

# Neural Networks for Information Retrieval

Владимир Гулин

# Word embeddings and convolution neural networks for NLP

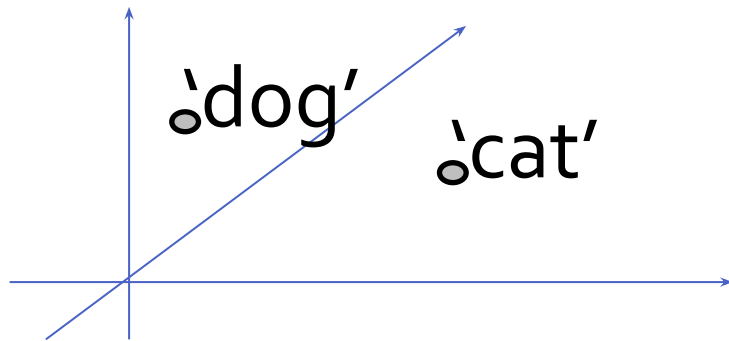
# Semantic Level NLP

- Поиск синонимов
- Ответы на вопросы
- Чат боты
- Машинный перевод
- Построение аннотаций
- ....

# Word-based NLP

Классический подход:

- Перенумеровываем слова ("один", "два")
- Используем *one-hot* encoding



$$X_{\text{мальчик}} : [0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0]$$

$$X_{\text{парень}} : [0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0]$$

$$X_{\text{олень}} : [0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]$$

**Проблема:**  $\|X_{\text{мальчик}} - X_{\text{парень}}\| = \|X_{\text{парень}} - X_{\text{олень}}\|$

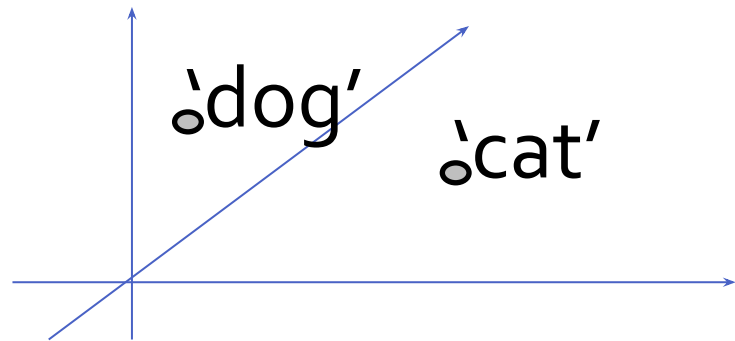
**Решение:** компактные представления

# Word-based NLP

$X_{\text{мальчик}}$  : [0 0 0 0 1 0 0 0 0]

$X_{\text{парень}}$  : [0 0 0 0 0 0 1 0 0]

$X_{\text{олень}}$  : [0 1 0 0 0 0 0 0 0]



**Проблема:** Очень высокая размерность пространства

**Костыль 1:** Лемматизация ("Парни -> парень", "Шёл" -> "Идти")

**Side-effect 1:** Лемматизация приводит к потере информации

**Костыль 2:** Заменить редкие слова заглушкой

**Side-effect 2:** Во многих задачах редкие слова крайне информативны

# Дистрибутивная гипотеза

## Гипотеза:

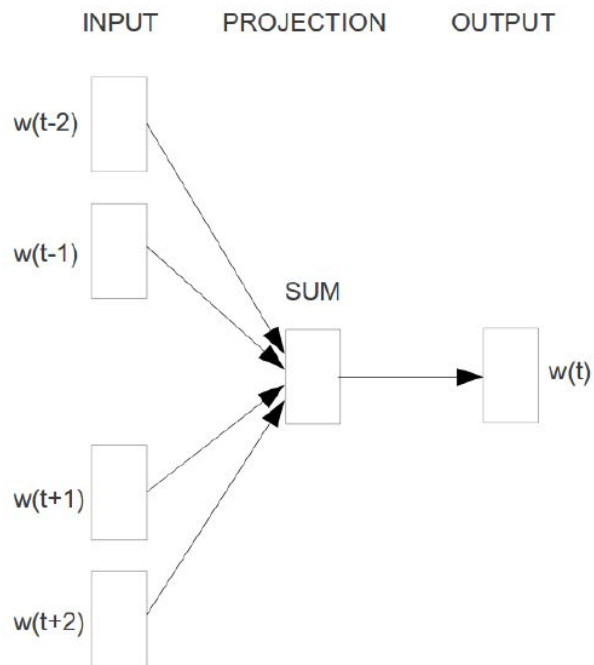
Лингвистические единицы, встречающиеся в схожих контекстах, имеют близкие значения.

## Вывод:

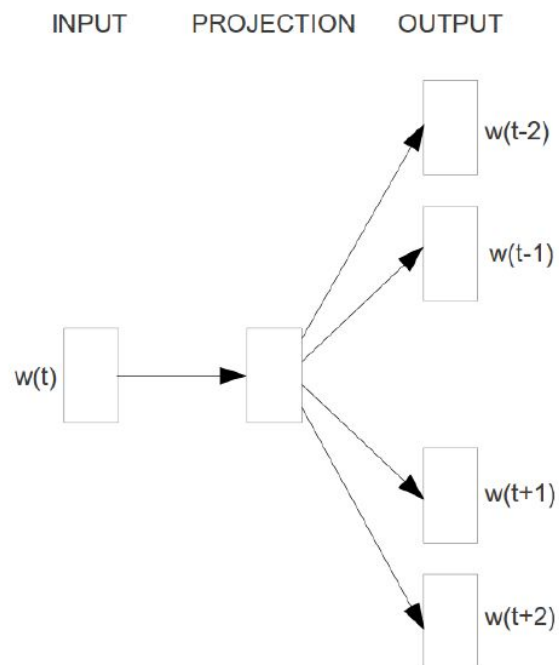
Значит, представления слов, можно построить с помощью контекстов, в которых эти слова употребляются.

# Word2Vec (2013)

## Архитектуры CBOW и Skip-gram

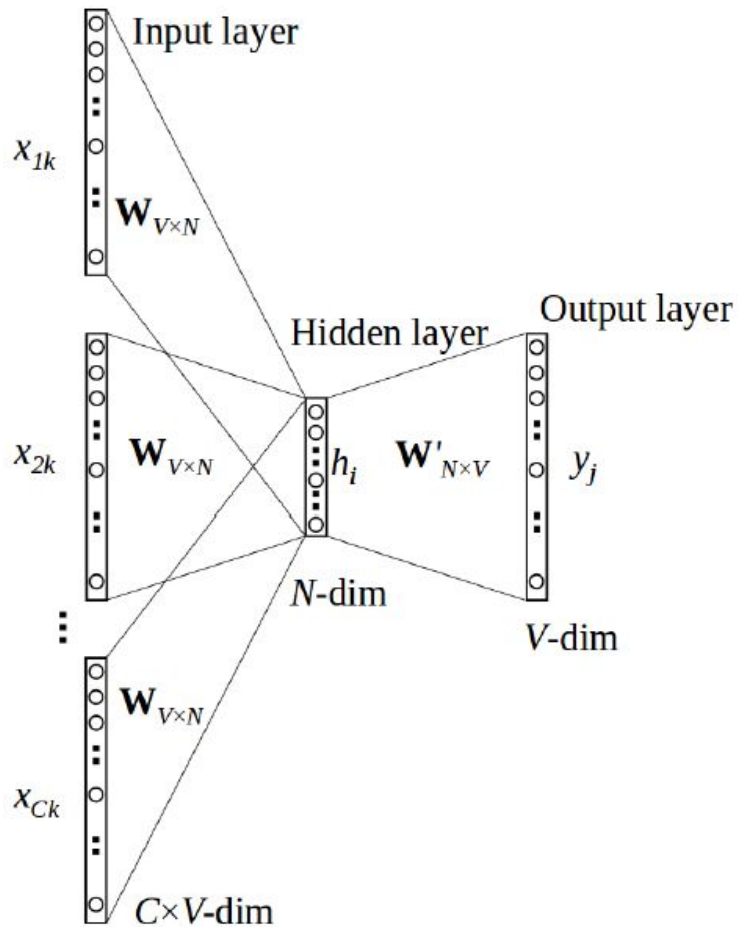


**CBOW**



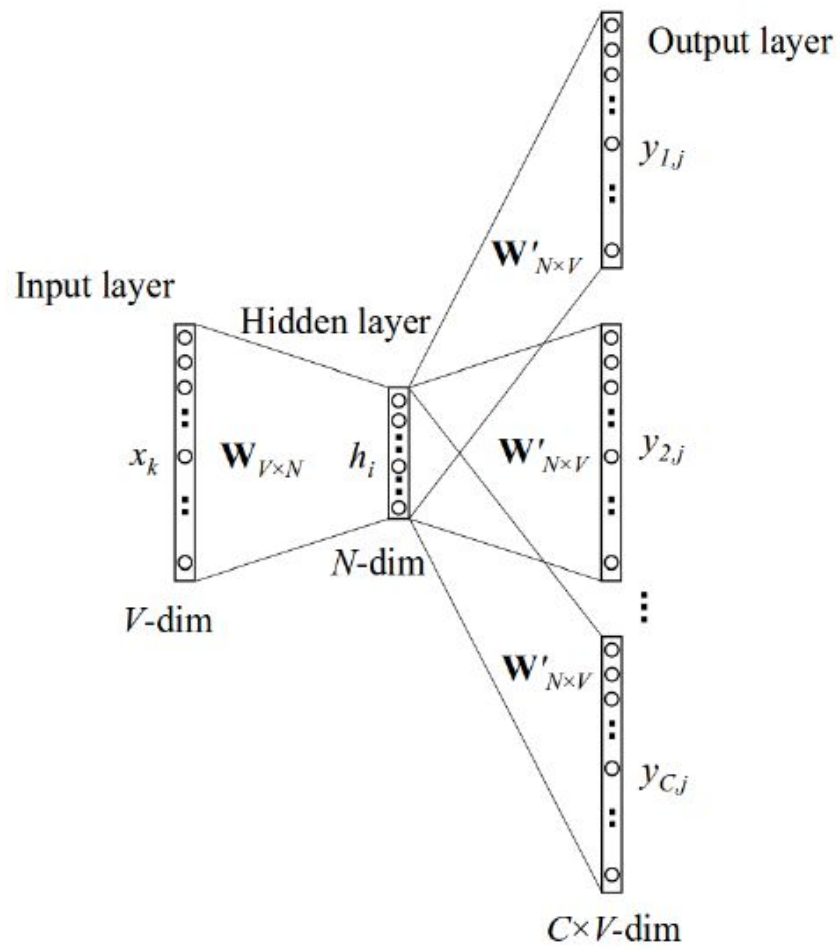
**Skip-gram**

# CBOW (Continuous Bag of Words)





# Skip-gram



# Skip-gram

Оптимизируем

$$\frac{1}{T} \sum_{t=1}^T \sum_{-c \leq j \leq c, j \neq 0} \log p(w_{t+j} | w_t)$$

