ENBISYS
ENTERPRISE BUSINESS SYSTEMS

Deep Learning & Ритейл

Максим Усов

План

- Big Data сегодня
- примеры в ритейл & тренды
- Machine learning & deep learning
- необычный пример от нас

Big Data сегодня

Michael Stephen Palmer, M.D.
(October 9, 1942 – October 30, 2013),
was an American physician and novelist.

«Data is just like crude. It's valuable, but if unrefined it cannot really be used. It has to be changed into gas, plastic, chemicals, etc., to create a valuable entity that drives profitable activity; so must data be broken down, analysed for it to have value.»

- Michael Stephen Palmer

Big Data сегодня

- 3V: Volume + Variety + Velocity
- от моделирования к обнаружению закономерностей
- сдвиг в область Machine Learning

Проблема: никто не хочет делиться данными

Где используется?

- финансы (махинации, скоринг, биржи, …)
- реклама (сегментация, кастомизация)
- здравоохранение (диагнозы, план лечения, прогнозы)

- ритейл (online + offline)
- безопасность (видеонаблюдение, сетевая, …)
- соцсети (куда пойти, с кем подружиться)
- выборы в США 2016 (Donald Trump)

Ритейл

- GEO-targeting (где открыть магазин)
- предсказание спроса
- предсказание ухода клиента
- персональные предложения
- распознавание паттернов поведения (на сайте)
- …

На каждом шаге воронки продаж

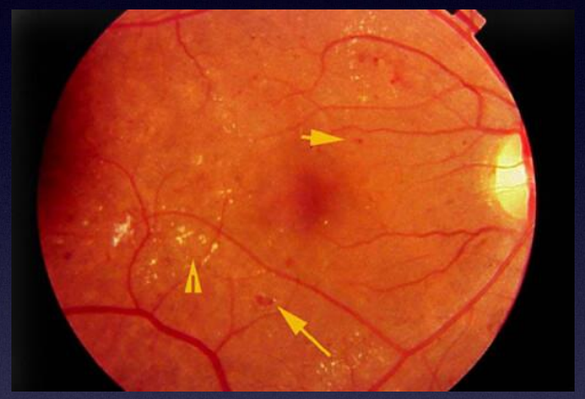
- реклама
- умные Landing Page (знает, кто смотрит)
- умный chat-bot - советует
- …

Тренды в offline ритейл

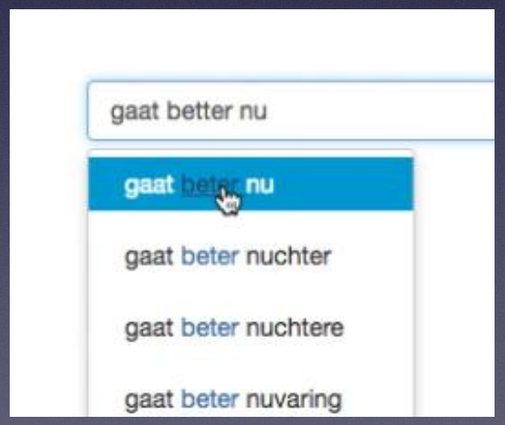
- тепловая карта
- распознавание кражи
- распознавание лица*
  - поиск профиля в соц.сети (findface.ru)
  - пол + возраст (86%)
- наличие товара на полке (фрукты кончаются)
- экзотика: магазин без продавцов «Amazon GO»

* https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/

Здравоохранение



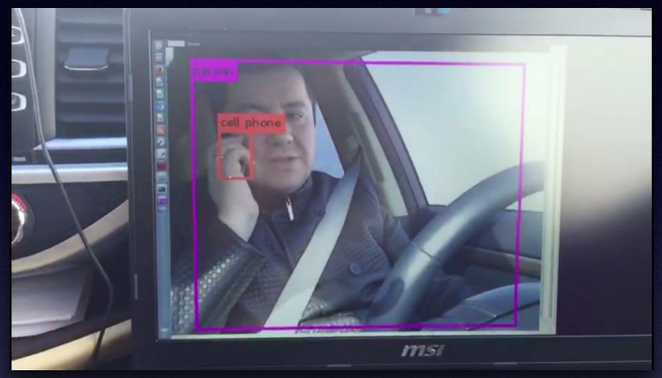
- обнаружение диабетической ретинопатии
- помощник ввода текста
- конвертирование «Субъективное» в «Объективное»

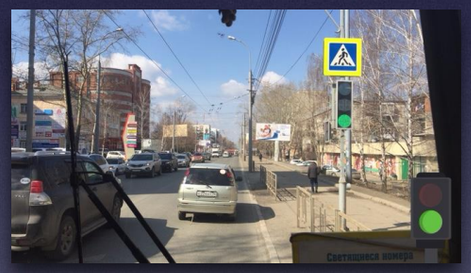


Со вчера у меня болит голова и я какой-то разбитый. За ночь бегал в туалет несколько раз

“Мигрень*[G43]* на фоне диареи*[A09]*. Общая слабость*[R53]*.”

