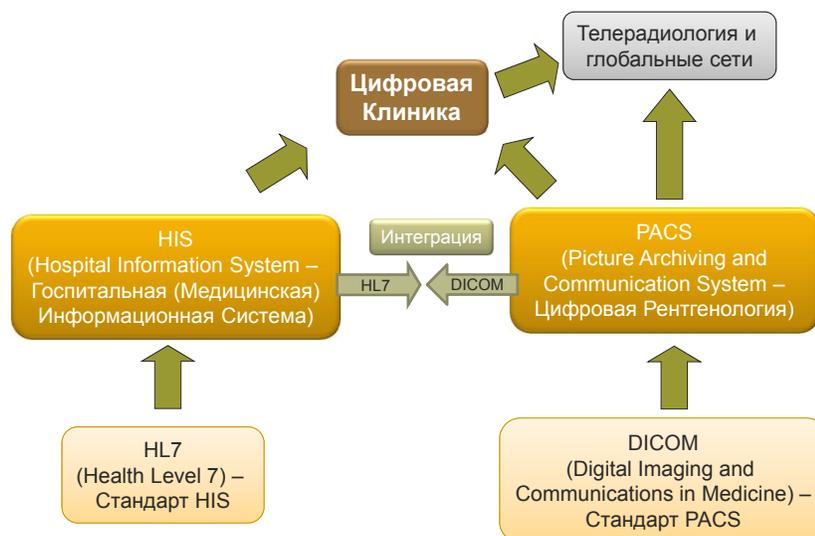


"Большие данные" в медицине: как заставить их работать?

Олег Пьяных
orpany@gmail.com

20.05.2014

Цифровая медицина



Олег Пьяных, orpany@gmail.com

Стандарт HL7

ADT (Admission, Discharge, Transfer) message

Pipe

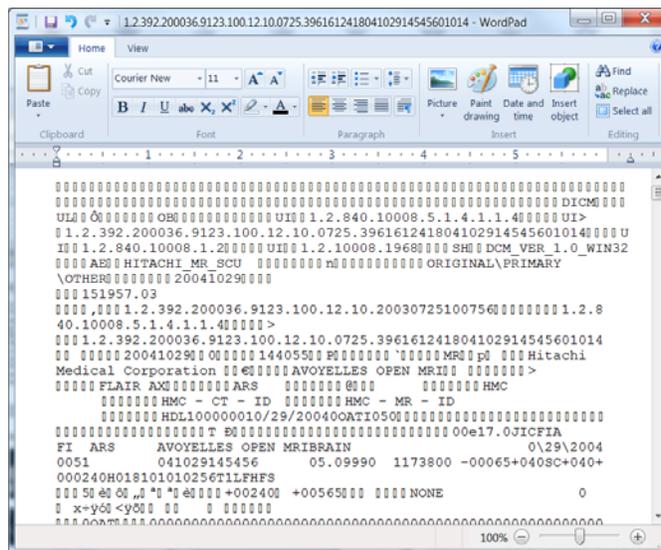
- MSH|^~\&|ADT|N|ADT|MEDSC|200601081527||ADT^A08|RE|P|3.2|||||ASCII|
- EVN|A08|200601080823|||||PID|1||3175875|1127278|S
AMPLE^JOE^|19901334|M||5400 Lake Villa
Dr^^Metairie^LA^70001-1230|(405)555-
2920||SINGLE|||||||||N|
- MSH|^~\&|ADT|N|ADT|MEDSC|200601081527||ADT^A08|RE|P|3.2|||||ASCII|EVN|A08|200601080832|||||PID|1||
3057088|1051999|INCOGNITO^MONICA^ANN^|197
80117|F||PO Box 1324^^Jefferson^LA^83625-
3184||(555)423-1423||OTHER||512-11-1425|||||||||N|

Найдите ошибку

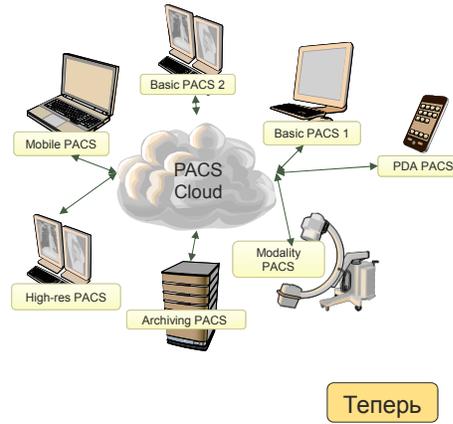
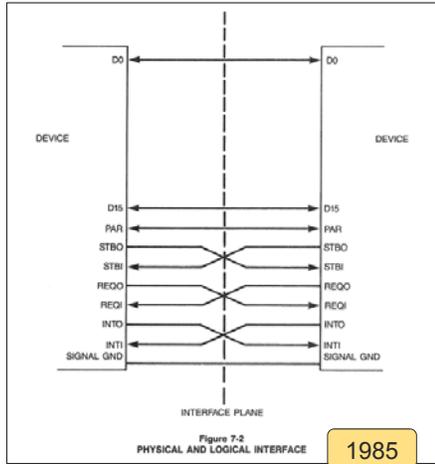
See http://www.hosinc.com/products/interfaces/interface_documentation.htm