# Skip-gram

## Вычисление вероятностей выходных слов

Используется softmax

$$p(w_O | w_I) = \frac{\exp \langle \mathbf{v}'_{w_O}, \mathbf{v}_{w_I} \rangle}{\sum_{w=1}^{|V|} \exp(\langle \mathbf{v}'_w, \mathbf{v}_{w_I} \rangle)}$$

На практике эту формулу применять сложно, так как вычисление градиента пропорционально  $|V|$ .

На практике применяют разные аппроксимации: иерархический softmax или negative sampling.

# Negative sampling

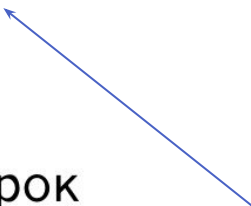
Идея:

Не будем рассматривать все слова из словаря, а учтем только рассматриваемое слово + подмешаем еще  $k$  отрицательных примеров.

Заменяем  $\log p(w_0|w_1)$  на

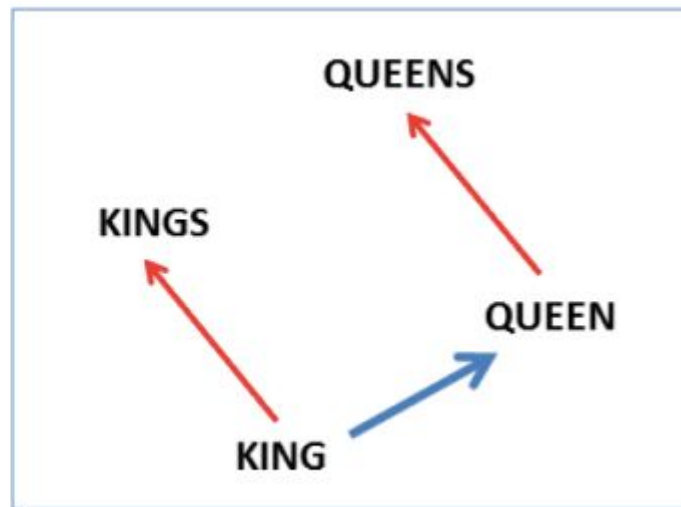
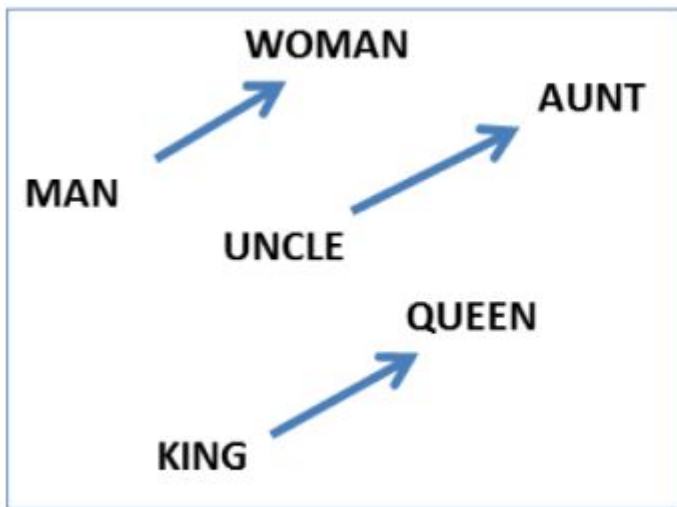
$$\log \sigma(\langle \mathbf{v}'_{w_0}, \mathbf{v}_{w_1} \rangle) + \sum_{i=1}^k \mathbb{E}_{w_i \sim P_n(w)} [\log \sigma(-\langle \mathbf{v}'_{w_i}, \mathbf{v}_{w_1} \rangle)]$$

- ▶  $k \approx 5 - 20$  для небольших выборок
- ▶  $k \approx 2 - 5$  для больших данных

$$\frac{1}{Z} P^{\text{empirical}}(w)$$


# Свойства выученных представлений

$$\text{vec}(\text{"man"}) - \text{vec}(\text{"king"}) + \text{vec}(\text{"woman"}) = \text{vec}(\text{"queen"})$$



# Пример модели

## Ближайшие по контекстам слова для *дефицит* (с косинусной близостью)

1. [нехватка](#) 0.65354729
2. [профицит](#) 0.62004328
3. [госдолг](#) 0.61431211
4. [несбалансированность](#) 0.59282905
5. [дефицитность](#) 0.58463109
6. [дисбаланс](#) 0.57497036
7. [недопоступление](#) 0.57044709

## Ближайшие по контекстам слова для *сингл* (с косинусной близостью)

1. [альбом](#) 0.91364485
2. [лонгплей](#) 0.88309568
3. [пластинка](#) 0.85516018
4. [кавер](#) 0.84436756
5. [студийник](#) 0.83440500

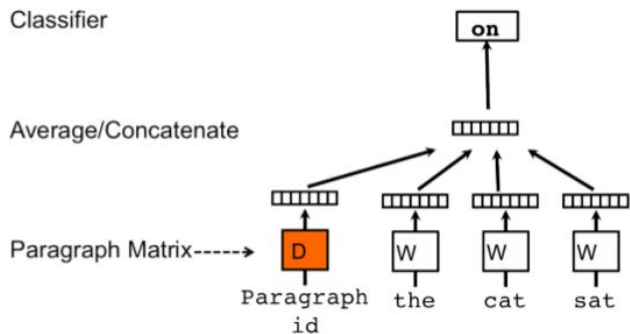
# Doc2vec (Mikolov, 2014)

- Обобщение word2vec на целые документы (фразы, предложения и т.д.)
- Преобразует текст произвольной длины в вектор фиксированного размера
- Distributed Memory (DM)
- Distributed Bag of Words (DBOW)

# Doc2vec (Mikolov, 2014)

## Distributed Memory (DM)

- ▶ Назначим и рандомно проинициализируем paragraph vector
- ▶ Будем предсказывать слово из текста используя контекст и paragraph vector
- ▶ Идем скользящим окном по всему документу, сохраняя при этом paragraph vector фиксированным (поэтому Distributed Memory)
- ▶ Обновление происходит при помощи SGD и backprop

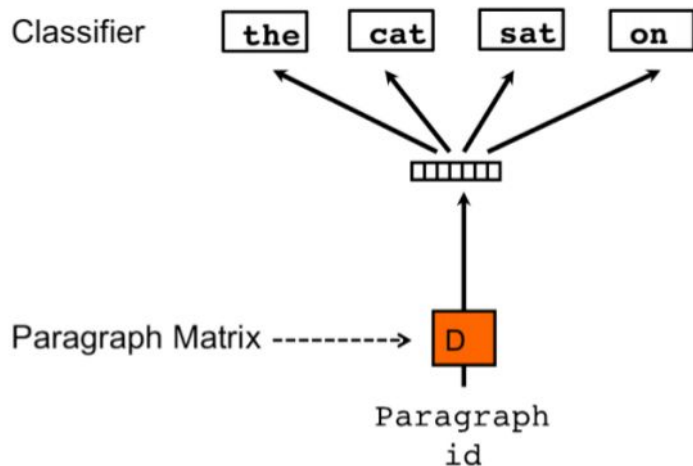




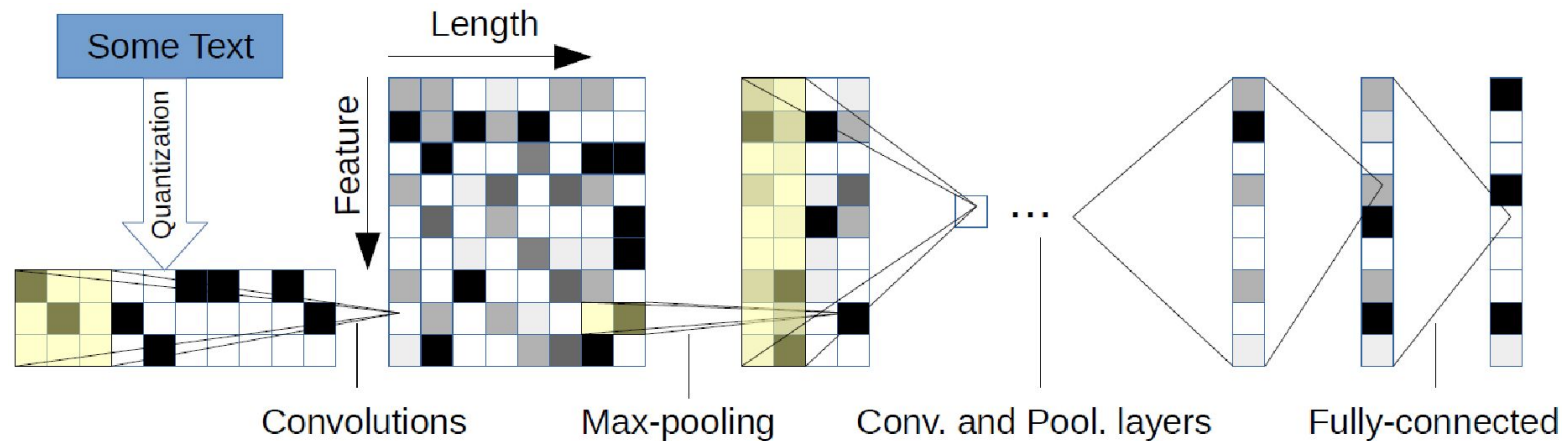
# Doc2vec (Mikolov, 2014)

