

# Введение в анализ данных

Лекция 15

Transfer learning

(обучение с переносом знаний)

Евгений Соколов

[esokolov@hse.ru](mailto:esokolov@hse.ru)

НИУ ВШЭ, 2017

# Named Entity Recognition

- Задача: выделить в тексте упоминания персон и организаций
- «Jim bought 300 shares of Acme Corp. in 2006»

# Named Entity Recognition

- Задача: выделить в тексте упоминания персон и организаций
- «**Jim** bought 300 shares of **Acme Corp.** in 2006»

# Named Entity Recognition

- Машинное обучение показывает высокие результаты
- Объект — одно слово
- Целевая переменная — «персона», «компания», «ничего»
- Признаки:
  - Текущее слово и его часть речи
  - Два слова до текущего
  - Два слова после текущего
- Методы — логистическая регрессия или другие

# Named Entity Recognition

## Quote from Wikipedia

“State-of-the-art NER systems produce near-human performance. For example, the best system entering MUC-7 scored 93.39% of f-measure while human annotators scored 97.60% and 96.95%”

- ...но задача не решена
- Почему?

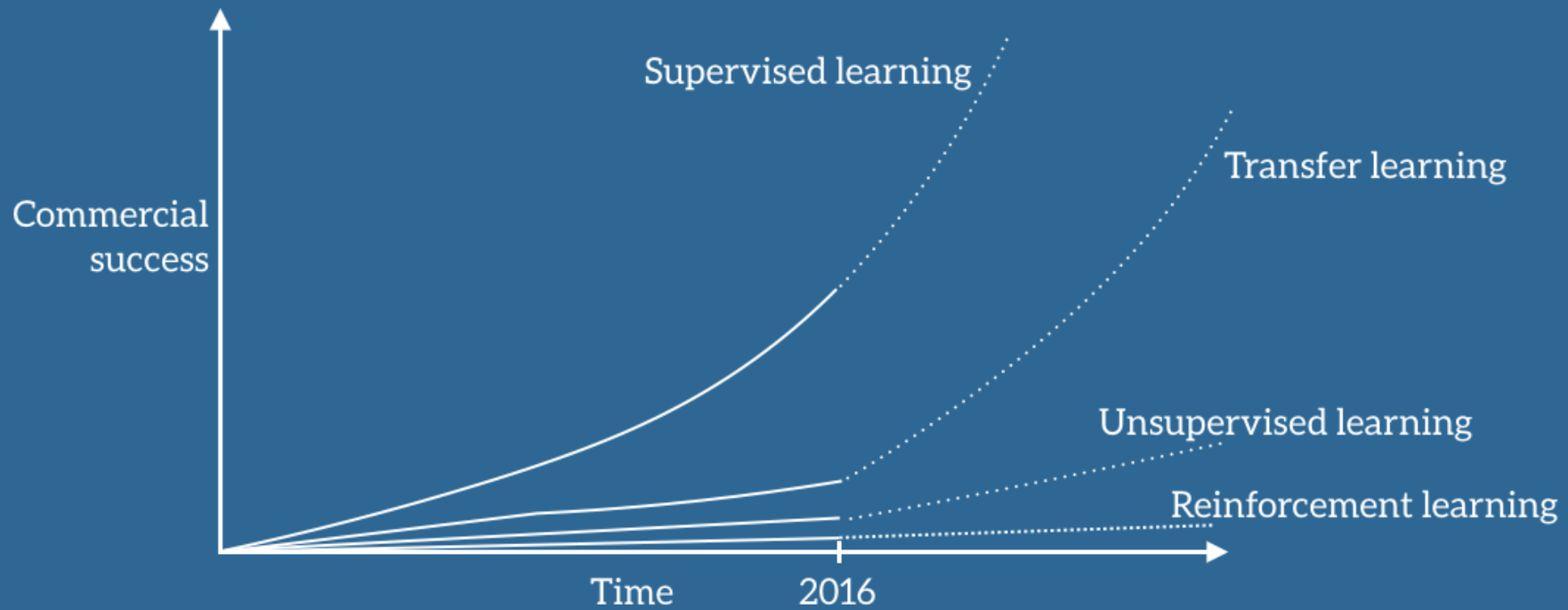
# Named Entity Recognition

## Quote from Wikipedia

“State-of-the-art NER systems produce near-human performance. For example, the best system entering MUC-7 scored 93.39% of f-measure while human annotators scored 97.60% and 96.95%”