Oleg Pianykh opiany@gmail.com

Стандарт DICOM

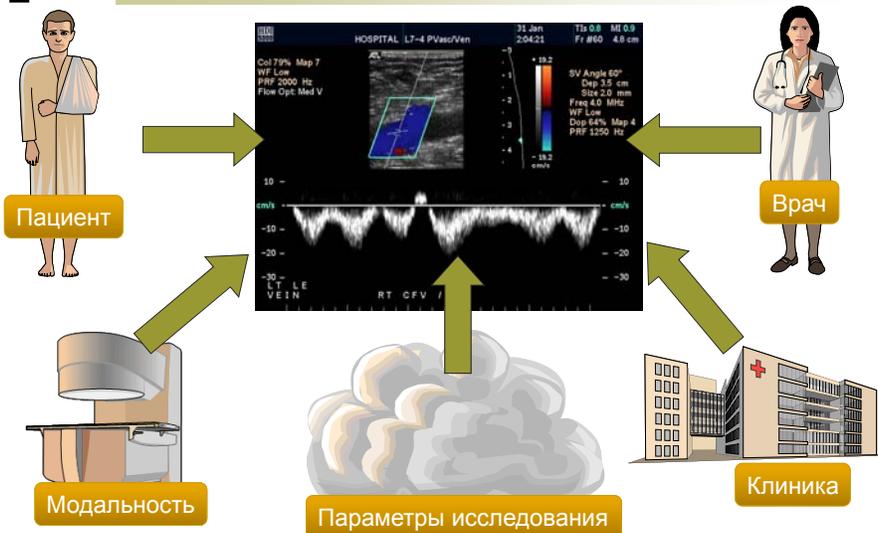


Эволюция цифровой медицины



Oleg Pinykh opiany@gmail.com

Сбор информации в медицине



Oleg Pinykh opiany@gmail.com

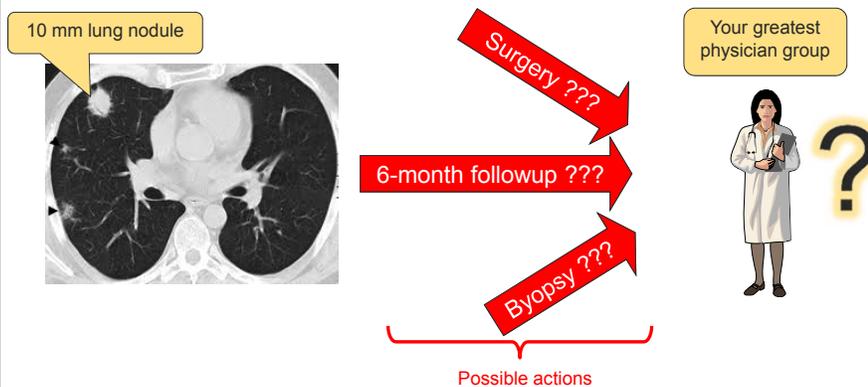
Информация собрана. Что дальше?

- Мы уже 20 лет как накапливаем информацию о всевозможных параметрах и процессах в здравоохранении. Не пора ли начать её использовать?
 - Показ отклонений от нормы
 - Анализ данных
 - Стандартизация здравоохранения
 - Оптимизация медицины
- Все это требует эффективных математических решений

Олег Пьяных, opiary@gmail.com

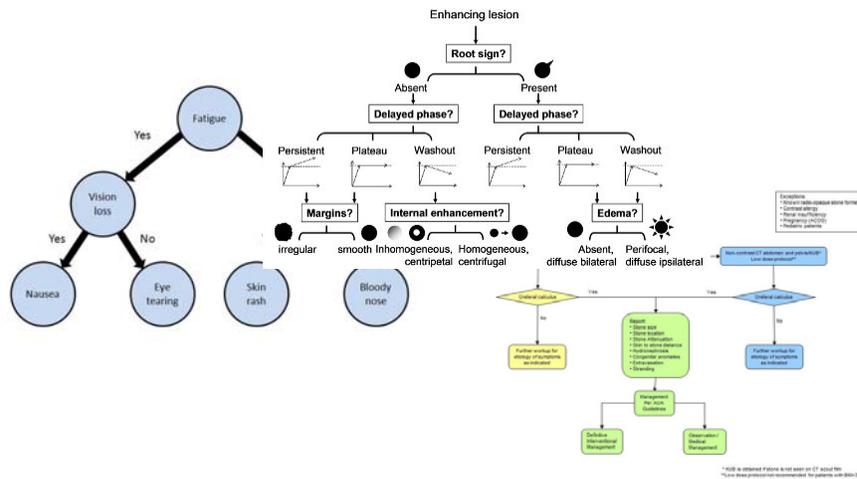
Очевидный пример: принятие решений в медицине

- The same clinical observations should lead to the same standard actions, but physicians' opinions may differ:

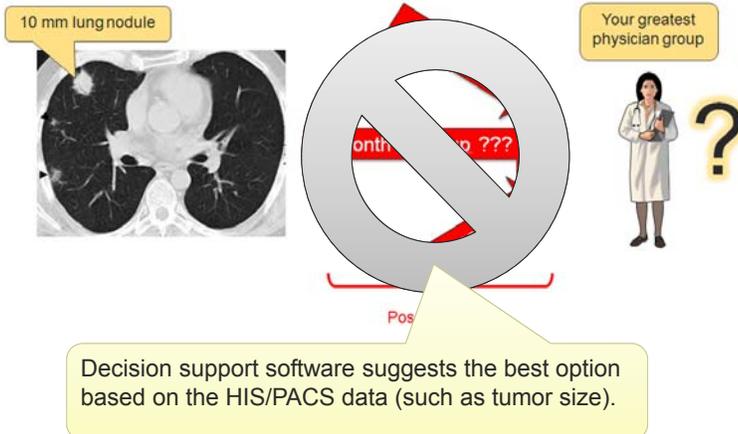


Олег Пьяных opiary@gmail.com

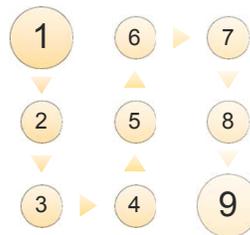
От «больших данных» к алгоритмам решений



Результат: алгоритмы принятия решений в медицине



Банальный пример: Обслуживание пациентов в клинике



Oleg Pinykh opiany@gmail.com

Зачем это нужно?

- Балансировка нагрузки
- Устранение проблем обслуживания, оптимизация работы
- Удовлетворенность пациентов и персонала. Качество обслуживания
- Отказ от бесплодных проектов («а давайте примем в два раза больше!»)
- Элементарное понимание того, что происходит

Олег Пьяных, opiany@gmail.com

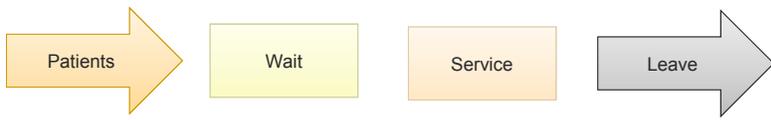
Классика жанра: Queuing theory (Теория массового обслуживания)



- Начало: A.K. Erlang, “The theory of probabilities and telephone conversations,” 1909
- Расчет параметров различных моделей очередей
- Универсальность: применима к любым очередям, будь то расчет задач на процессоре или очередь в кассу магазина

