m \alpha_j f_j) \quad (3.5)$$

Здесь α_j - параметры модели, f_j - различные признаки контакта, например

- Длительность контакта
- Уровень сигнала устройства, зарегистрировавшего контакт
- Количество контактов за отрезок времени

В этой модели учитываются что различный характер контакта по-разному влияет на вероятность инфицирования в момент t , но не учитывается, что вероятность инфицирования зависит еще и от того, болен ли или здоров второй участник контакта.

4.3 Рекурсивное оценивание рисков

Добавим в оценку количества контактов (формула 3.3) вероятность что второй участник контакта болен, $\tilde{p}_t(v)$:

$$k_t(x) = \sum_{\langle t:(x,v) \rangle} [t - 1 \leq t' \leq t] a_{t'}(x, v) \tilde{p}_{t'}(v) \quad (3.6)$$

$$\tilde{p}_{t'}(v) = \begin{cases} 1, & \text{if } y_{t'}(v) = \mathbf{I} \\ p_{t'}(v), & \text{if } y_{t'}(v) \neq \mathbf{I} \end{cases} \quad (3.7)$$

Для всех предыдущих моделей вероятность инфицирования на шаге t , $p_t(x)$, зависела только от вероятности инфицирования на шаге $t-1$ (формула 3.1). Теперь добавляется зависимость на шаге t , $p_t(v)$. В качестве значения $p_t(v)$ можно либо брать значение с предыдущего шага $p_{t-1}(v)$, либо рекурсивно распространять градиент через суперпозицию функций, ограничившись определенной глубиной суперпозиции d .

4.4 Модель распространения рисков по сети

Предыдущие модели не учитывают, что при смене состояния индивида (появлении результата тестирования на наличие инфекции), например, $y_t(x) = \mathbf{I}$, должны рекурсивно измениться оценки рисков для всех индивидов, контактировавших с x для любого t' из отрезка времени $[t-d, t]$ («контактных» индивидов). После этого по цепочке должны измениться оценки рисков для индивидов, контактировавших с «контактными» индивидами, для моментов времени t'' из интервала $(t', t]$, и так далее. Добавим в формулу оценивания вероятности инфицирования в момент t (формула 3.2) вероятность того, что x будет инфицирован в ближайшем будущем, на интервале $(t, t+d]$:

$$q_t(x) = \sigma(w_1 k_t(x) + w_2 b_t(x) - w_0) \quad (3.8)$$

$$b_t(x) = [t' : t < t' \leq t + d \text{ и } y_{t'}(x) = \mathbf{I}] \quad (3.9)$$

При появлении информации о смене состояния индивида x на I , запускается *Алгоритм распространения рисков по сети*, при этом риск для x скачком увеличивается до 1, $\Delta p_t(x) = 1 - p_t(x)$, а затем увеличиваются оценки рисков для всех контактировавших с x , $p_t(v)$, и так по цепочке контактов.

4.5 Алгоритм распространения рисков по сети

Алгоритм состоит из двух функций: $BakwardUpdate(x, t)$ и $ForwardUpdate(x, t_0, t)$. При смене состояния индивида x запускается функция $BakwardUpdate(x, t)$. В ней для всех контактов x в момент t_0 из интервала $t-d, t$ пересчитывается вероятность $p_{t_0}(x)$ по формуле (3.1). Так как в этой формуле теперь $b_{t_0} = 1$, вероятность $p_{t_0}(x)$ увеличивается на $\Delta p_{t_0}(x)$. Функция $ForwardUpdate(x, t_0, t)$ обновляет оценку рисков для всех контактов x вперед с t_0 до t . Если приращение $\Delta p_{t'}(u)$ было достаточно велико, то запускается пересчет весов и для контактов

$u(\text{ForwardUpdate}(u, t', t))$. Чтобы не допустить зацикливание, вводится множество просмотренных индивидов, и пересчет запускается только для непросмотренных индивидов. Иллюстрация работы алгоритма и псевдокод функций приведен ниже.