- ...но задача не решена
- Почему?
- Для каждой новой задачи нужно собирать разметку и обучать новую модель

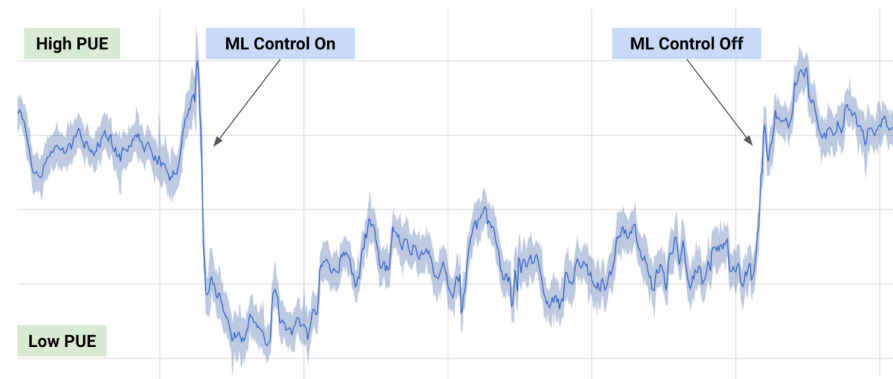
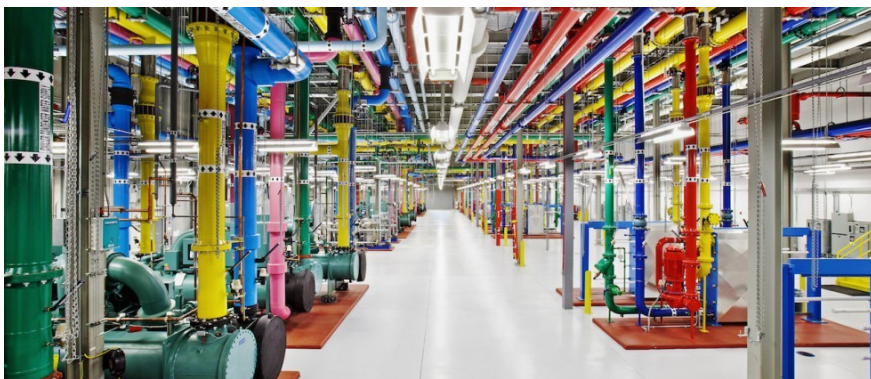
## Drivers of ML success in industry



- Andrew Ng, NIPS 2016 tutorial

# Отступление: охлаждение дата-центров

- Одна из главных компонент дата-центра — система охлаждения
- Результат работы системы сложным образом зависит от её параметров
- Необходимо быстро адаптироваться под изменение условий (нагрузка на серверы, погода)
- Все дата-центры разные — эвристические правила одного центра не работают в другом
- Машинное обучение позволило сократить затраты электричества на охлаждение на 40%