Oleg Pinykh opianykh@mgh.harvard.edu

QT в медицине



Demand
(λ patients per minute)

Queue

Capacity
(μ patients per minute,
 $\rho = \lambda / \mu$ – utilization,
 $\rho < 1$ – stable system)

The formula states that the mean queue length L is given by

$$L = \rho + \frac{\rho^2 + \lambda^2 \text{Var}(S)}{2(1 - \rho)}$$

where

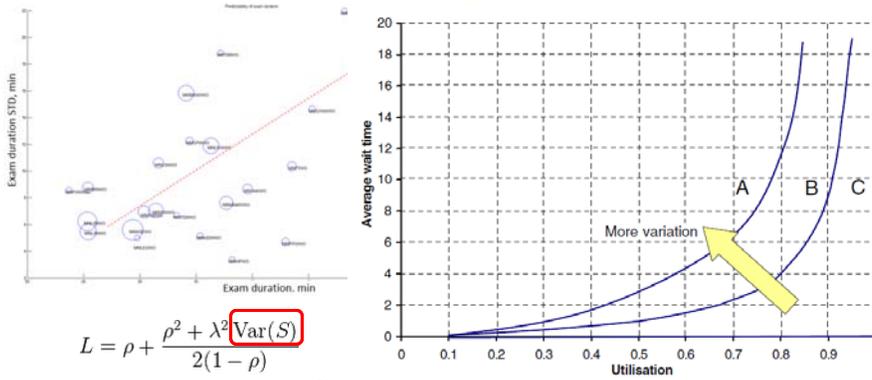
- λ is the arrival rate of the Poisson process
- $1/\mu$ is the mean of the service time distribution S
- $\rho = \lambda/\mu$ is the utilization
- $\text{Var}(S)$ is the variance of the service time distribution S .

See “Queuing for Healthcare”, R. Kannapiran Palvannan & Kiok Liang Teow

Oleg Pinykh opianykh@mgh.harvard.edu

Уроки QT

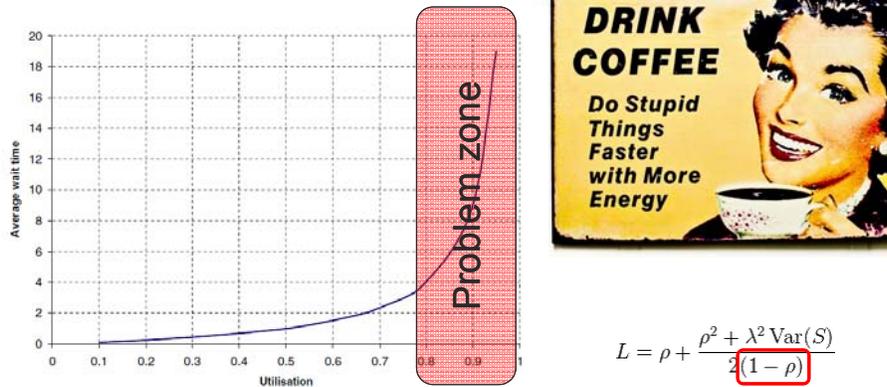
- L1: Queues are formed even when service rate *exceeds* demand. Variation in demand and service rate affects patient wait time significantly



Oleg P. pianykh@mgm.harvard.edu

Уроки QT

- L2: Short wait time requires low system utilization and high cost structure.

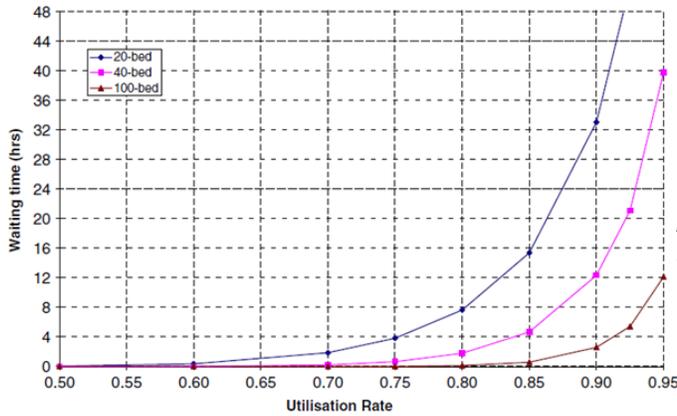


Oleg P. pianykh@mgm.harvard.edu

Уроки QT

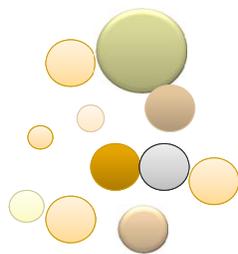
- L3: Resource availability and allocation strategies matter:

Can you always allocate more, really?



Реальность: всё гораздо сложнее

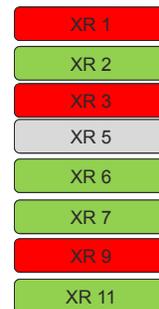
- Простейшая задача:



