

Обучение моделей распространения рисков по графам

Рогозина Анна

Московский Физико-технический институт

Физтех-школа прикладной математики и информатики

Кафедра интеллектуальных систем

Научный руководитель:

д. ф.-м. н.

Воронцов Константин Вячеславович

16 июня 2021 г.

Задача

Поток данных:

- $\langle t, (u, v) \rangle$ - контакт индивидов u и v в момент t
- (допущение) для каждого индивида x и момента t знаем состояние индивида

Обучающая выборка:

$$X^l = (x_i, y_i)_{i=1}^l, x_i \in R, y_i \in \{-1, 1\}$$

Найти:

$p(y | x, t)$ - вероятность инфицирования индивида x в момент t

Критерий

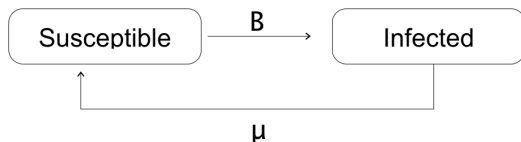
Максимизация логарифмического правдоподобия

$$\sum_{t,x} ([y_t(x) = I] \log(p_t(x)) + [y_t(x) \neq I] \log(1 - p_t(x))) \rightarrow \max$$

Компартментные модели

Модель SIS

Делит популяцию на группы. Позволяет отследить распространение инфекции только «в среднем»



Распространение инфекции по модели SIS

$$\begin{cases} \frac{dS}{dt} = -\beta IS + \mu I \\ \frac{dI}{dt} = -\mu I + \beta IS \end{cases}$$

Адаптируем модель SIS для оценки индивидуального риска

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

$p_t(x)$ - $P(I | x, t)$, вероятность что x инфицирован в момент t

$q_t(x)$ - вероятность что x передали инфекцию в момент t

Оценим вероятность передать инфекцию логистической регрессией на количестве контактов

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

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

Модификации

Заменяем число контактов на вероятность передачи инфекции в отрезке $[t - 1, t]$:

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

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

Здесь f_j - информация о контакте:

- Количество контактов за день
- Длительность контактов

Модификации

Вводим вероятность того то контактирующего индивид инфицирован

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

$$\tilde{p}_t(v) = \begin{cases} 1, & \text{если } y_{t-1}(v) = I \\ p_{t-1}(v), & \text{иначе} \end{cases}$$

Распространение рисков по графам

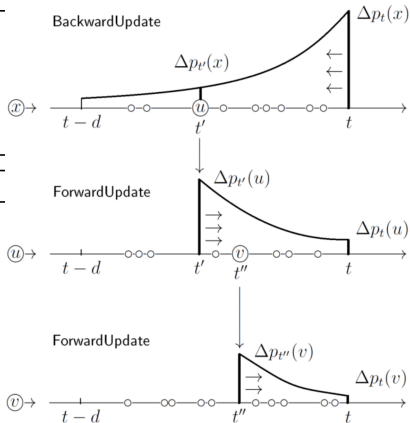
- Когда из потока мы узнаем о смене статуса $y_t(x)$ с S на I , увеличивается риск $p_t(x)$, а так же риски $p_{t'}(x)$ в прошлом
- Введем индикатор что x будет инфицирован в $(t, t + d]$:
 $b_t(x) = [t' : t < t' \leq t + d, y_{t'}(x) = I]$
 $q_t(x) = \sigma(w_1 k_t(x) + b_t(x) w_2 - w_0)$
- При изменении $y_t(x)$ с S на I будем пересчитывать риски в прошлое для x и контактных индивидов
алгоритмом распространения рисков

Algorithm 1: BackwardUpdate(x, t)

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

Algorithm 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: **end for**
- 5: **for** $t \in [t - d, t] : (x, u \notin U)$ **do**
- 6: **if** $\Delta p_t(u) > \epsilon$ **then**
- 7: **ForwardUpdate**(u, t', t)
- 8: **end if**
- 9: **end for**