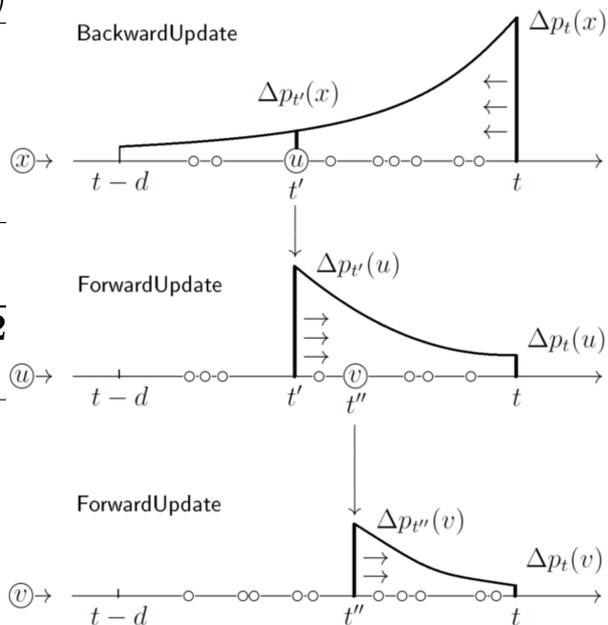
Algorithm 4.1 BackwardUpdate(x, t)

- 1: $U = \emptyset$
 - 2: **for** $t \in [t - d, t] : (x, v)$ **do**
 - 3: Пересчитать $p'_t(x)$
 - 4: **ForwardUpdate**(x, td, t)
-

Algorithm 4.2

ForwardUpdate(x, t_0, t)

- 1: $U := U \cup \{x\}$
 - 2: **for** $t \in [t - d, t] : (x, u \notin U)$ **do**
 - 3: Пересчитать $p'_t(u)$
 - 4: **for** $t \in [t - d, t] : (x, u \notin U)$ **do**
 - 5: **if** $\Delta p'_t(u) > \epsilon$ **then**
 - 6: **ForwardUpdate**(u, t', t)
-



5 Эксперименты

5.1 Данные

Данные - информация о контактах на предприятии с маячков системы Amuleit. Данные в виде $\langle t, (u, v) \rangle$ - контакт индивида u с индивидом v в момент времени t , поток контактов отсортирован по времени. Данных о распространении инфекции нет, поэтому таргеты - статусы индивидов в каждый момент времени - генерируются имитационной моделью SEIRS. Таким образом, в качестве данных есть поток контактов и набор семплов состояний индивидов в каждый момент времени (таргетов).

5.2 Сравнение моделей на полусинтетических данных

5.2.1 Имитационная модель

Так как данных о распространении инфекции нет, мы моделируем распространение инфекции с помощью имитационной модели SEIRS. Задаются такие параметры модели SEIRS:

- Вероятность передачи инфекции $\alpha = 0.9$
- Нельзя заразиться извне популяции, $\mu = 0$
- Вероятность перехода от состояния зараженного (E) к заболевшему (I), $\sigma = 0.64$
- Вероятность выздоровления и приобретения иммунитета $\gamma = 0.14$, то есть в среднем человек болеет 7 дней
- Вероятность потери иммунитета $\xi = 0.09$, то есть иммунитет в среднем приобретает на 11 дней.
- При возникновении заболевания человек продолжает ходить на работу.

5.2.2 Детали процесса обучения

Модели обучаем в парадигме онлайн-обучения: на каждом шаге мы

- Получаем элемент из потока

- Делаем предсказание вероятности инфицирования $p_t(x)$
- Получаем состояние индивида $y_t(x)$
- Считаем ошибку на одном элементе
- Делаем градиентный шаг

В каждой модели итоговая вероятность инфицирования считается по формуле (3.1): $p_t(x) = (1 - \mu)p_{t-1}(x) + \beta(1 - p_{t-1}(x))q_t(x)$. В ней в левой части присутствует вероятность, а в правой - вероятности складываются с множителями-параметрами модели $(1 - \mu)$ и β . Поэтому мы накладываем на μ и β ограничения - $0 \leq \mu, \beta \leq 1$. В качестве оптимизирующего алгоритма используется проективный градиентный спуск.