Пациенты



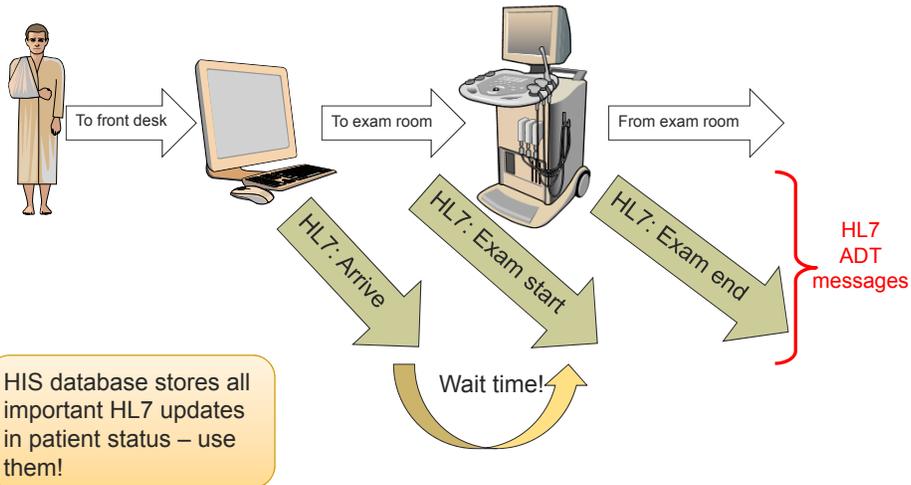
Ассистенты



Рентген-станции

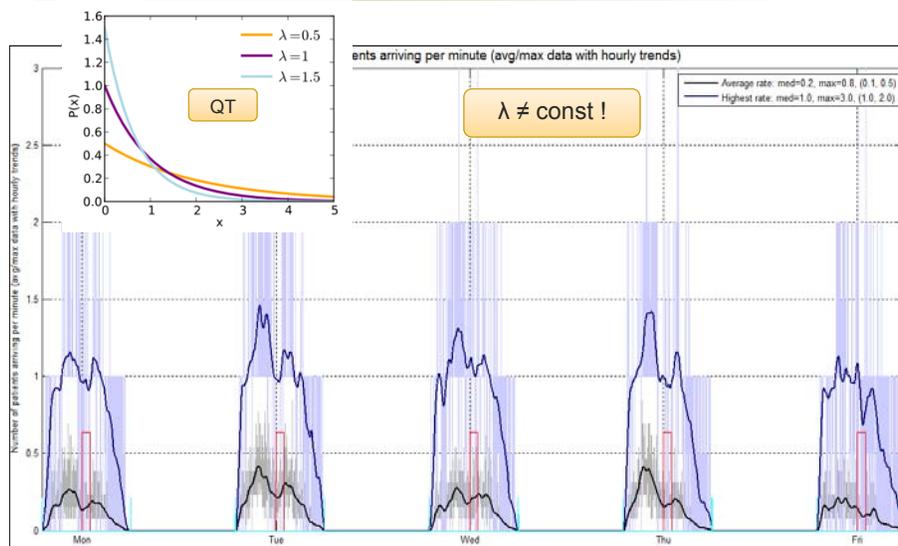
- Можем ли мы контролировать время ожидания пациента?

Источник «больших данных»: ГИС



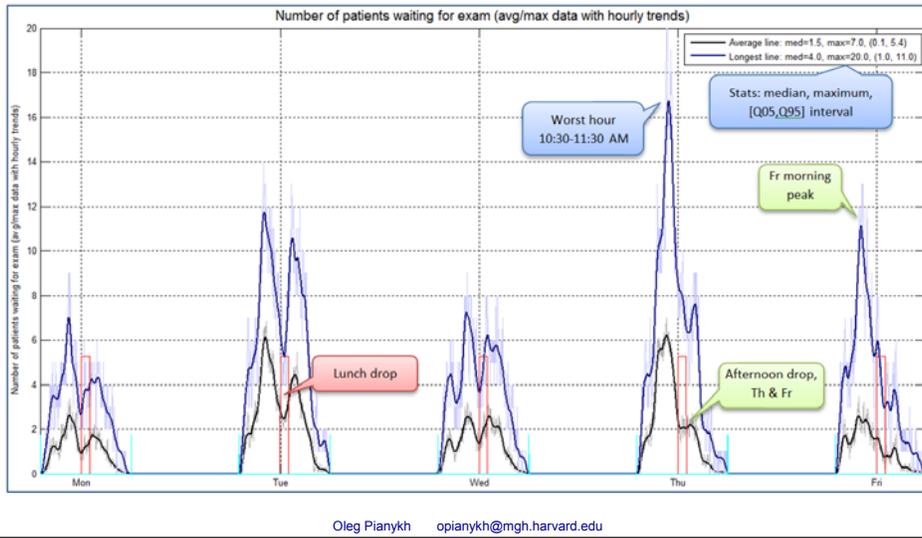
Oleg Pinykh opiany@gmail.com

Реальность: Patient arrival rate λ

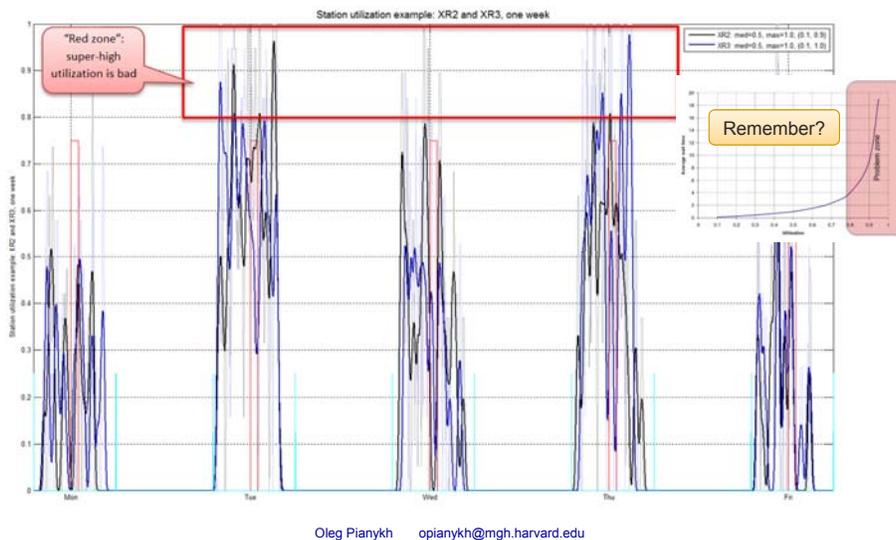


Oleg Pinykh opianykh@mgh.harvard.edu

Реальность: Patient waiting time W



Реальность: Station utilization



Итак, на самом деле

- “Big Data” значительно отклоняется от классических распределений и моделей. Например, в нашем случае корреляция QT с наблюдаемым временем оказалась и вовсе отрицательной
- Применение моделей сильно ограничено тем, что мы знаем на данный момент, и достоверностью этого знания.
- Значительное присутствие непредсказуемых и случайных факторов (особенно в медицине, где многое меняется «на лету»)

Олег Пьяных, oriany@gmail.com

Неутешительный вывод

- Классические модели обслуживания малоприменимы в сложных реальностях клиники
- Приходится *начинать* с простых, но надежных эмпирических моделей – например, линейной регрессии. Они строятся на основе накопленных «больших данных», и ложатся в основу практических алгоритмов обслуживания.