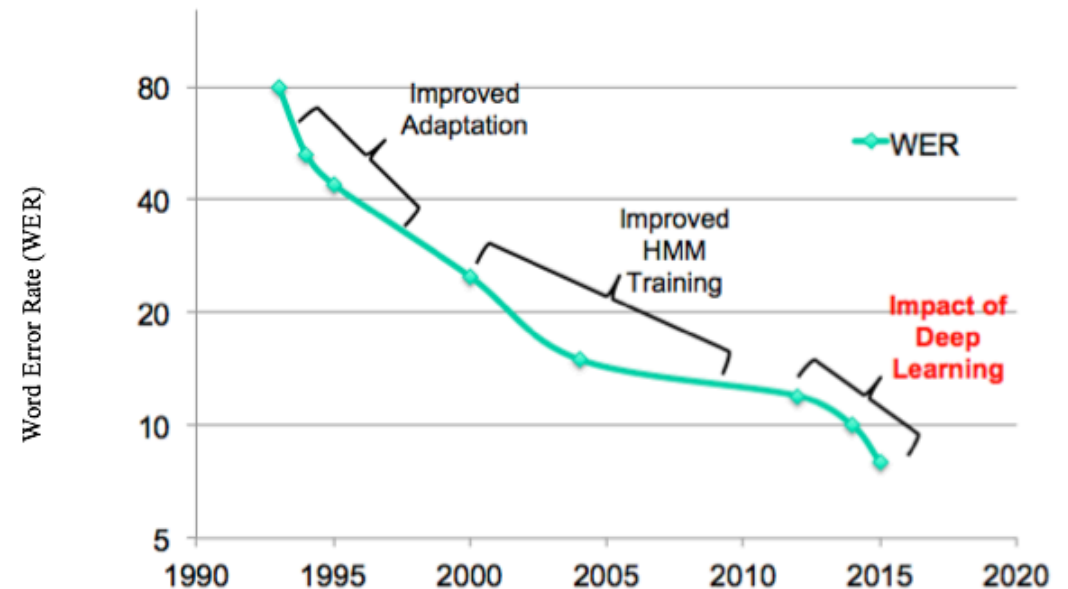


<https://deepmind.com/blog/deepmind-ai-reduces-google-data-centre-cooling-bill-40/>



# Supervised learning

- Классификация изображений — свёрточные сети (ResNet)
- Классификация текстов — рекуррентные сети
  - Google Smart Reply
- Машинный перевод — сложные рекуррентные архитектуры
- Распознавание речи
- В общем случае — GBM



# Supervised learning

- Основные результаты достигнуты на датасетах большого объёма
- Пример: ImageNet
  - Задача распознавания объектов на изображении
  - Более 10 миллионов размеченных вручную изображений (с помощью краудсорсинга)
- Что, если нужно решать задачу распознавания объектов, которые не входят в ImageNet?

# Supervised learning

- Применение классификатора к данным новой природы приводит к деградации качества

NER Task	Train → Test	F1
to find PER, LOC, ORG from news text	NYT → NYT	0.855
	Reuters → NYT	0.641
to find gene/protein from biomedical literature	mouse → mouse	0.541
	fly → mouse	0.281

# Подходы к решению

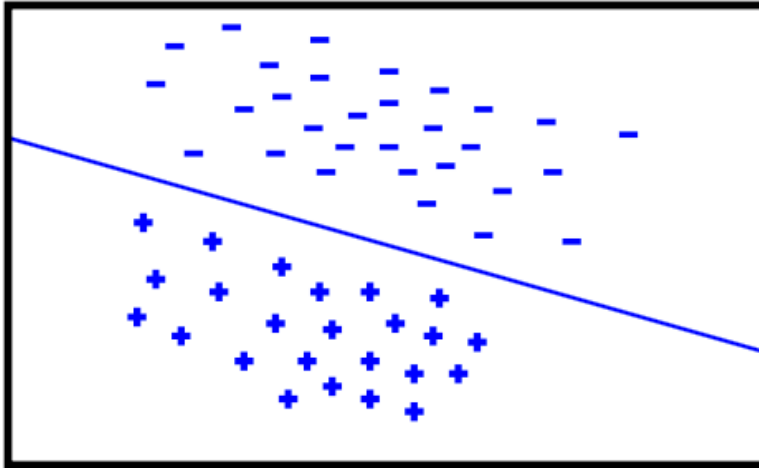
- Ничего не делать
  - Непредсказуемо низкое качество
- Разметить данные для новой задачи
  - Дорого
  - Краудсорсинг — много проблем с качеством разметки
- Собрать универсальный датасет
  - Wikipedia? ImageNet?
  - Невозможно покрыть все случаи
- Перенос знаний между задачами

# Transfer learning

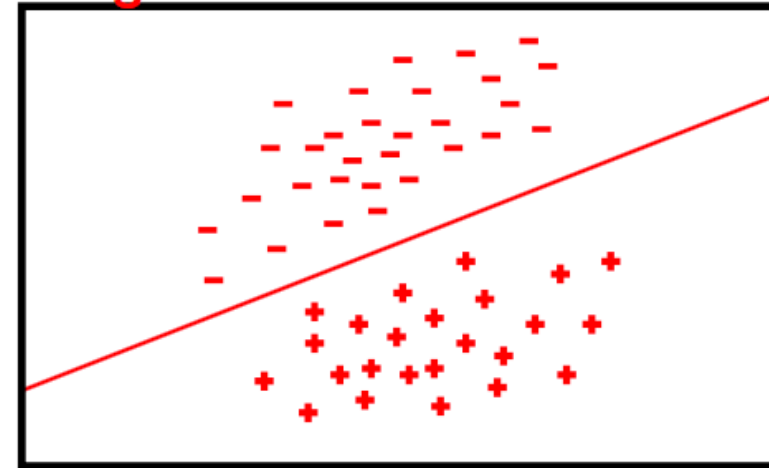
- Domain adaptation
  - Адаптация к смене распределения признаков
- Дообучение моделей
  - Пример — свёрточные сети, предобученные на ImageNet
- Перенос моделей между языками
  - Пока что мало успехов