Автомобильная безопасность

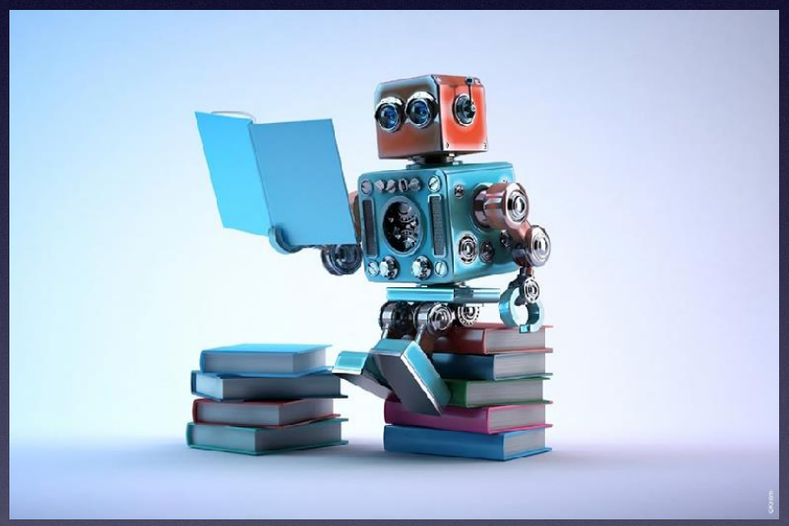




Контроль водителя такси

- пользование сотовым телефоном
- курение
- проезд на красный свет
- пересечение сплошной
- не пропустил пешехода

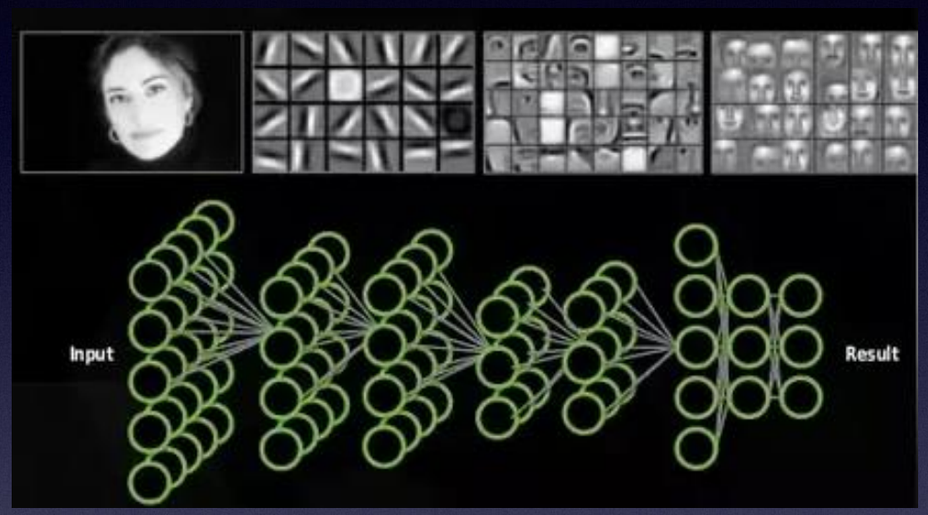
Machine Learning



- байесовы методы
- генетические алгоритмы
- reinforcement learning
- gradient boosting (XGBoost)
- deep learning

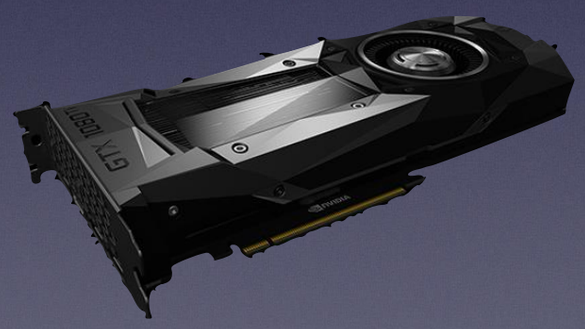
…

Deep learning



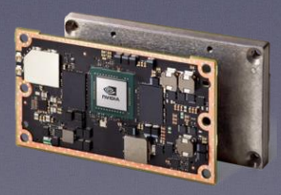
Моделирование высокоуровневых абстракций