## Distributed Bag of Words (DBOW)

- ▶ Используем только paragraph vector (вектора слов не используем)
- ▶ Берем окно из слов в параграфе и случайно семплируем какое из слов предсказать используя paragraph vector (игнорируем порядок слов)
- ▶ Очень просто и требует меньше ресурсов

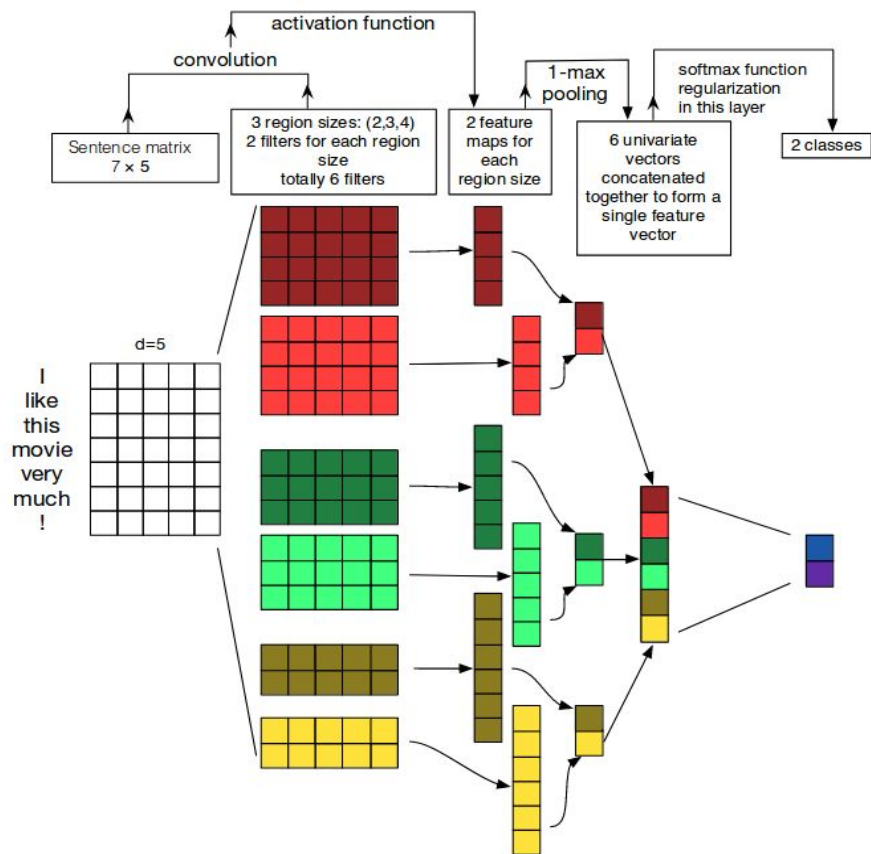


# Character-level ConvNet

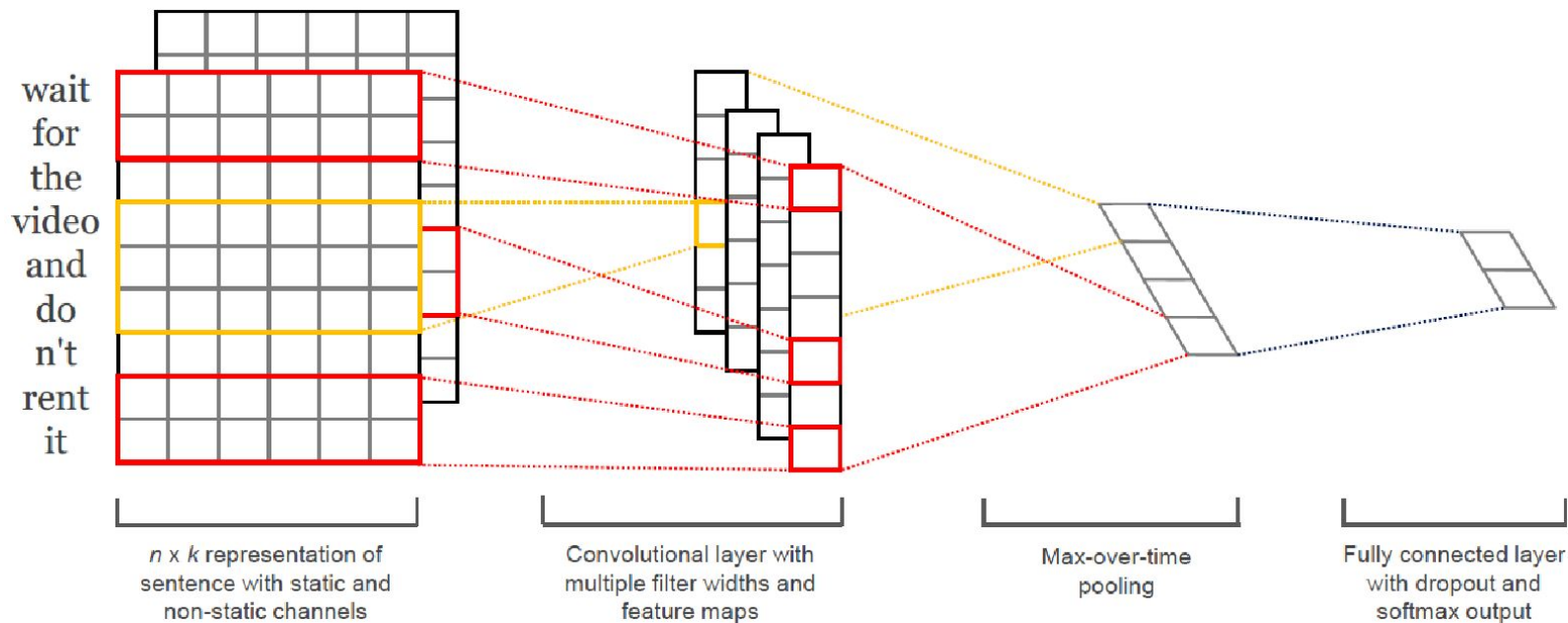


Trained for text classification

# Векторное представление текста как изображения



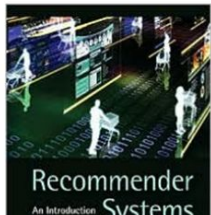
# Convolutional Neural Networks for Sentence Classification



Обычная сверточная сеть, проинициализированная word2vec представлениями слов

# Neural networks for recommendation systems

# Рекомендательные системы



## Recommender Systems: An Introduction

by Dietmar Jannach, Markus Zanker, Alexander Felfernig, Gerhard Friedrich

AVERAGE CUSTOMER RATING:

☆☆☆☆☆ ( Be the first to review )

Gefällt mir

Registrieren, um sehen zu können, was deinen Freunden gefällt.

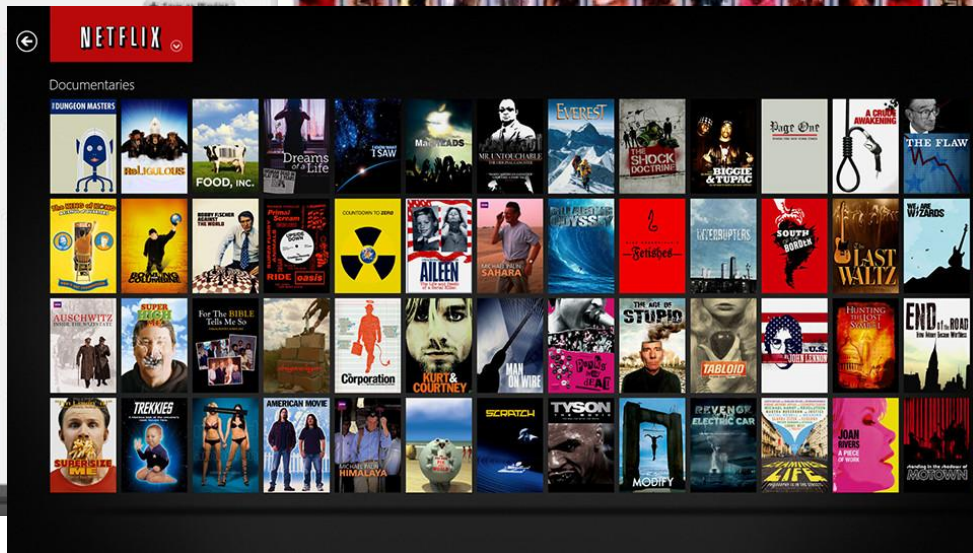
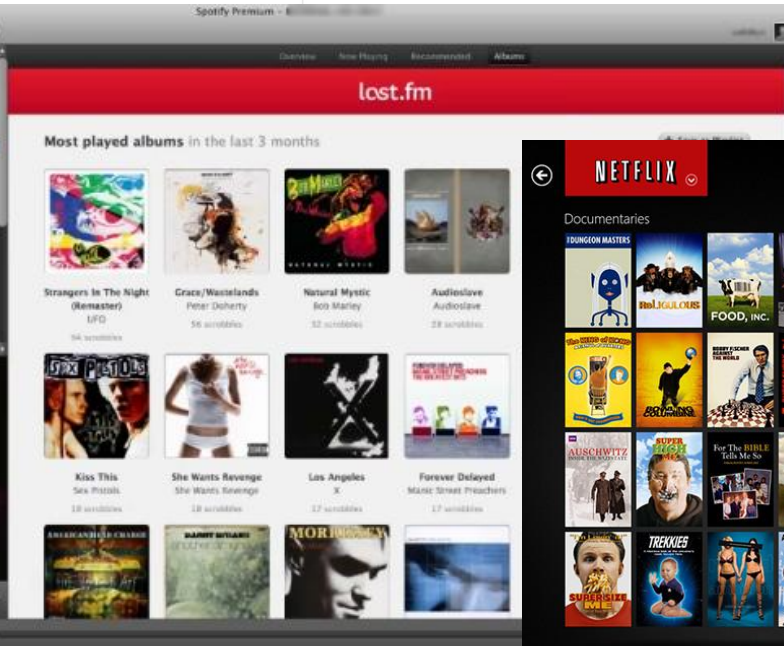
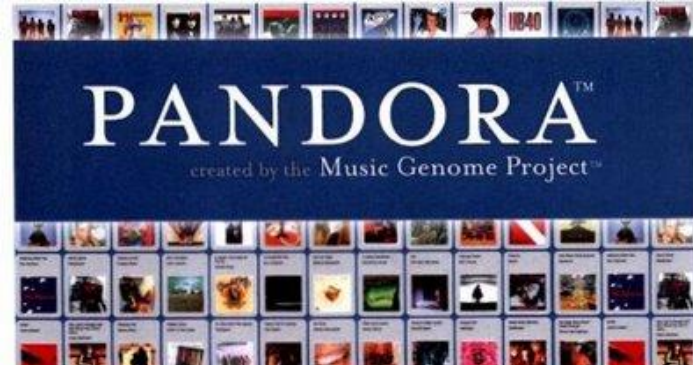
FORMAT:  
Hardcover