# Domain adaptation

source domain



target domain



# Domain adaptation

- Объекты по-разному распределены на обучении и на контроле
- Идея: взвешивать объекты при обучении

$$\sum_{i=1}^{\ell} s_i (a(x_i) - y_i)^2 \rightarrow \min$$

- Большие веса будем ставить объектам, которые похожи на объекты из тестовой выборки

# Взвешивание объектов

- Можно показать:

$$s_i = \frac{P_{test}(x_i)}{P_{train}(x_i)}$$

- Обучим классификатор  $b(x)$ , который отличает объекты обучающей выборки от объектов тестовой выборки
- Классификатор  $b(x)$  должен выдавать вероятность принадлежности обучающей выборке (напр., логистическая регрессия)
- Тогда:

$$s_i = \frac{P_{test}(x_i)}{P_{train}(x_i)} = \frac{|X_{train}|}{|X_{test}|} \frac{b(x_i)}{1 - b(x_i)}$$



# Адаптация под новый словарь

- Пример: задача определения тональности отзыва

**Running with Scissors: A Memoir**

**Title: Horrible book, horrible.**

This book was horrible. I **read half** of it, **suffering from a headache** the entire time, and eventually **i lit it on fire**. One less copy in the world...don't waste your money. I wish i had the time spent reading this book back so i could use it for better purposes. This book wasted my life

**Avante Deep Fryer, Chrome & Black**

**Title: lid does not work well...**

I love the way the Tefal deep fryer cooks, however, I am **returning** my second one due to a **defective** lid closure. The lid may close initially, but after a few uses it no longer stays closed. I **will not be purchasing** this one again.

# Адаптация под новый словарь

- Если обучаться на отзывах на книги, то некоторые слова из отзывов на технику окажутся незнакомыми
- При этом по смыслу новые слова будут соответствовать каким-то уже известным

**read-half, headache** ↔ **defective, returned**

# Адаптация под новый словарь

## Unlabeled **kitchen** contexts

- Do **not buy** the Shark portable steamer .... Trigger mechanism is **defective**.
- the very nice lady assured me that I must have a **defective** set .... What a **disappointment**!
- Maybe mine was **defective** .... The directions were **unclear**

## Unlabeled **books** contexts

- The book is so **repetitive** that I found myself yelling .... I will definitely **not buy** another.
- A **disappointment** .... Ender was talked about for **<#> pages** altogether.
- it's **unclear** .... It's repetitive and **boring**

# Адаптация под новый словарь

- Идея — найти опорные слова
  - Часто встречаются и в обучении, и на контроле
  - На обучении их появление в тексте коррелирует с целевой переменной
  - Неспецифичные для конкретного датасета слова, характеризующие задачу

Unlabeled **kitchen** contexts

· Do **not buy** the Shark portable steamer .... Trigger mechanism is **defective**.  
· the very nice lady assured me that I must have a **defective** set .... What a **disappointment**!  
· Maybe mine was **defective** .... The directions were **unclear**

Unlabeled **books** contexts

· The book is so **repetitive** that I found myself yelling .... I will definitely **not buy** another.  
· A **disappointment** .... Ender was talked about for **<#> pages** altogether.  
· it's **unclear** .... It's repetitive and **boring**

# Адаптация под новый словарь

- Для каждого опорного слова построим классификатор, который определяет вхождение этого слова в текст
- Классификатор выявит слова, близкие по смыслу к опорному

(1) The book is so **repetitive** that I found myself yelling .... I will definitely [REDACTED] another.