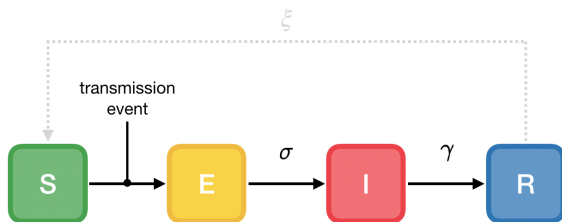
Эксперимент

Данные

- Граф контактов на предприятии (один цех) за 2 месяца

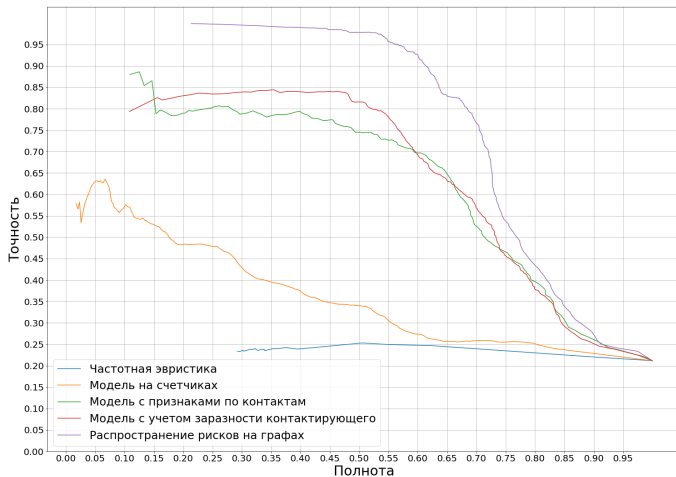
Информация о заражениях

- Генерируется моделью *SEIRS*: $\sigma = 0.64$, $\lambda = 0.14$,
 $\xi = 0.09$, $p(E | \text{contact}) = 0.9$



<https://github.com/ryansmcgee/seirsplus>

Сравнение качества моделей



Вакцинация на полусинтетических данных

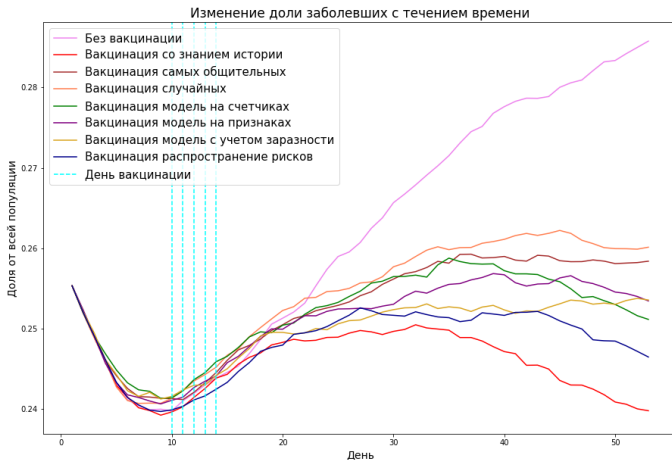
Схема вакцинации:

- Вакцинируем на протяжении 5 дней
- Каждый день по 2% выборки
- Индивидов для вакцинации выбираем из здоровых по определенной *стратегии*

Стратегии:

- Вакцинировать случайных
- Вакцинировать «самых общительных» (наибольшее суммарное количество контактов к дню вакцинации)
- Вакцинировать индивидов с наибольшим предсказанным моделью риском
- Вакцинировать «самых общительных» на всей истории наблюдений

Различные стратегии вакцинации



Вакцинация на синтетических данных

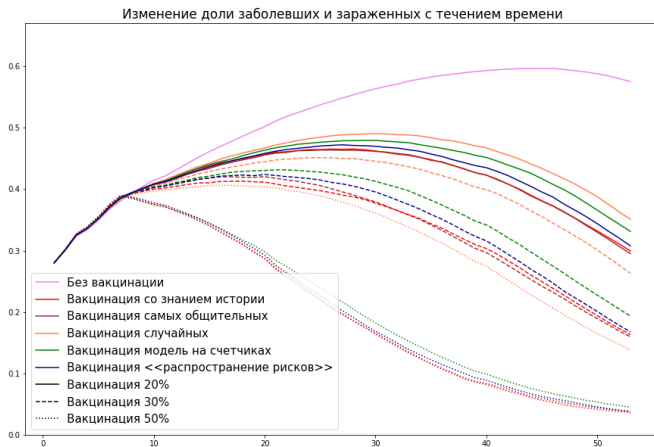
Генерация графов контактов

- Генерируем стартовый граф
- Для всех вершин стартового графа генерируем время контакта: выбираем временную группу контакта (короткий, средний, длинный), затем время из группы
- Чтобы получить граф контактов i го дня, случайно изменяем каждое ребро стартового графа: либо сдвигаем временную группу, либо удаляем ребро

Параметры имитационной модели

- Вероятность перехода $E \rightarrow I$, $\sigma = 0.3$
- Вероятность перехода $I \rightarrow R$, $\lambda = \frac{1}{16}$
- Вероятность перехода $R \rightarrow S$, $\xi = \frac{1}{180}$
- Вероятность перехода $S \rightarrow E$ зависит от времени контакта
- Переход от n к $n + 1$ происходит 3 раза в день

Стратегии вакцинации на синтетических данных



Результаты, выносимые на защиту

- Предложены модели оценки индивидуального риска во времени и алгоритм их обучения
- Показано, что итеративное усложнение моделей улучшает качество предсказания
- Исследована эффективность использования предсказаний моделей для вакцинации