NEW FROM BN.COM

\$65.00 List Price

**\$52.00** Online Price  
(You Save 20%)

**Add to Cart**





# E-commerce

MTS RUS 16:30

Смартфоны на Android  
Покупай по безопасной сделке

Юля  
Прод

2,9 км

4 900 ₺

Куртка зимняя му...

3,2 км

900 ₺

Книга Гарри потер...

2,5 км

25 000 ₺

Mercedes Brabus

39,1 км

6 850 000 ₺

Квартира, 3 комнаты,...

MTS RUS 16:28

PANDAO

Поиск товаров в Китае

Лучшее Товары за баллы Красота и здс

-50%

Наушники с  
экранированным пр...

62 205 заказов

98 ₺ 49 ₺

-88%

Подставкой для  
мобильных телефон...

30 146 заказов

133 ₺ 16 ₺

-51%

-70%

Q

♡

🛒

📄

👤

MTS RUS 16:29

Кооперативная улица, 3к5

Сортировка и фильтры

IL Патио

Заказ от 300 ₺ · Итальянская

Delivery Express

45 мин

BB&Burgers

45 мин

# Recommender systems - The task

- Нужно построить модель, предсказывающую насколько user нравятся items
- У типичной рекомендательной системы имеется:
  - Матрица users и items
  - Рейтинги, отражающие их прошлые предпочтения
  - Задача состоит в предсказании будущих предпочтений
- В общем случае, все **НЕ** ограничивается предсказанием рейтинга



2			4	5	
5		4			1
		5		2	
	1		5		4
		4		2	
4	5		1		



# Approaches to recommender systems

- Collaborative filtering

Основан на анализе предпочтений пользователей таких как рейтинги фильмов, лайки и т.д.

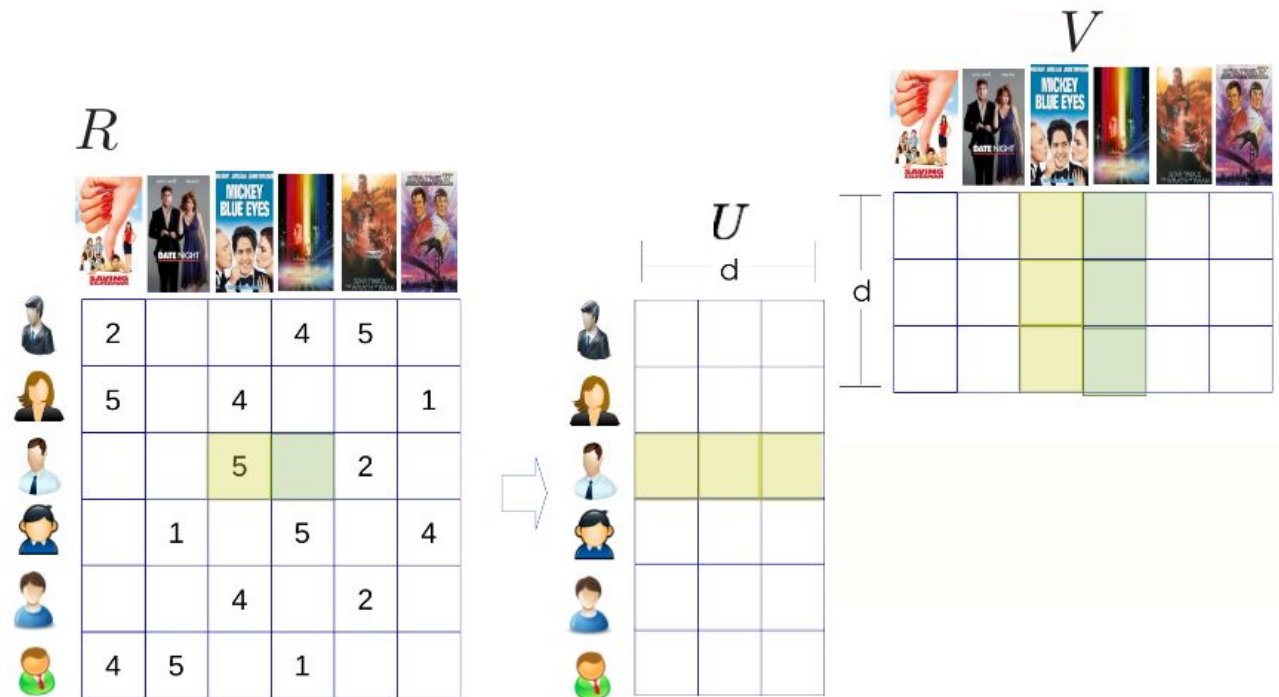
- Content-based filtering

Основан на матчинге представлений items и users

Пользовательский профиль может представляться предыдущими рейтингами, посещенными страницами в интернете, поисковыми запросами и т.д.

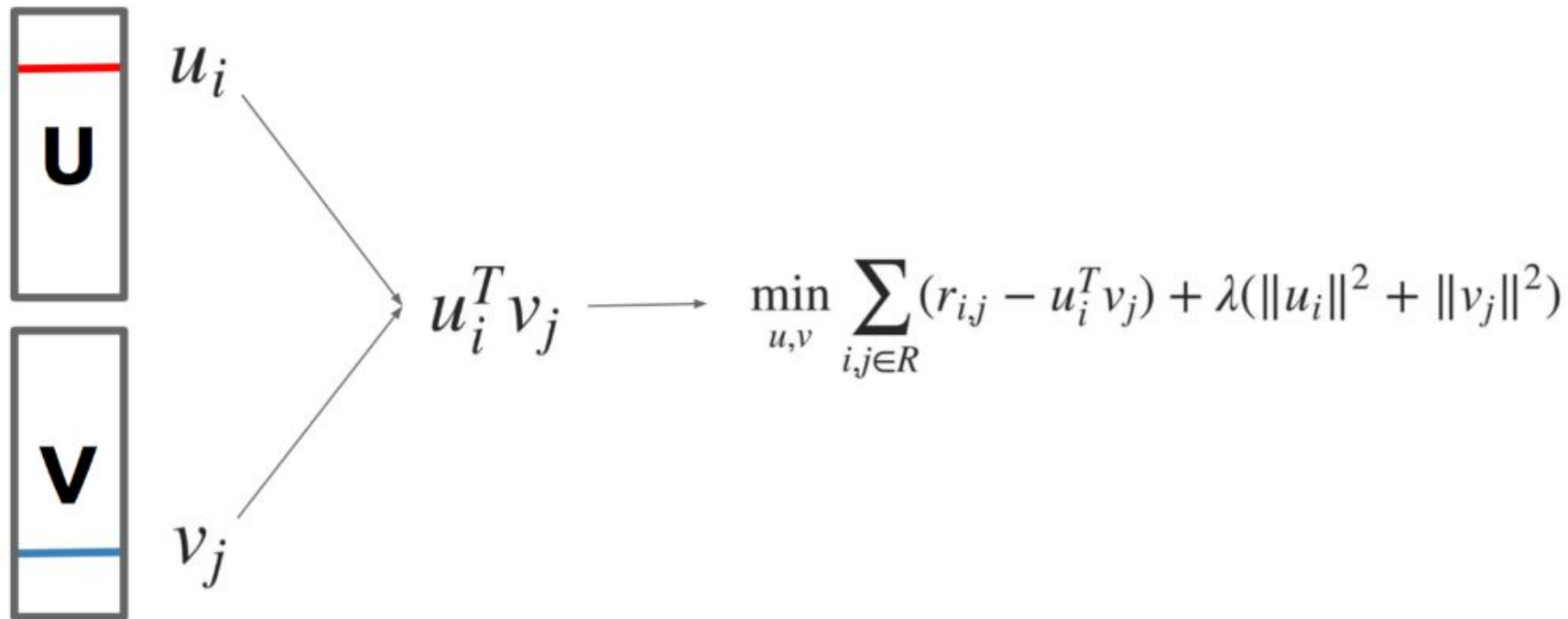
- A hybrid approach

# Matrix factorization



$$\min_{u,v} \sum_{i,j \in R} (r_{i,j} - u_i^T v_j) + \lambda(\|u_i\|^2 + \|v_j\|^2)$$