- Tools: Tensorflow, Caffe, Keras, nVidia DIGITS …
- Hardware: nVidia CUDA



Обучение:

- Titan X (90 000 руб)
- GTX 1080 ti (57 000 руб)



Runtime:

- nVidia Jetson TX2 (27 000 руб)

Deep learning

```
52
53          self.model.add(LSTM(200, input_shape=(3, self.dictionary_size), return_sequences=True, implementation=2))
54          self.model.add(LSTM(200, return_sequences=False, implementation=2))
55          self.model.add(RepeatVector(1))
56          self.model.add(TimeDistributed(Dense(self.dictionary_size-1)))
57          self.model.add(Activation('softmax'))
58          optimizer = RMSprop(lr=0.01)
59
60          self.model.compile(
61              loss='categorical_crossentropy',
62              #loss='mse',
63              optimizer=optimizer,
64              metrics=['accuracy']
65          )
```

Необычный пример

Mix & Match

“Купи по одной бутылке вина из категории (German+French+USA) и получи самую дешевую из них - бесплатно!”

Задача: рассчитать максимальную скидку (POS!)

Проблема: в супермаркетах в чеке может быть свыше 100 позиций x 10 акций

NP-полная задача (задача о ранце)
>> 1 000 000 комбинаций за 0.5 сек!

Матрица покрытия

	A1	A2	A3	A4	A5
G1	1	0	1	0	1
G2	0	1	1	0	0
G3	0	0	0	1	0

Нет градиента - порядок не важен

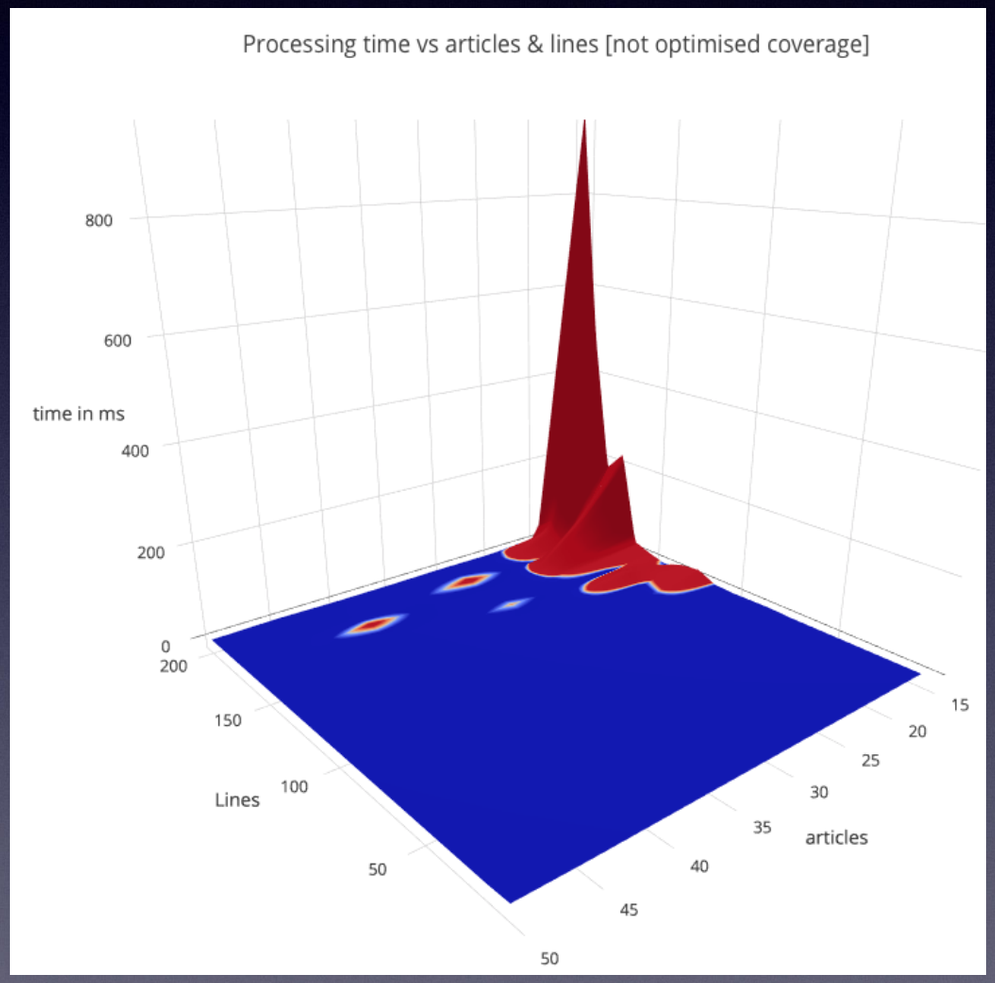
Brute force

Java 8

Core i7-4710HQ

(2.5/3.5 GHz)