Мы получаем данные потоком по одному элементу. Так как функция ошибки (формула (2.1.1)) аддитивная относительно элементов выборки, то на каждом шаге мы считаем градиент экспоненциальным скользящим средним.

5.2.3 Результаты

Сравниваются модели со всеми фиксированными гиперпараметрами, кроме параметра скользящего среднего и шага градиента - эти параметры мы оптимизируем по средней точности. При сравнении моделей гиперпараметры настраиваются на одном семпле таргетов, затем считается среднее качество обучения моделей с зафиксированными гиперпараметрами на наборе семплов.

На графике сравнения моделей (Рис. 1) мы видим что усложнения моделей последовательно улучшают их качество. Кроме моделей, на графике представлена следующая частотная эвристика: для каждого индивида x считается количество контактов за последние 5 дней, $k(x, t)$. Вероятностью заражения индивида $p_t(x) = \sigma(k(x, t))$, где $\sigma()$ - эвристическое распределение значений. $k(x, t)$ для всех x за первые 5 дней. Видно, что качество моделей гораздо выше, чем качество эвристики.

5.3 Противоэпидемиологические меры

В качестве противоэпидемиологической меры возьмем вакцинацию. Вакцинация моделируется через имитационную модель: в день вакцинации состояние индивида x меняется на R (выздоровевший) и далее продолжается генерация распространения инфекции.

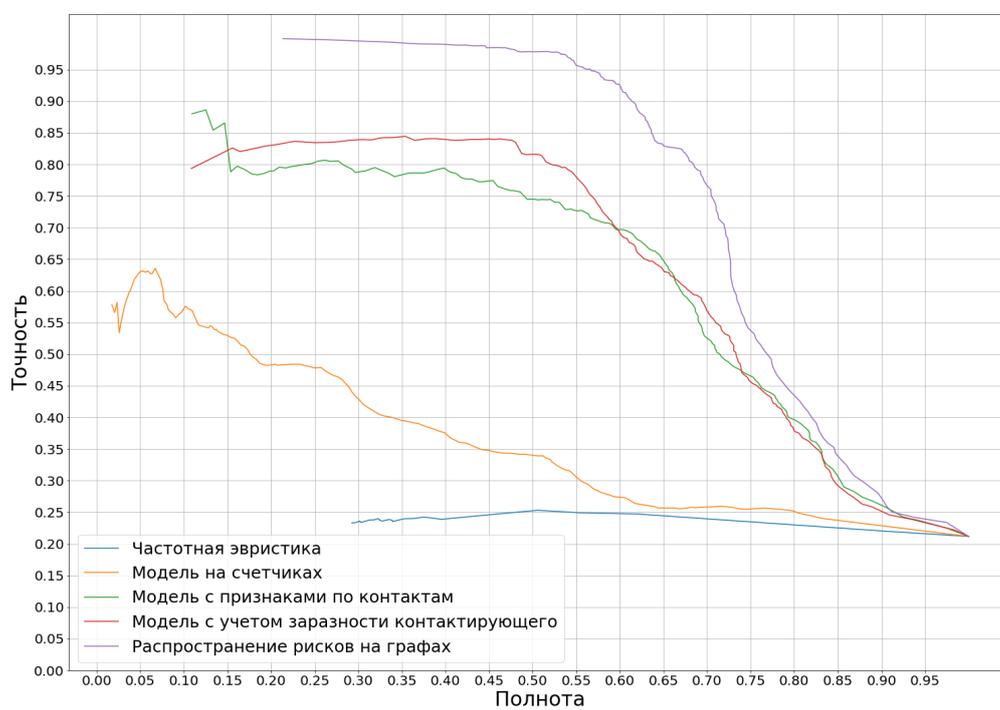


Рис. 1: Сравнение моделей по средней точности-полноте

5.3.1 Стратегии вакцинации

Исследуется эффект от вакцинации популяции. Для этого проварьируем количество вакцинируемых и *стратегии вакцинации* - согласно *стратегии вакцинации* мы выбираем множество индивидов для вакцинации из всех множеств одинаковой мощности. Сравниваются такие *стратегии* вакцинации:

- Отсутствие вакцинации
- Случайный выбор индивидов
- Вакцинация «самых общительных» индивидов, то есть индивидов с наибольшим количеством контактов к моменту вакцинации.
- Вакцинация индивидов с наибольшим риском по предсказаниям модели
- Вакцинация «самых общительных» за весь период наблюдений

Последняя стратегия используется, чтобы понять верхнюю оценку возможной эффективности вакцинации.

5.3.2 Вакцинация на полусинтетических данных

Имитационная модель распространения инфекции такая же, как в разделе. сравнения качества моделей. Вакцинируем 4% населения на 10 й день и по 2% населения с 11 по 14, всего 12%. Сравнение стратегий вакцинации представлено на Рис. 2. На графике видим долю зараженных (E) и заболевших (I) в популяции во времени. Видно, что разница от вакцинации эвристикой с заглядыванием в будущее вакцинацией честной эвристикой очень большая - это значит, что схема контактов неоднородна по дням. Кроме того, видно, что модель распространения рисков приближает качество вакцинации к эвристике со знанием будущего. Однако такой вакцинации недостаточно для того, чтобы побороть эпидемию. Оценим зависимость распространения инфекции от долю вакцинируемой популяции на полностью синтетических данных.

5.3.3 Генерирование графа

Для отслеживания эффекта от противоэпидемиологических мер была сгенерирована последовательность более плотных графов. Статистики длительности контактов берутся из реального графа контактов.

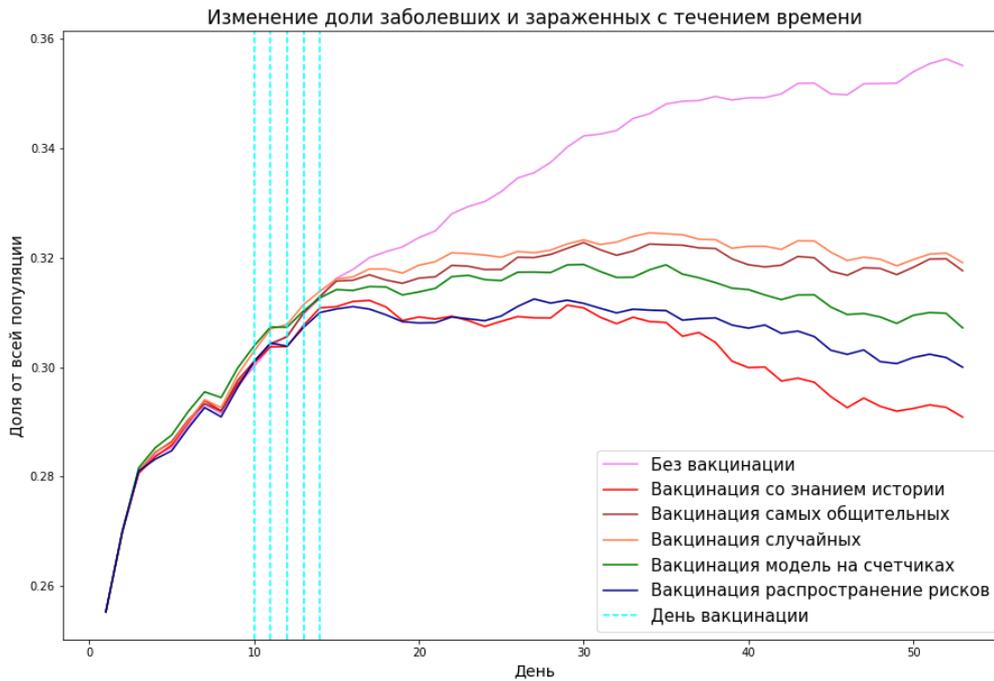


Рис. 2: Сравнение стратегий вакцинации на полусинтетических данных

Для этого сначала был сгенерирован стартовый граф G на $n = 100$ на $n = 100$ вершинах:

Множество вершин V делится случайным нормальным разбиением на кластеры, размеры которых берутся из нормального распределения со средним $s = 3$ и дисперсией $v = 1.5$ (как описано в [5]). Затем внутри кластеров ребра генерируются с вероятностью $p_{in} = 0.06$, вне кластеров ребра генерируются с вероятностью $p_{out} = 0.03$.