# A feed-forward neural network view



# Pytorch example

```
import torch
from torch.autograd import Variable

class MatrixFactorization(torch.nn.Module):
    def __init__(self, n_users, n_items, n_factors=20):
        super().__init__()
        # create user embeddings
        self.user_factors = torch.nn.Embedding(n_users, n_factors,
                                              sparse=True)

        # create item embeddings
        self.item_factors = torch.nn.Embedding(n_items, n_factors,
                                              sparse=True)

    def forward(self, user, item):
        # matrix multiplication
        return (self.user_factors(user)*self.item_factors(item)).sum(1)

    def predict(self, user, item):
        return self.forward(user, item)

model = MatrixFactorization(n_users, n_items, n_factors=20)
loss_fn = torch.nn.MSELoss()
optimizer = torch.optim.SGD(model.parameters(),
                              lr=1e-6)

for user, item in zip(users, items):
    # get user, item and rating data
    rating = Variable(torch.FloatTensor([ratings[user, item]]))
    user = Variable(torch.LongTensor([int(user)]))
    item = Variable(torch.LongTensor([int(item)]))

    # predict
    prediction = model(user, item)
    loss = loss_fn(prediction, rating)

    # backpropagate
    loss.backward()

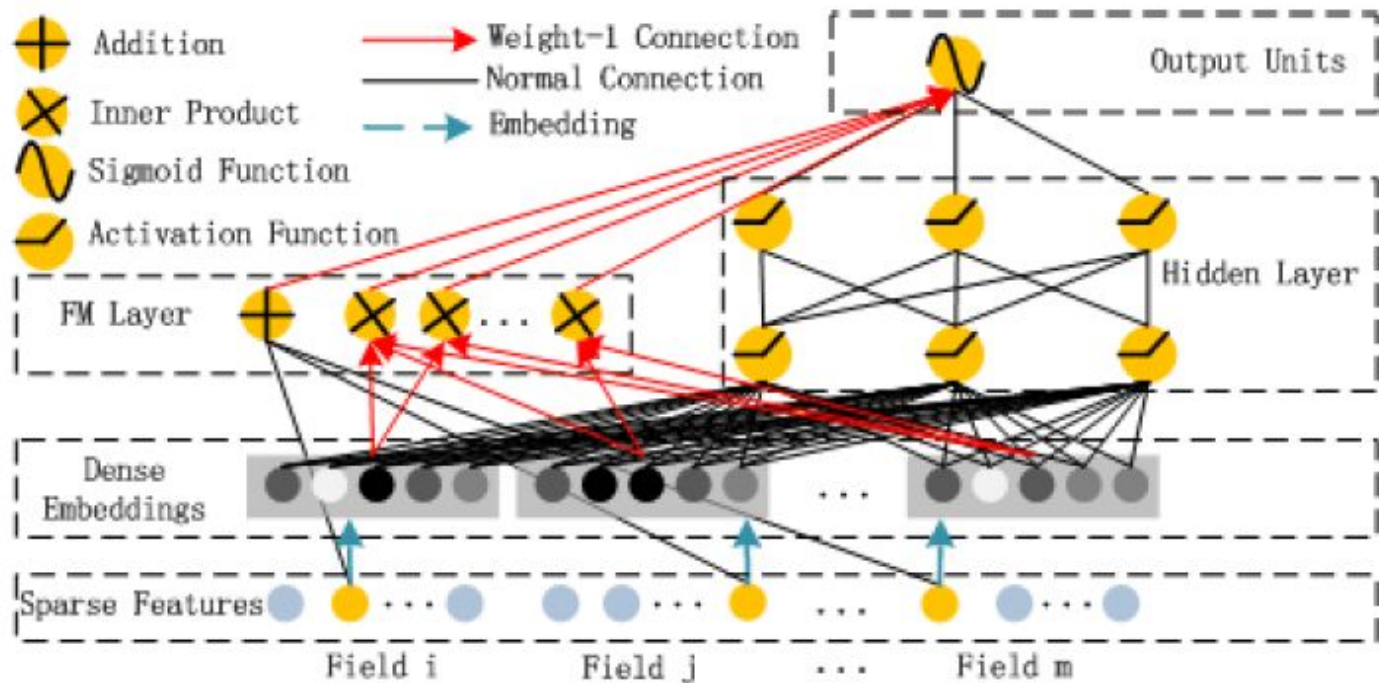
    # update weights
    optimizer.step()
```

# Factorization Machines

$$\hat{y}(x) = w_0 + \sum_{j=1}^p w_j x_j + \sum_{j=1}^p \sum_{j'=j+1}^p x_j x_{j'} \mathbf{v}_j^T \mathbf{v}_{j'}$$

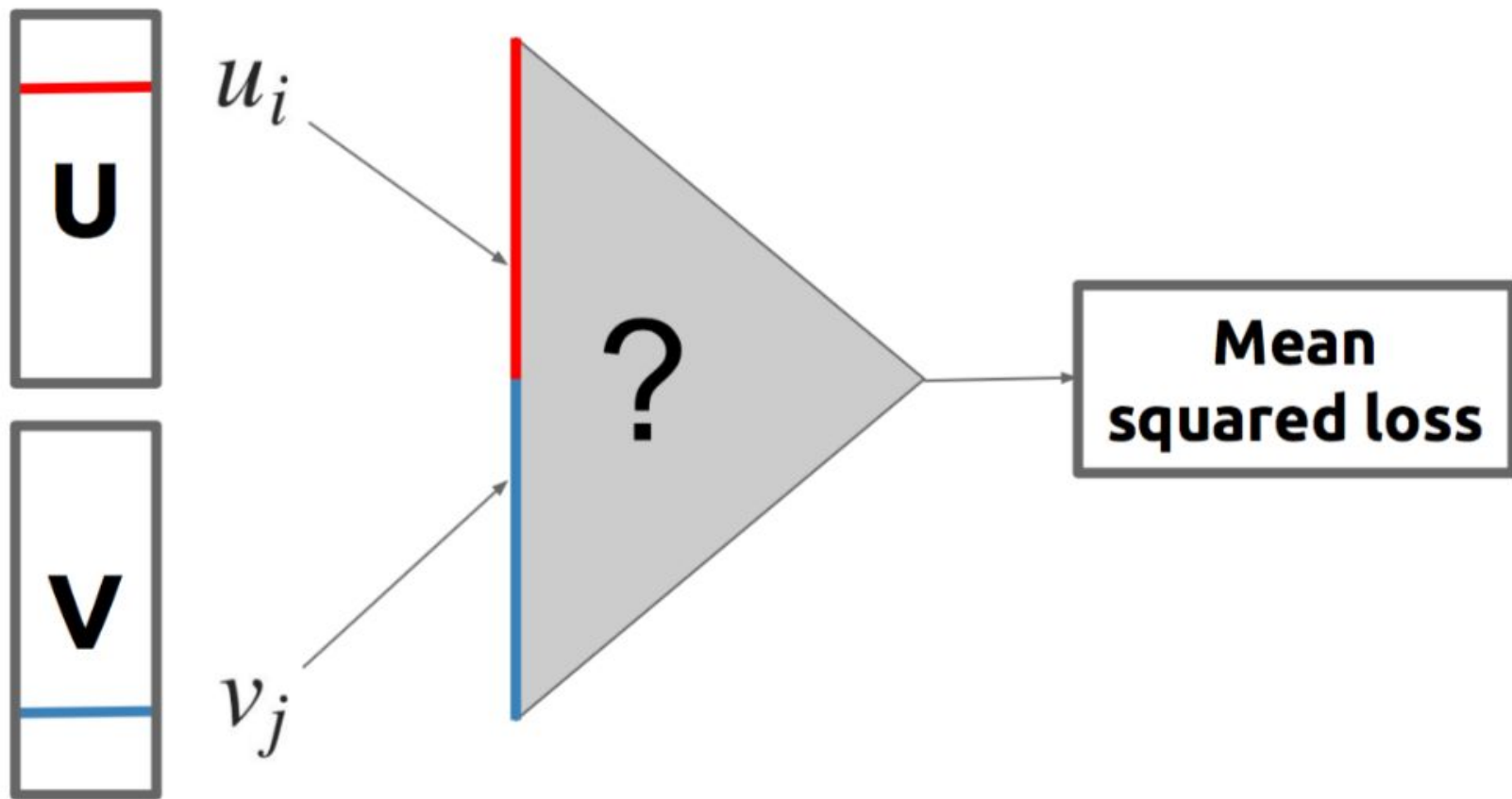
Feature vector $\mathbf{x}$															Target $y$							
$x_1$	1	0	0	...	1	0	0	0	...	0.3	0.3	0.3	0	...	13	0	0	0	0	...	5	$y_1$
$x_2$	1	0	0	...	0	1	0	0	...	0.3	0.3	0.3	0	...	14	1	0	0	0	...	3	$y_2$
$x_3$	1	0	0	...	0	0	1	0	...	0.3	0.3	0.3	0	...	16	0	1	0	0	...	1	$y_3$
$x_4$	0	1	0	...	0	0	1	0	...	0	0	0.5	0.5	...	5	0	0	0	0	...	4	$y_4$
$x_5$	0	1	0	...	0	0	0	1	...	0	0	0.5	0.5	...	8	0	0	1	0	...	5	$y_5$
$x_6$	0	0	1	...	1	0	0	0	...	0.5	0	0.5	0	...	9	0	0	0	0	...	1	$y_6$
$x_7$	0	0	1	...	0	0	1	0	...	0.5	0	0.5	0	...	12	1	0	0	0	...	5	$y_7$
	A	B	C	...	TI	NH	SW	ST	...	TI	NH	SW	ST	...	Time	TI	NH	SW	ST	...		
	User				Movie					Other Movies rated						Last Movie rated						