16GB RAM



Average time

Max time x 100

Sometimes crashes

on OOM

(>8GB heap)

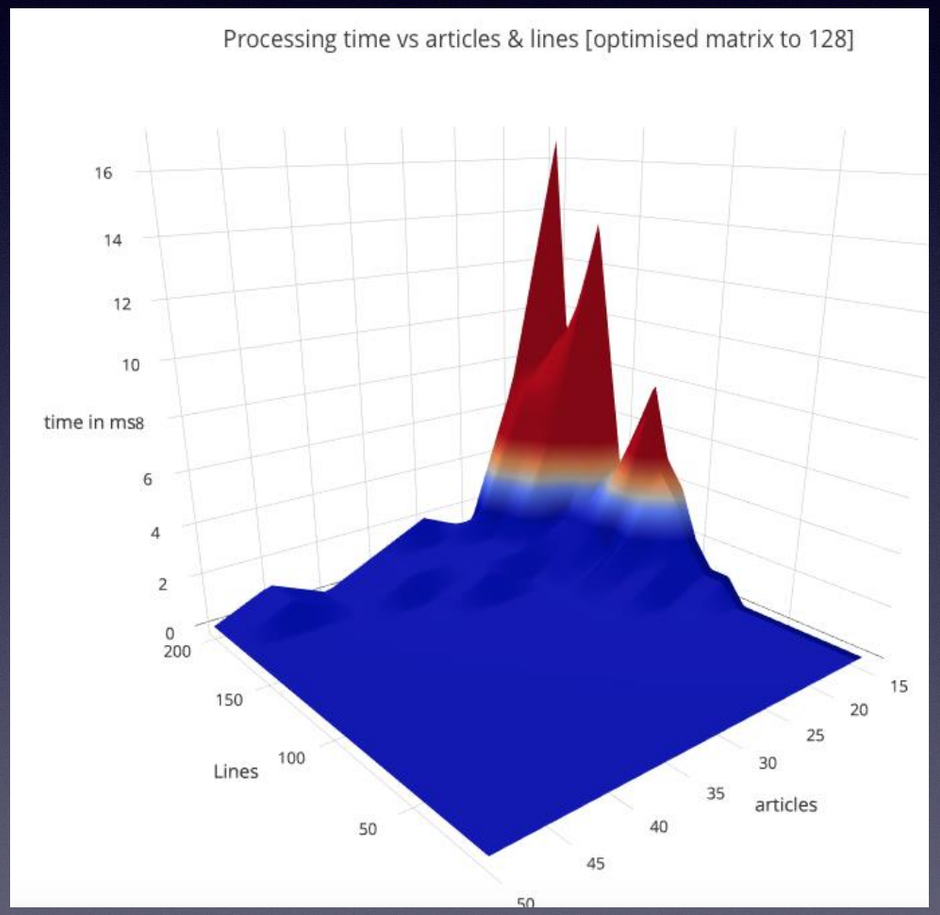
Стохастический подход

Java 8

Core i7-4710HQ

(2.5/3.5 GHz)

16GB RAM



Average time

Max time x 100

No crashes

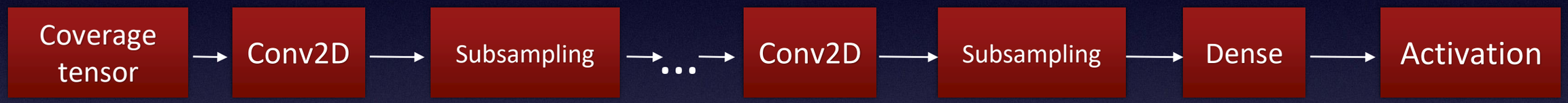
(~200 MB heap)

Accuracy: ~0.75

Вызов принят

- считаем, что в матрице есть скрытые закономерности
- пусть нейросеть будет предсказывать комбинацию
- уменьшить количество переборов