Затем ребрам графа присваиваются веса - длительности контактов: выделяется три временные группы: короткие $[15, 30]$, средние $[30, 6]$. С вероятностью соответственно 0.55, 0.35, 0.1 для ребра выбирается одна из временных групп. Затем время из временной группы генерируется равномерно.

Для описания графов контактов каждого дня мы меняем все ребра стартового графа следующим образом:

- С вероятностью 0.42 временная группа ребра остается прежней
- С вероятностью 0.45 временная группа ребра сдвигается на одну вверх или вниз (равновероятно)
- С вероятностью 0.05 временная группа сдвигается вверх или вниз на две

- С вероятностью 0.08 ребро удаляется

5.3.4 Имитационная модель

Для моделирования распространения инфекции в модели SEIRS берутся такие параметры (подставляются в формулу (3.4.2)):

- Вероятность передачи инфекции α от заболевшего (I) к восприимчивому (S) зависит от длительности контакта: при контакте до 1 часа $\alpha = 0.65$, от часа до двух $\alpha = 0.81$, от двух до трех $\alpha = 0.9$ от четырех до пяти $\alpha = 0.95$, далее $\alpha = 0.98$.
- Вероятность заразиться извне популяции $\mu = 0.02$.
- Вероятность перехода от состояния зараженного (E) к заболевшему (I), $\sigma = 0.3$
- Вероятность выздоровления и приобретения иммунитета $\gamma = \frac{1}{16}$, то есть в среднем человек болеет 16 дней
- Вероятность потери иммунитета, $\xi = \frac{1}{180}$, то есть иммунитет в среднем приобретается на 6 месяцев.
- При возникновении заболевания человек с вероятностью 0.3 плохо себя чувствует, уходит домой и ни с кем не контактирует.

В эксперименте для ускорения течения инфекции переход от n к $n + 1$ в формуле (3.4.2) происходит 3 раза в день.

5.3.5 Обучение моделей

Необходимо применить предсказания модели для выбора множества вакцинации. Поэтому изменим таргеты: ранее мы в момент t оценивали вероятность того, что индивид x болен в момент t , $p_t(x)$. Для определения эффективности вакцинации будем в момент t предсказывать вероятность того, что индивид станет заболевшим (I) в течение отрезка $[t + \text{incubation period}, t + \delta]$. Так мы смоделируем условия, в которых хотим проводить вакцинацию: мы можем вакцинировать только здоровых восприимчивых (S) индивидов, и мы хотим вакцинировать индивидов, которые с большой вероятностью заболеют через некоторое время. В эксперименте $\text{incubation period} = 2$, $\delta = 30$.

5.3.6 Результаты

На Рис. 3 сравнивается вакцинация 50% популяции, 30% популяции и 20% популяции различными стратегиями вакцинации. Видно, что при вакцинации 50% популяции результат для осмысленных стратегий различается в абсолютных величинах меньше всего. Это происходит, так как вакцинируются практически все восприимчивые индивиды.

Кроме того, на графике видно, что при уменьшении доли вакцинации уменьшается зазор между осмысленными стратегиями и случайной стратегией - чем меньше влияния на систему, тем меньше разница.

Видно, что различие в зазоре между осмысленными стратегиями и случайной сильно падает при переходе от 30% к 20%. Это позволяет предположить, что доля «самых общительных» индивидов, заражающих большинство других индивидов, находится между 20% и 30% - при вакцинации в 30% осмысленные стратегии покрывают множество «общительных», а при вакцинации 20% уже не покрывают.