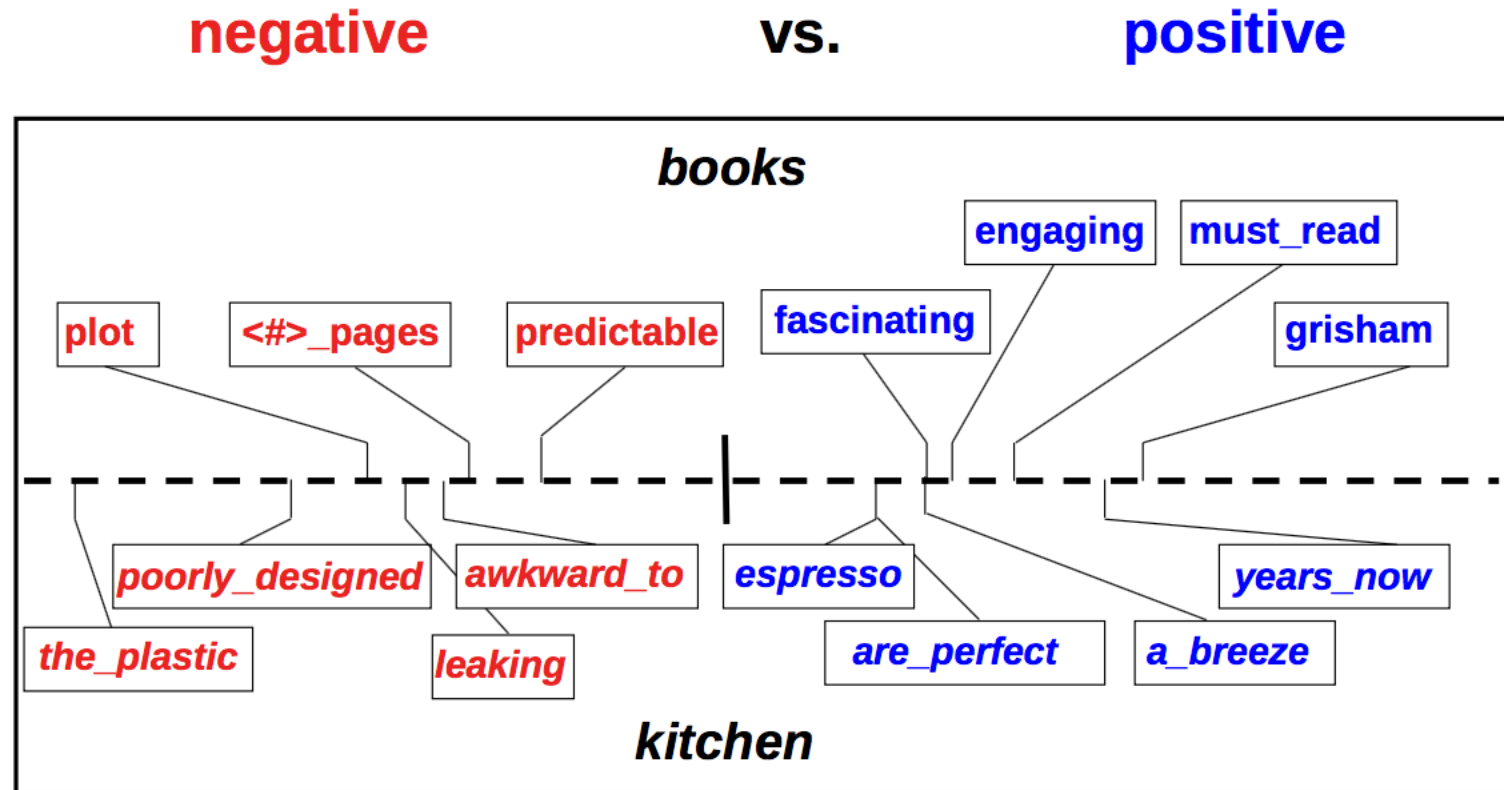
(2) Do [REDACTED] the Shark portable steamer .... Trigger mechanism is **defective**.

**Binary problem:** Does “**not buy**” appear here?

# Адаптация под новый словарь

- Всего  $N$  опорных слов
- Получим  $N$  классификаторов
- Новые признаки —  $N$  прогнозов от данных классификаторов
- Обучаем над новыми признаками модель
- Новая модель будет менее зависимой от конкретной задачи

# Адаптация под новый словарь



# Резюме

- Transfer learning — перенос знаний из задач с большим количеством размеченных данных на новые задачи
- При смене распределения признаков можно взвешивать объекты
- Адаптация под новые слова — подход с опорными словами