Олег Пьяных, oriany@gmail.com

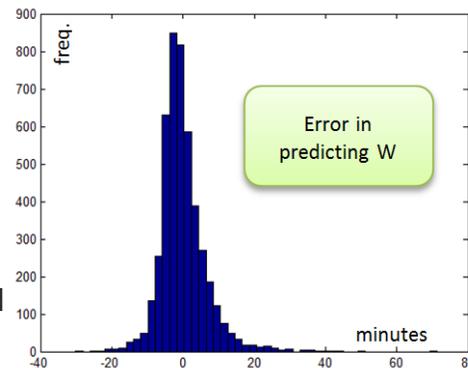
Как же предсказать время ожидания пациента W ?

- QT formula $W=L/\lambda$ – based on idealized model

- Empirical formula, derived with best-predictor regression analysis from 3 months of real clinic data:

$$W = 4.5 + L + 0.5L_5 + 0.25L_{10},$$

where L_5 is the size of the waiting line 5 minutes ago, and L_{10} – 10 minutes ago.



Oleg Pinykh opianykh@mgh.harvard.edu

Проблемы с простыми моделями

- Статичны: не подстраиваются под изменяющуюся реальность
- Ограничены по точности
- Работают лишь в простых процессах

Олег Пьяных, opiany@gmail.com

Можем ли мы улучшить нашу модель?

- Регрессия – ограничена
- Нелинейная регрессия – не получилось!
- Искусственный интеллект (нейронные сети) – не получилось!
- Почему?

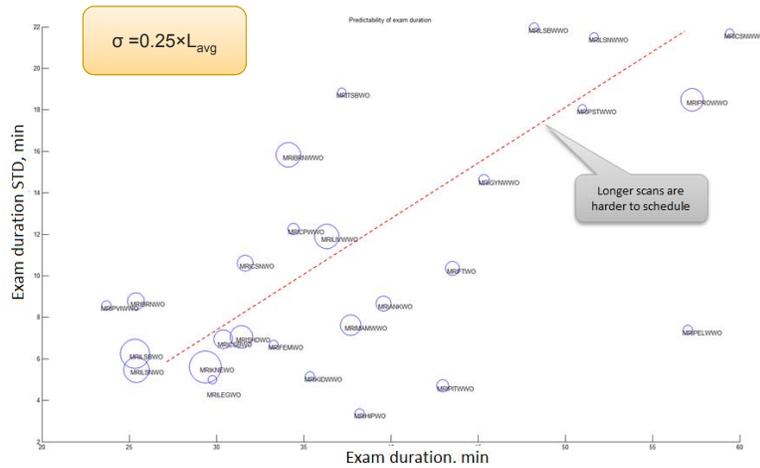
Олег Пьяных, oriany@gmail.com

Пределы применения Big Data

- Никаких иллюзий: «большие данные» не гарантируют «большой смысл»! В них отсутствует структура, корректность, достоверность. *«Большие данные» - это просто всё то, что у нас есть, и с чем надо как-то работать.*
- Большие данные могут содержать большие горы мусора:
 - Случайные, непредсказуемые события
 - Неверные данные (ручной ввод!!!)
 - Данные не в реальном времени (естественная задержка в получении данных, или данные post factum)
 - Беспольные или неполные данные

Олег Пьяных, oriany@gmail.com

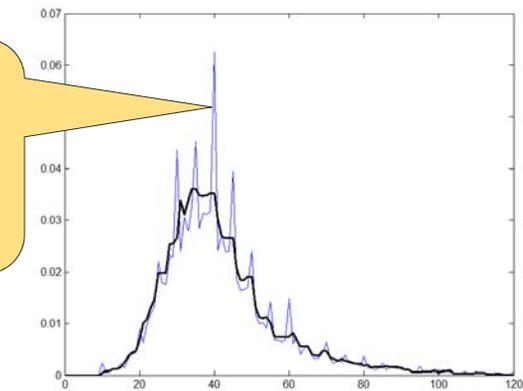
Накопление случайных событий



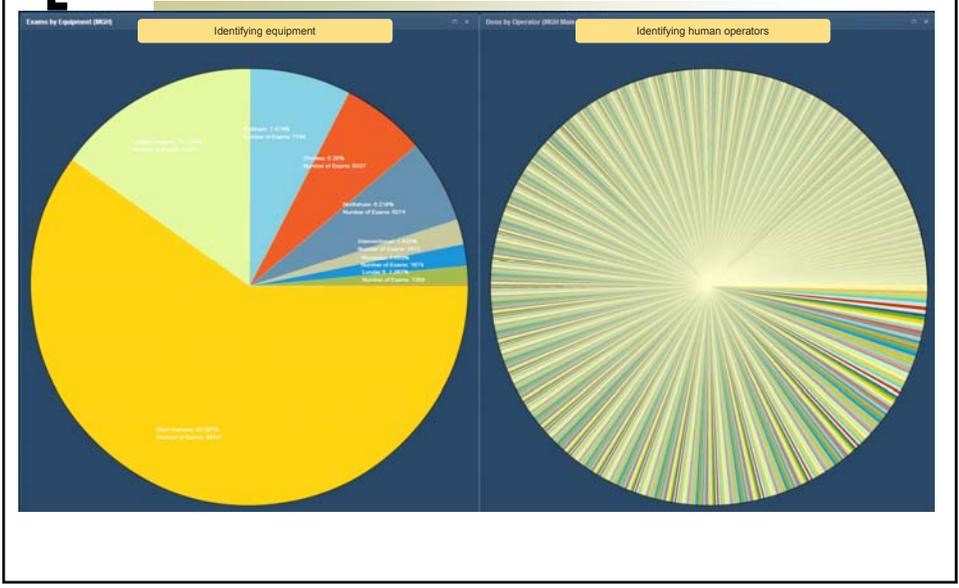
Неверные данные

- Manual entry supplies most of HIS data!
- Result: incorrect or strangely-biased numerical results