Видно, что, при всех схемах вакцинации самые эффективные стратегии - эвристики со знанием будущего и без знания будущего. Из того, что эти две эвристики дают одинаковый результат, можно сделать вывод, что сгенерированная последовательность графов слишком однородна по времени: по первым 5 ти дням можно полностью предсказать контакты последующих 50 ти.

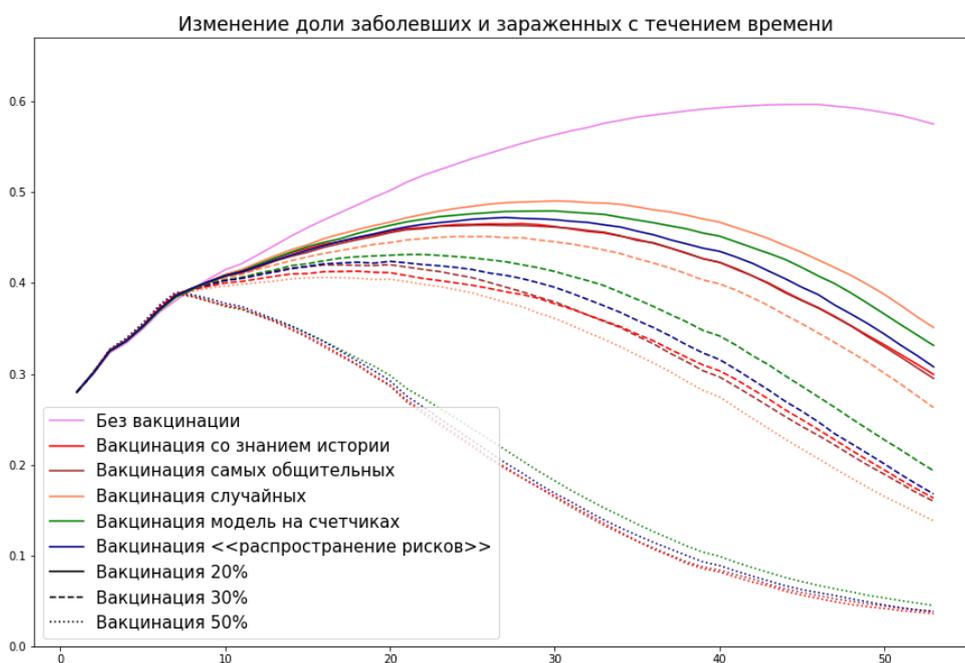


Рис. 3: Сравнение стратегий вакцинации на синтетических данных

6 Заключение

В работе были предложены модели оценки индивидуального риска во времени и алгоритм их обучения.

На полусинтетических данных показано, что итеративное усложнение моделей улучшает качество предсказания.

На синтетических данных исследована эффективность моделей для организации противоэпидемиологических мер. Показано, какой процент популяции достаточно вакцинировать, чтобы предотвратить распространение эпидемии. Оценено влияние предсказаний оценки индивидуального риска на эффективность вакцинации

Список литературы

- [1] *Kermack W., McKendrick A.* A contribution to the mathematical theory of epidemics. Proc.R. Soc. 1927. Vol. A115. P. 700–721.
- [2] *L., Song W., Lv W., Lo S.* Simulation of emotional contagion using modified SIR model: A cellular automaton approach //Physica A: Statistical Mechanics and its Applications. 2014. Vol. 405. P. 380–391
- [3] *Chen N., Zhu X., Chen Y.* Information spreading on complex networks with general group distribution // Physica A: Statistical Mechanics and its Applications. 2019. Vol. 523. P. 671–676.
- [4] *Pastor-Satorras R., Vespignani A.* Epidemic Spreading in Scale-Free Networks // Physical Review Letters. 2001, Apr. Vol. 86. P. 3200–3203.
- [5] *Brandes, U. et al.* Experiments on Graph Clustering Algorithms. ESA (2003)