Model

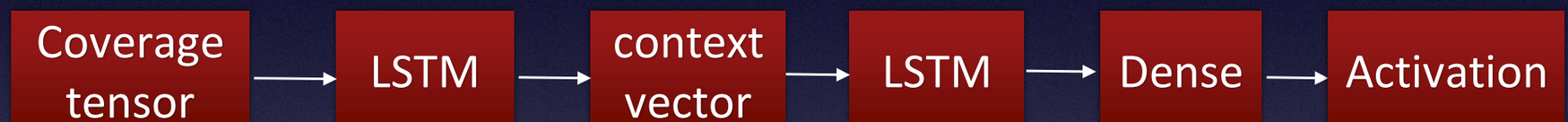


$$\begin{vmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{vmatrix}$$

$$\begin{vmatrix} 0.10 & 0.04 & -0.00 & -0.10 \\ 0.28 & -0.40 & 0.02 & -0.07 \\ -0.07 & 0.02 & 0.08 & 0.47 \end{vmatrix}$$

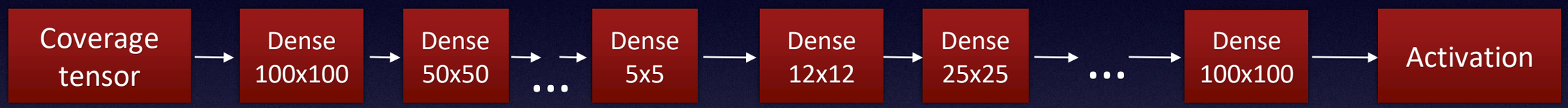
Implementation: Keras + tensorflow + nVidia Pascal X

Model 2



Implementation: Keras + tensorflow + nVidia Pascal X

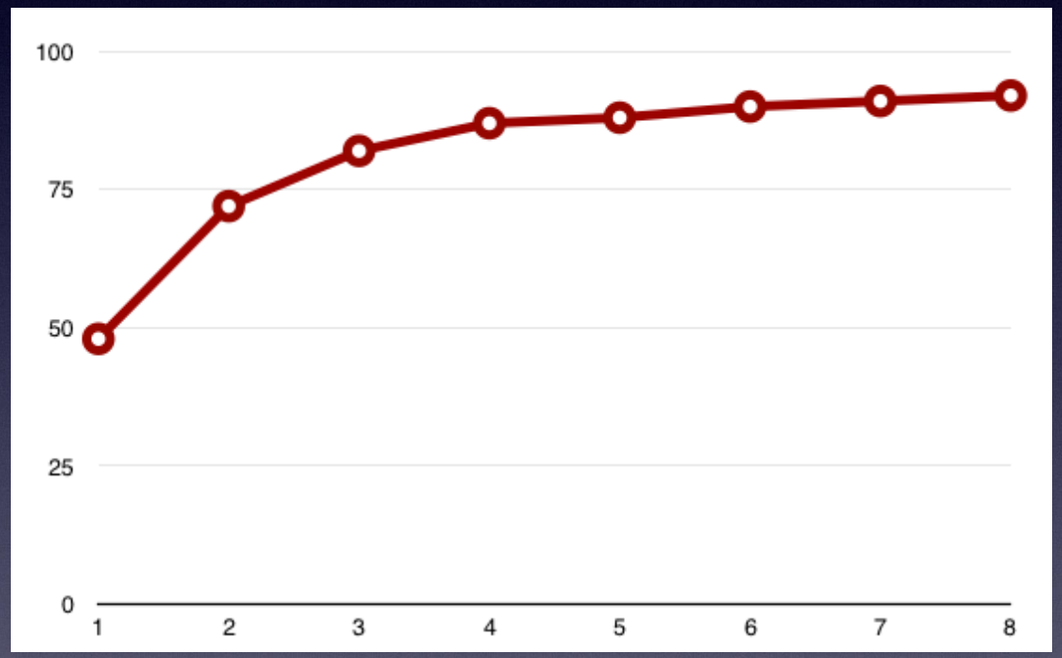
Model 3



1 0 0 0	0.10 0.04 -0.00 -0.10
1 1 0 0	0.28 -0.40 0.02 -0.07
0 0 1 1	-0.07 0.02 0.08 0.47

Implementation: Keras + tensorflow + nVidia Pascal X

Результат



Заключение

- Данные = сырая нефть. Ими не делятся
- Много интересного в Offline ритейле
- Deep Learning - умное дополнение к корпоративным системам
- Deep Learning - стоит изучать и инвестировать

Спасибо!

Максим Усов

+7 3822 48 85 83

m.usov@enbisys.com

vk.com/enbisys

facebook.com/enbisys