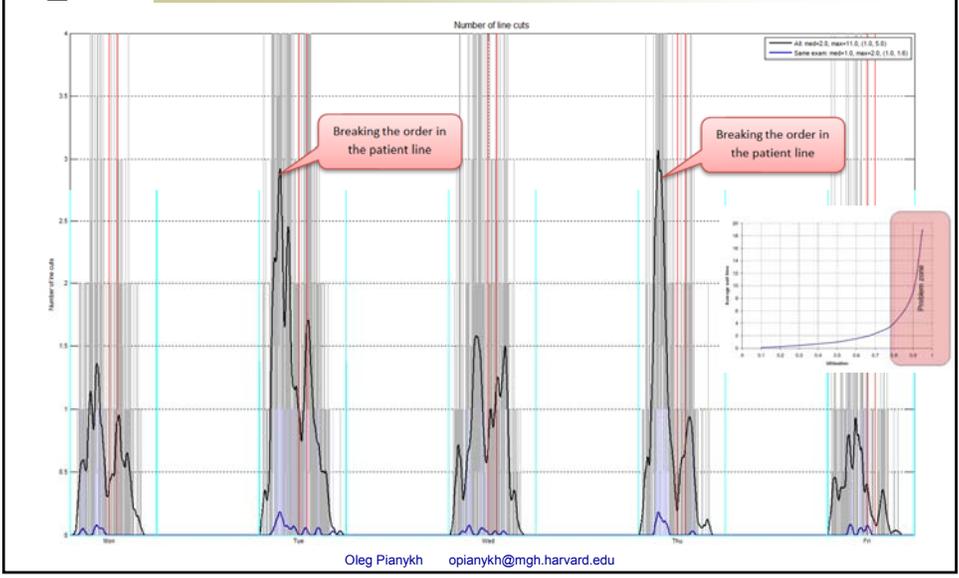
Example: Histogram for MRI exam length distribution (blue) shows unusual spikes for multiple of 5 minutes: data entry with 5-min accuracy. They have to be filtered out (black).



Неверные данные



Хаос в данных (и в работе)



Oleg Pianykh opianykh@mgh.harvard.edu

[Что делать?]

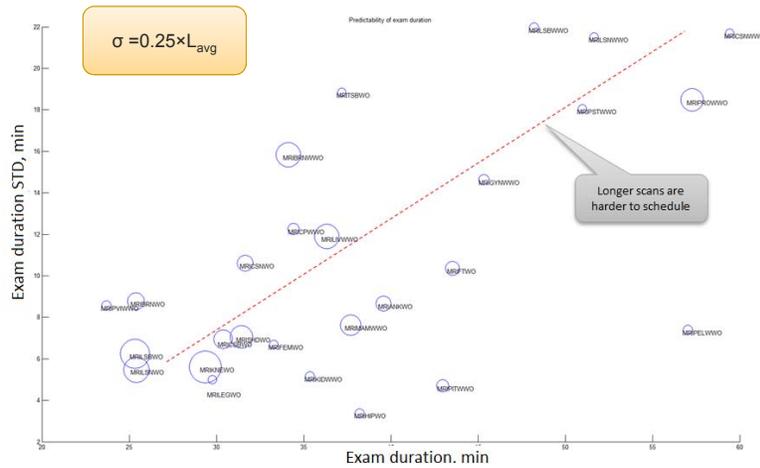
- Случайность, хаос, несвоевременность являются основными врагами отлаженного рабочего процесса
- Логика работы с большими данными должна включать в себя устранение этих аномалий, и превращение любого процесса в наиболее детерминированный
- → Анализируя «большие данные», необходимо:
 - Понимать их
 - Развивать не только логику, но и оптимальные стратегии работы

Олег Пьяных, oriany@gmail.com

[Пример: понижение случайности]

Олег Пьяных, oriany@gmail.com

Накопление случайных событий



Oleg Pinykh opianykh@mgh.harvard.edu

Как ускорить обслуживание пациентов ?

- Стандартный подход: уменьшить время обслуживания



- Правильный подход: изменить стратегию обслуживания для понижения случайности

Олег Пьяных, opianykh@gmail.com

Выбор стратегии обслуживания

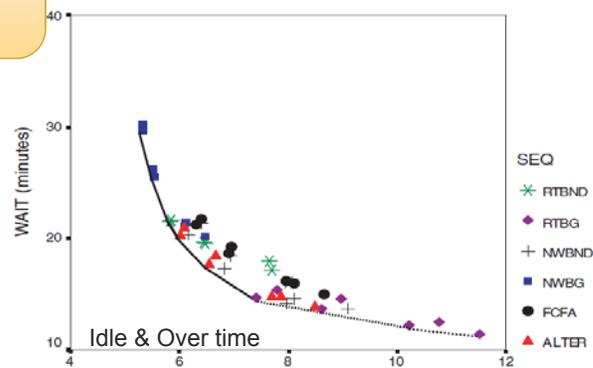
- Основные правила:
 - Никогда не начинать дня с наиболее длинного исследования
 - Чередовать длинные и короткие
 - Назначать исследования пропорционально их реальной длине (средней для данного типа исследования)

Олег Пьяных, opiany@gmail.com

Выбор стратегии обслуживания

Sequencing rules:

FCFS
ALTER (long/short)
Long first (NWBG)
Short first (RTBG)
Long first and last (NWBND)
Short first and last (RTBND)



"Designing appointment scheduling systems for ambulatory care services", Tugba Cayirli · Emre Veral · Harry Rosen

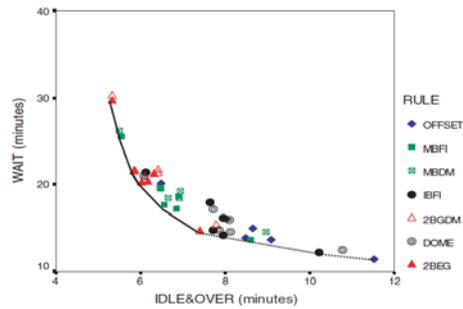
Oleg Pinykh opianykh@mgh.harvard.edu

Выбор стратегии обслуживания

Appointment rules:

Table 1 Appointment rules

Symbol	Description	Formulation ^a
IBFI	Individual block/first-interval rule: calls patients individually at intervals equal to the mean service times of patients	$n = 0$ $t_i = t_{i-1} + \mu$ for $i > 1$
OFFSET	Individual block/variable interval rule, where initial $(k_1 - 1)$ patients are scheduled earlier, and the rest are scheduled later compared to IBFI	$t_i = (i - 1)\mu - \beta_1(k_1 - i)\mu$ $t_i = (i - 1)\mu + \beta_2(i - k_2)\mu$ for $i \leq k_1$, and for $i > k_1$
DOME	Individual block/variable interval rule, where initial $(k_1 - 1)$ patients are scheduled earlier, patients $(k_1 + 1)$ through $(k_2 - 1)$ are scheduled later, and the rest earlier compared to IBFI	$t_i = (i - 1)\mu - \beta_1(k_1 - i)\mu$ $t_i = (i - 1)\mu + \beta_2(i - k_2)\mu$ for $i \leq k_1$, for $k_1 < i < k_2$, and for $i \geq k_2$
2BEG	Individual block/first-interval rule with an initial block of two patients	$n = n_1 = 0$ $t_i = t_{i-1} + \mu$ for $i > 2$
MBFI	Multiple block/first-interval rule: calls patients two-at-a-time with intervals set equal to twice the mean service time	$t_i = t_{i-1} + 2\mu$ for $i = 1, 3, 5$
2BGDM	Combination of the 2BEG and the DOME rules	$t_i = t_{i-1} + 2\mu - \beta_1(k_1 - i)\mu$ $t_i = t_{i-1} + 2\mu + \beta_2(i - k_2)\mu$ for $i = 1$, for $2 \leq i \leq k_1$, for $k_1 < i < k_2$, and for $i \geq k_2$
MBDM	Combination of the MBFI and the DOME rules	$t_i = t_{i-1} + 2\mu - \beta_1(k_1 - i)\mu$ $t_i = t_{i-1} + 2\mu + \beta_2(i - k_2)\mu$ for $i = 1, 3, \dots$, for $k_1 < i < k_2$, and for $i \geq k_2$; all k_2 are odd

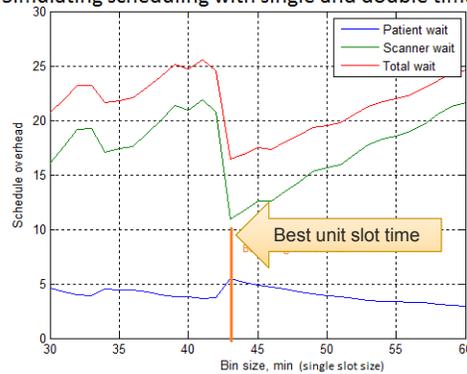


Copyright © 2010 Massachusetts Institute of Technology. All rights reserved. <http://www.mit.edu/~anilkumar/papers/2010-01-01-appointment-rules.pdf>

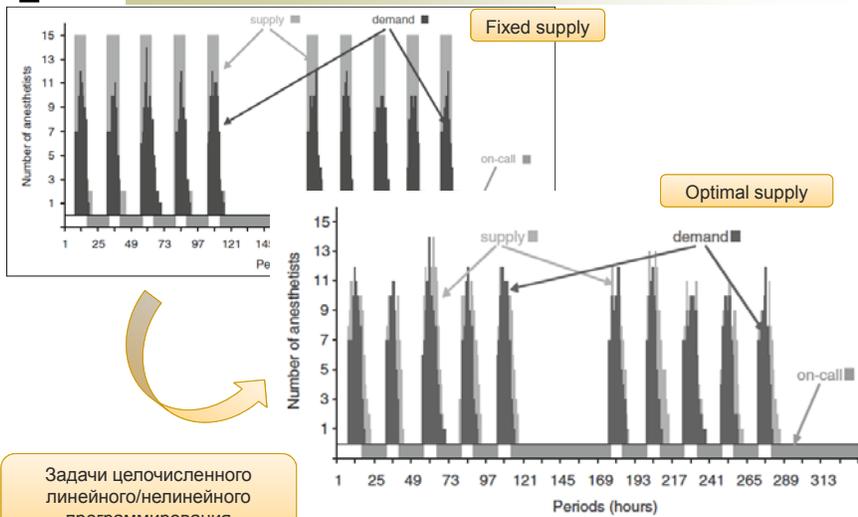
Оптимизация расписания исследований

- Пример: найти оптимальный интервал времени T так что все исследования можно назначать либо на T , либо на $2T$ времени

Simulating scheduling with single and double time slots:



Оптимизация расписания ресурсов



Задачи целочисленного линейного/нелинейного программирования

See: Brunner JO, Bard JF, Kolisch R., "Flexible shift scheduling of physicians", Health Care Manag Sci. 2009 Sep;12(3):285-305.
 Oleg Pinykh opiany@gmail.com

Оптимизация расписания ресурсов

Model

Parameters

- d_i^t demand in period t $d_i^t = a_i - (1 - \alpha_{i,t})$ $\forall i, t \in T$ (2)
- c_i^t cost per hour of paid-out time for physician i $c_i^t = a_{i,t} - (1 - \alpha_{i,t})$ $\forall i, t \in T$ (3)
- c_o^t cost per hour for outside physicians $c_o^t = \sum_{i \in I} a_{i,t} - (1 - \alpha_{i,t})$ $\forall t \in T$ (4)
- O_i^{max} maximal allowed overtime for a physician i in a week $\sum_{t \in T} a_{i,t} \leq \min(\sum_{t \in T} (1 - \alpha_{i,t}) c_i^t, O_i^{max})$ $\forall i \in I$ (5)
- L_i^{min} minimum shift length $\forall i \in I, t \in T$ (6)
- L_i^{max} maximum shift length after a shift ends $\forall i \in I, t \in T$ (7)
- W_i regular working hours per week for physician i according to his individual contract $\sum_{t \in T} (1 - \alpha_{i,t}) c_i^t \leq W_i$ (8)
- F^{max} number of periods without a week $F^{max} = \lfloor T / W \rfloor$ (9)
- F^{min} number of periods in a day $F^{min} = \lceil T / W \rceil$ (10)
- l length of an on-call service $\forall i \in I, t \in T$ (11)
- N maximal number of on-call services for a physician in a week $\sum_{t \in T} a_{i,t} + 4l \leq N$ $\forall i \in I$ (12)

Function

- $f(O)$ indicate the number of hours that are charged for any on-call service for each $i \in I$, so regular working time per week $\sum_{t \in T} a_{i,t} + \sum_{t \in T} (1 - \alpha_{i,t}) c_i^t \leq W_i + n_{i,t} - n_{i,t-1}$ $\forall i \in I, t \in T$ (13)