# Factorization Machines as Neural Network

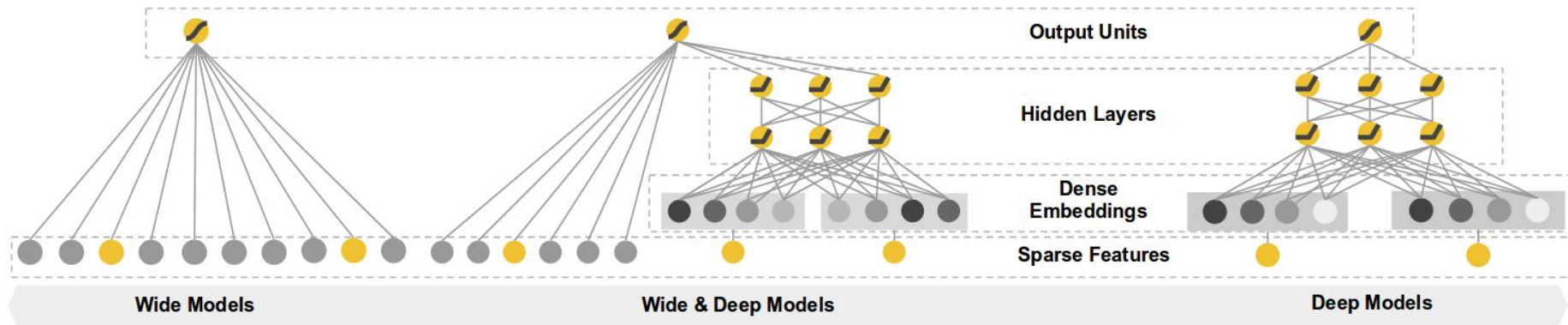


$$\hat{y}(x) = w_0 + \sum_{j=1}^p w_j x_j + \sum_{j=1}^p \sum_{j'=j+1}^p x_j x_{j'} \mathbf{v}_j^T \mathbf{v}_{j'}$$

# A deeper view

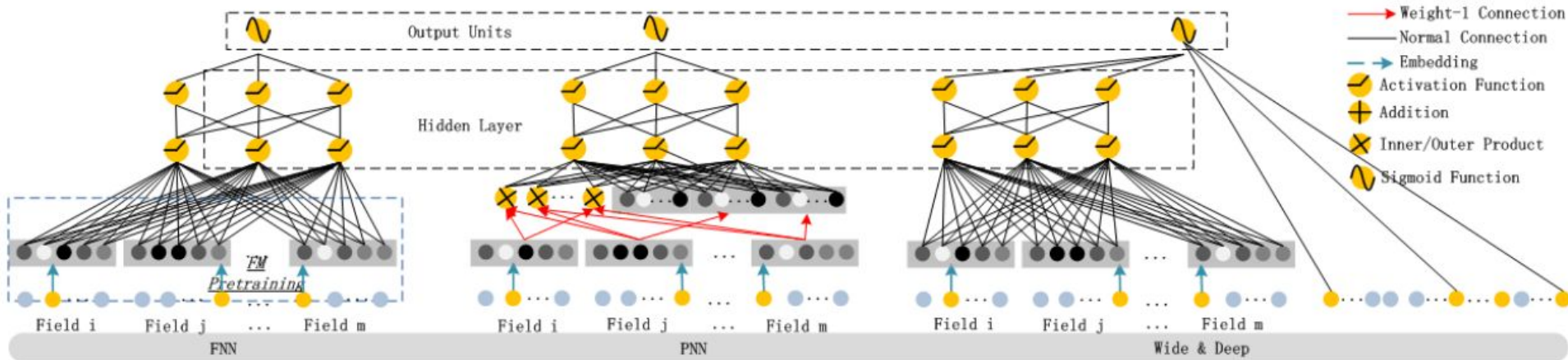


# Wide and Deep (Google 2016)





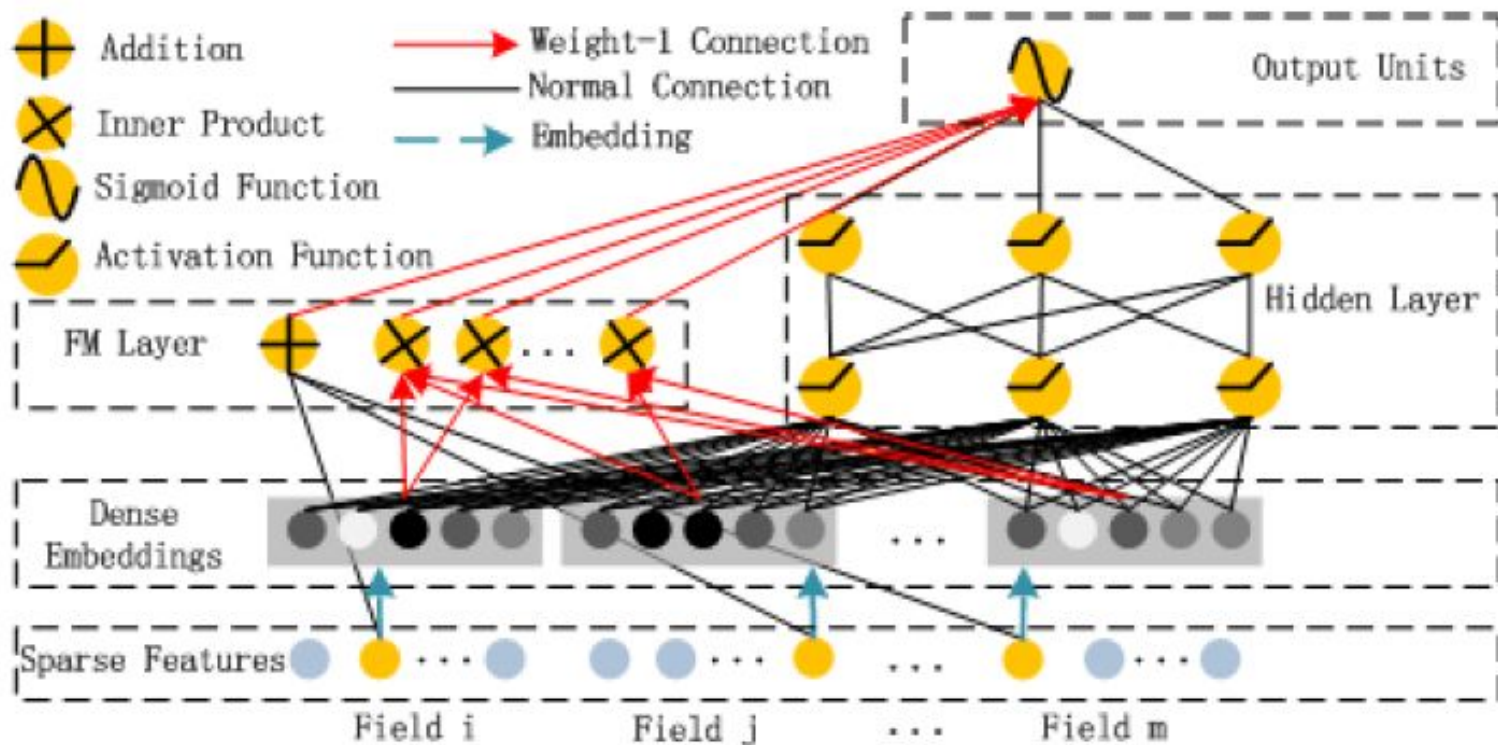
# Нейросетевые архитектуры для CF



FNN - FM initialized feed forward neural network

PNN - Product based neural network

# DeepFM (2017)



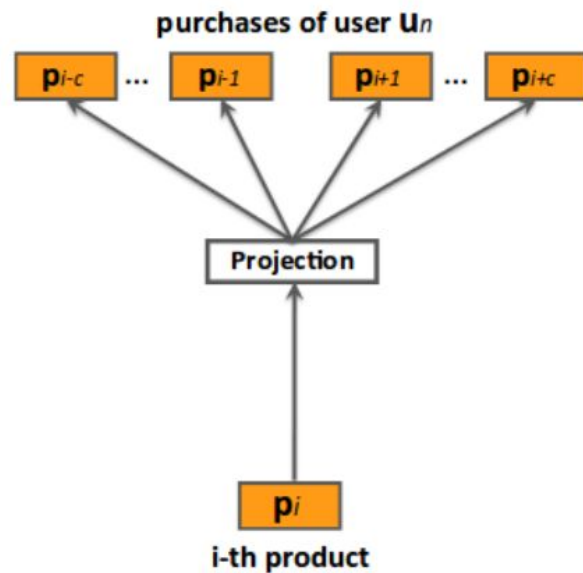
# DeepFM (2017)

Результаты:

	Company*		Criteo	
	AUC	LogLoss	AUC	LogLoss
LR	0.8640	0.02648	0.7686	0.47762
FM	0.8678	0.02633	0.7892	0.46077
FNN	0.8683	0.02629	0.7963	0.45738
IPNN	0.8664	0.02637	0.7972	0.45323
OPNN	0.8658	0.02641	0.7982	0.45256
PNN*	0.8672	0.02636	0.7987	0.45214
LR & DNN	0.8673	0.02634	0.7981	0.46772
FM & DNN	0.8661	0.02640	0.7850	0.45382
DeepFM	<b>0.8715</b>	<b>0.02618</b>	<b>0.8007</b>	<b>0.45083</b>