Binary decision variables

- $\alpha_{i,t}$ 1, if physician i works in period t , 0 otherwise $\alpha_{i,t} \in \{0, 1\}$ $\forall i \in I, t \in T$ (14)
- $\beta_{i,t}$ 1, if physician i begins a shift in period t , 0 otherwise $\beta_{i,t} \in \{0, 1\}$ $\forall i \in I, t \in T$ (15)
- $\gamma_{i,t}$ 1, if rest period begins for physician i in period t , 0 otherwise $\gamma_{i,t} \in \{0, 1\}$ $\forall i \in I, t \in T$ (16)
- $\delta_{i,t}$ 1, if physician i begins an on-call service in period t , 0 otherwise $\delta_{i,t} \in \{0, 1\}$ $\forall i \in I, t \in T$ (17)

General integer decision variables

- $n_{i,t}$ amount of overtime for physician i in week w $n_{i,t} \geq \sum_{t \in T} a_{i,t} - \sum_{t \in T} (1 - \alpha_{i,t}) c_i^t$ $\forall i \in I, t \in T$ (18)
- $n_{i,t}$ amount of paid-out time for physician i in week w $n_{i,t} \geq \sum_{t \in T} a_{i,t} - \sum_{t \in T} (1 - \alpha_{i,t}) c_i^t$ $\forall i \in I, t \in T$ (19)
- $n_{i,t}$ number of outside physicians hours in period t $n_{i,t} \geq \sum_{t \in T} a_{i,t} - \sum_{t \in T} (1 - \alpha_{i,t}) c_i^t$ $\forall i \in I, t \in T$ (20)

Model

- Minimize $\sum_{i \in I} \sum_{t \in T} c_i^t n_{i,t} + \sum_{t \in T} c_o^t n_{o,t}$ (21)
- $n_{i,t} + \gamma_{i,t} \leq 1$ $\forall i \in I, t \in T$ (22)

Optimal roster

See: Brunner JO, Bard JF, Kolisch R., "Flexible shift scheduling of physicians", Health Care Manag Sci. 2009 Sep;12(3):285-305.
 Oleg Pinykh opiany@gmail.com

ОК, алгоритм готов. Что дальше?

Олег Пьяных, oriany@gmail.com

Разработка приложений

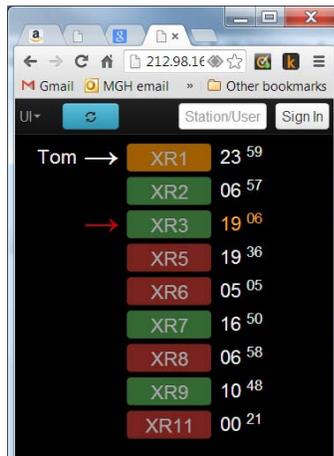
- Прекрасно: мы проанализировали «большие данные» и создали улучшенный алгоритм работы

НО

- Любой алгоритм необходимо внедрить:
 - Доказать преимущества решения на практике
 - Создать надежное, устойчивое решение
 - Создать удобное решение
 - Отработать все возможные случаи, когда решение не срабатывает

Олег Пьяных, oriany@gmail.com

Пример: распределение пациентов по рентгенам



The screenshot shows a web browser window with a URL of 212.98.16. The page displays a list of X-ray stations (XR1 to XR11) with their respective times. A patient named Tom is assigned to XR1. The interface includes a 'Station/User' dropdown and a 'Sign In' button.

Station	Time
XR1	23:59
XR2	06:57
XR3	19:06
XR5	19:36
XR6	05:05
XR7	16:50
XR8	06:58
XR9	10:48
XR11	00:21

- Web application to view and book X-ray stations”
 - Faster patient dispatching
 - Fairness in exam-to-tech assignments
 - Optimal station utilization

→ Reduction of variability
→ Reduction of over-utilization and stress

Oleg Pinykh opianykh@mgh.harvard.edu

Всё, проблема решена?

- Нет!
 - Интеграция с госпитальными системами (источники данных)
 - Надежность в работе (что делать, если села батарейка, исчез WiFi, ввели не тот пароль, ...?)
 - Масштабируемость; переносимость на другие процессы
 - Обучения персонала
 - Работа с пациентами

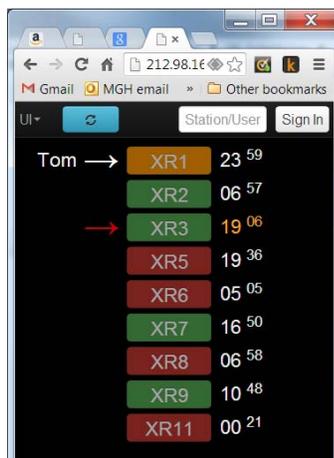
Олег Пьяных, opiany@gmail.com

Человеческий фактор

- Любое массовое внедрение алгоритмов неизбежно превращается в «социальную игру»: «Обмани машину»
 - Люди пытаются обмануть алгоритм
 - Люди пытаются обмануть друг друга
 - Люди далеко не просто «интегрируются» с компьютерами и гаджетами!
- Ваше приложение должно включать пользователей-антогонистов

Олег Пьяных, opiany@gmail.com

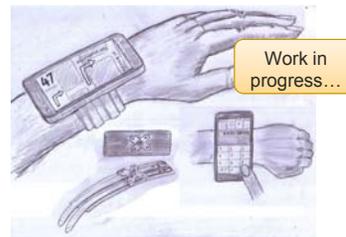
Пример: распределение пациентов по рентгенам



Tom →	XR1	23 ⁵⁹
	XR2	06 ⁵⁷
→	XR3	19 ⁰⁶
	XR5	19 ³⁶
	XR6	05 ⁰⁵
	XR7	16 ⁵⁰
	XR8	06 ⁵⁸
	XR9	10 ⁴⁸
	XR11	00 ²¹

Олег Пьяных opianykh@mgh.harvard.edu

Пример: «интеграция» с человеком



http://www.kickstarter.com/projects/1958024164/connector-smartphone-iphone-gps-wearable-mounting?ref=recently_launched

Oleg Pinykh opianykh@mgh.harvard.edu

Заключение

- Да, «большие данные» в медицине абсолютно необходимо анализировать и использовать для оптимизации работы любого процесса.
- Но:
 - Начинать надо с задачи, а не решения
 - Надо вкладываться в понимание накопленной информации еще до того, как создавать какие-либо решения на ее основе
 - В медицине наиболее важно законченное, до конца внедренное решение. Это налагает жесткие требования на качество алгоритма.
 - «Не навреди» → «Сделай значительно лучше»

Олег Пьяных, opiany@gmail.com