# Prod2Vec and Item2Vec

- Prod2Vec & item2vec: факторизация матрицы совместной покупки товаров
- Аналог word2vec за исключением игнорирования пространственной составляющей



Prod2vec skip-gram model

$$\frac{1}{K} \sum_{i=1}^K \sum_{-c \leq j \leq c, j \neq 0} \log p(w_{i+j} | w_i) \rightarrow \frac{1}{K} \sum_{i=1}^K \sum_{j \neq i}^K \log p(w_j | w_i)$$

# Pandao2Vec

- Вектора Pandao2Vec отражают семантику
- Для них даже (иногда) работают аналогии:



Чехол  
для  
iPhone

—



Стекло  
для  
iPhone

+



Стекло  
для  
Samsung

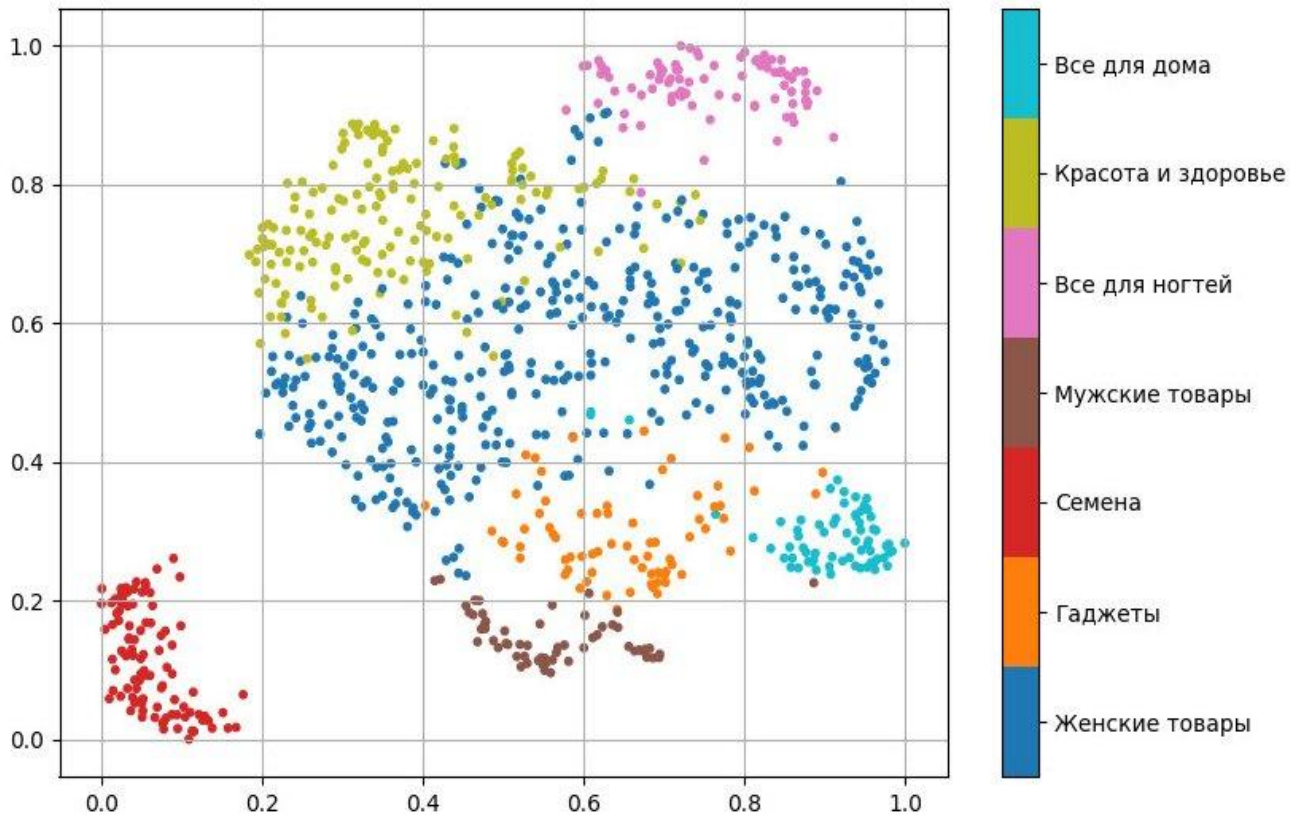
=



Чехол  
для  
Samsung

# Pandao2Vec

Топ 1000 товаров



# Neural networks for information retrieval

# Задача ранжирования в поиске

Цель:

Хотим научиться считать “правильные” динамические докумено-запросные факторы.

NDCG

$$DCG = \sum_{i=1}^{N_q} \frac{2^{rel_i} - 1}{\log_2 i + 1}, \quad NDCG = \frac{DCG}{DCG_{max}}$$



## Котики — ВКонтакте

[vk.com](#) > [v.kote](#)

Дата рождения: 19 янв 2011

О себе: Комиксы: #котмиксы@v.kote Анимашки: #котогифки@v.k...

[Читать ещё >](#)



## Кошка



[Все](#) [Картинки](#) [Видео](#) [Новости](#) [Карты](#) [Ещё](#) [Настройки](#) [Инструменты](#)

Результатов: примерно 5 580 000 (0,42 сек.)

### Видео



1:08:39

Космос -  
Вселенная от  
начала до конца  
(1 из 4)

History and Economics  
[YouTube](#) - 20 авг. 2012 г.



53:44

Вселенная.  
Путешествие в  
Космос. В поисках  
ответов ...

SDF 35  
[YouTube](#) - 15 окт. 2017 г.



25:08

Край вселенной  
Где заканчивается  
космос  
Документальный ...

Galaxy Channel  
[YouTube](#) - 2 июн. 2017 г.



ньонов». С точки  
иства кошачьих



Каракал

## Мстители (2012) смотреть онлайн бесплатно

[kinogo.net/110-mstiteless-2012....](#)

«Мстители» – фантастический боевик о новых приключениях супергероев. Сюжет картины посвящен извечному противостоянию добра и зла.

## Мстители 3: Война бесконечности смотреть онлайн фильм бесплатно в...

[kino-1080hd.ru > Боевик](#)

Пока правительство Соединенных Штатов Америки пытается восстановить разрушения после вторжения пришельцев, а Мстители устраивают разборки из-за...

## Смотреть Первый мститель: Противостояние в хорошем качестве hd 720

[kinoabr.net > Фантастика](#)

Все это заканчивается тем, что между членами Мстителей происходит серьезное противостояние, имеющее массу проблемных последствий в будущем.

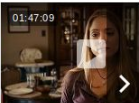
### Больше видео по запросу: Мстители



Мстители Гримм (2015)  
[ok.ru](#) 23.01.16



Горные мстители (2009)  
[ok.ru](#) 02.05.17



Мстители (2007) комедия  
[ok.ru](#) 28.04.16

## Мстители (2012) смотреть онлайн бесплатно в хорошем hd 720 качестве

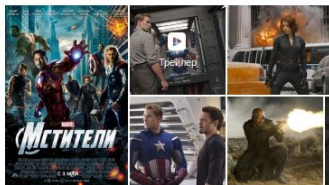
[kinogid.me > Фильмы 2012](#)

Жанр: Боевики, Приключения, Фантастика, Фэнтези. Мстители 2: Эра Альтрона. Безумный Макс 3: Дорога ярости. Тор 2: Царство тьмы

## Мстители — трейлеры, даты премьер — КиноПоиск

[kinoportok.ru/film/11175](#)

Для того чтобы добавить рецензию на фильм Мстители, необходимо. Звездные войны:



## Мстители

The Avengers, США, 2012, 12+

★ Кино Mail Ru 7,9 ★ КиноПоиск 7,9 ★ IMDb 8,1

Фильм Marvel «Мстители» является продолжением эпическ берущих начало в фильмах «Железный человек», «Неверо «Железный человек 2», «Тор» и «Первый мститель». В про супергерои объединяются в уникальную команду. Когда че. [Читать дальше >](#)

Жанр: фэнтези, боевик, приключения, фантастика

Режиссёр: Джосс Уидон

Продюсеры: Стэн Ли, Джон Фавро, Кевин Файги, Алан Фэй Луис Д'Эспозито, +

Длительность: 2 ч. 22 мин.

### Актёры



Роберт  
Дауни мл.

Крис Эванс

Марк Руффало

Хемсворт

### Картинки по запросу космос вселенная



[→ Другие картинки по запросу "космос вселенная"](#)

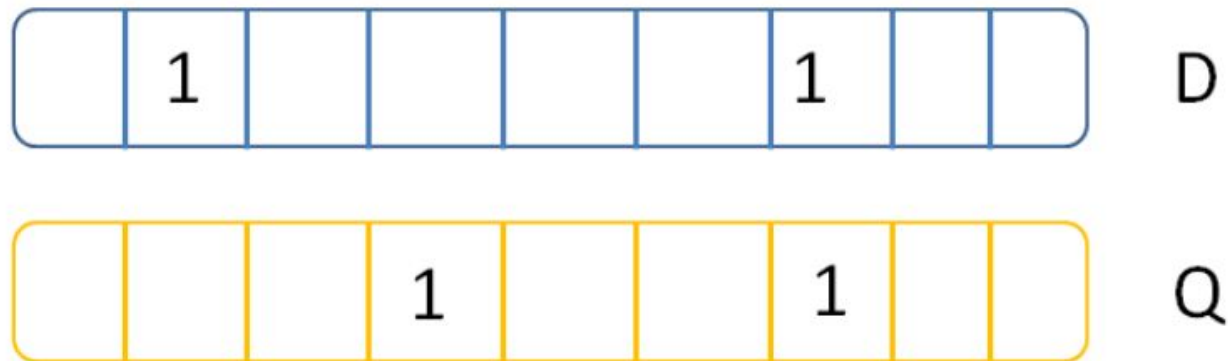
[Пожаловаться на картинку](#)

### Почему космос такой темный, если во Вселенной миллиарды ...

[https://hi-news.ru > Темы > Космос >](#)

1 июн. 2016 г. - Если Вселенная безгранична и полна звезд и галактик, почему мы не видим их повсюду, куда бы ни посмотрели? Иногда самые ...

# Vector space ranking model



TfIdf

$$TfIdf = \sum_{w \in q} tf_w * Idf_w$$

# BM25

## BM25

$$BM25 = \sum_{w \in q} \frac{(k_1 + 1)tf_w}{k_1((1 - b) + b\frac{dl}{avdl}) + tf_w} ldf_w$$

## BM25F

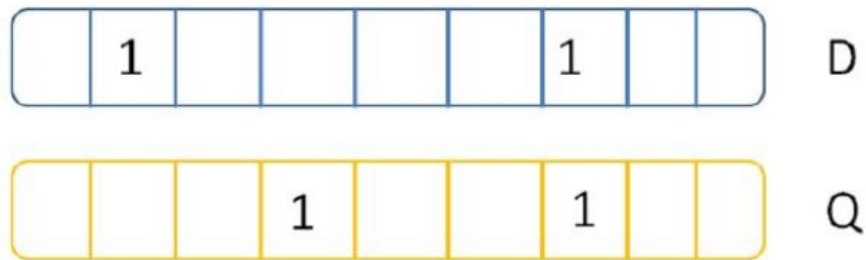
$$BM25F = \sum_{w \in q} \frac{\sum_E rank(E)}{\sum_E rank(E) + k} ldf_w$$

# Классические модели

## Недостатки

- ✘ Соответствие происходит по полному совпадению слов из запроса в документе
- ✘ Не понятно как определить важность слов
- ✘ Не понятно как учитывает порядок слов и контекст, в котором они использованы
- ✘ Приходится думать на лингвистикой (морфология, синонимы, транслиты и т.д)

# Векторная модель



$$\text{Sim}(Q,D) = Q^T D = \sum_{w \in q} tf d_w \text{imp}_w$$

Обучаем важность слов на кликах

# Векторная модель

## deep learning

W [Deep learning — Wikipedia](#)

[en.wikipedia.org/wiki/Deep\\_learning](https://en.wikipedia.org/wiki/Deep_learning)

**Deep learning** (also known as **deep structured learning** or **hierarchical learning**) is part of a broader family of machine learning methods based on learning data representations, as opposed to task-specific algorithms. Learning can be supervised, semi-supervised or unsupervised. Deep learning models...

W [Глубокое обучение — Википедия](#)

[ru.wikipedia.org/wiki/Глубокое\\_обучение](https://ru.wikipedia.org/wiki/Глубокое_обучение)

Глубокое обучение — совокупность методов машинного обучения (с учителем, с частичным привлечением учителя, без учителя, с подкреплением), основанных на обучении представлениям, а не на специализированных алгоритмах, разработанных для конкретных задач. Многие методы глубокого обучения были...

[История](#) [Определения](#) [Содержание](#)

$$\text{Sim}(Q,D) = Q^T D = \sum_{w \in q} tf d_w idf_w$$

Надо подправить модель так, чтобы смошь выучить такие зависимости из данных

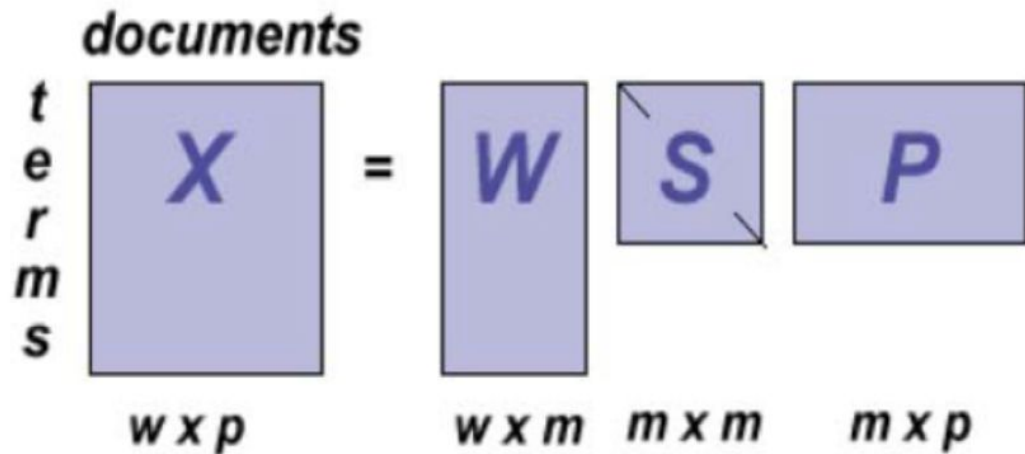
# Векторная модель

Идея: Будем рассматривать не только пересечение слов запроса и документа, но все пары слов из запроса и документа



$$\text{Sim}(Q,D) = \sum_{q \in Q} \sum_{w \in D} tf d_{qw} idf_{qw}$$

# Latent Semantic Analysis



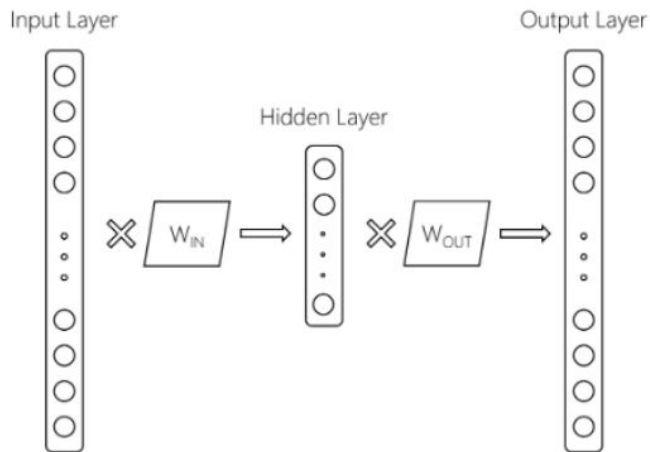
$$\hat{D} = A^T D, \quad \hat{Q} = A^T Q,$$

$$A = W_k S^{-1}$$

$$\text{sim}(Q, D) = \frac{\hat{Q} \hat{D}}{\|Q\| \|D\|}$$



# Dual Embedding Space Model (DESM)



- ▶ Word2vec optimizes IN-OUT dot product which captures co-occurrence statistics of words from the training corpus
- ▶ We can gain by using these two embeddings differently

# Dual Embedding Space Model (DESM)

<b>yale</b>			<b>seahawks</b>			<b>eminem</b>		
IN-IN	OUT-OUT	IN-OUT	IN-IN	OUT-OUT	IN-OUT	IN-IN	OUT-OUT	IN-OUT
yale	yale	yale	seahawks	seahawks	seahawks	eminem	eminem	eminem
harvard	uconn	faculty	49ers	broncos	highlights	rihanna	rihanna	rap
nyu	harvard	alumni	broncos	49ers	jerseys	ludacris	dre	featuring
cornell	tulane	orientation	packers	nfl	tshirts	kanye	kanye	tracklist
tulane	nyu	haven	nfl	packers	seattle	beyonce	beyonce	diss
tufts	tufts	graduate	steelers	steelers	hats	2pac	tupac	performs

- ▶ IN-IN and OUT-OUT cosine similarities are high for words that are similar by function or type
- ▶ IN-OUT cosine similarities are high between words that often co-occur in the same query or document

# Dual Embedding Space Model (DESM)

- ▶ A document is represented by the centroid of its word OUT\_vectors:

$$\vec{v}_{d,OUT} = \frac{1}{|d|} \sum_{t_d \in d} \frac{\vec{v}_{t_d,OUT}}{\|\vec{v}_{t_d,OUT}\|}$$

- ▶ Query-document similarity is average of cosine similarity over query words:

$$\text{DESM}_{\text{IN-OUT}}(q, d) = \frac{1}{q} \sum_{t_q \in q} \frac{\vec{v}_{t_q,IN}^\top \vec{v}_{t_d,OUT}}{\|\vec{v}_{t_q,IN}\| \|\vec{v}_{t_d,OUT}\|}$$

- ▶ IN-OUT captures more Topical notion of similarity than IN-IN and OUT-OUT.
- ▶ DESM is effective at, but only at, ranking at least somewhat relevant documents.

# Dual Embedding Space Model (DESM)

*Albuquerque* is the most populous city in the U.S. state of New Mexico. The high-altitude city serves as the county seat of Bernalillo County, and it is situated in the central part of the state, straddling the Rio Grande. The city population is 557,169 as of the July 1, 2014, population estimate from the United States Census Bureau, and ranks as the 32nd-largest city in the U.S. The Metropolitan Statistical Area (or MSA) has a population of 902,797 according to the United States Census Bureau's most recently available estimate for July 1, 2013.

(a)

Allen suggested that they could program a BASIC interpreter for the device; after a call from Gates claiming to have a working interpreter, MITS requested a demonstration. Since they didn't actually have one, Allen worked on a simulator for the Altair while Gates developed the interpreter. Although they developed the interpreter on a simulator and not the actual device, the interpreter worked flawlessly when they demonstrated the interpreter to MITS in Albuquerque, New Mexico in March 1975; MITS agreed to distribute it, marketing it as Altair BASIC.

(b)

**Figure 3: Two different passages from Wikipedia that mentions "Albuquerque" (highlighted in orange) exactly once. Highlighted in green are all the words that have an IN-OUT similarity score with the word "Albuquerque" above a fixed threshold (we choose -0.03 for this visualization) and can be considered as providing supporting evidence that (a) is *about* Albuquerque, whereas (b) happens to only *mention* the city.**

# Lexical and Semantic matching

Query: united states president

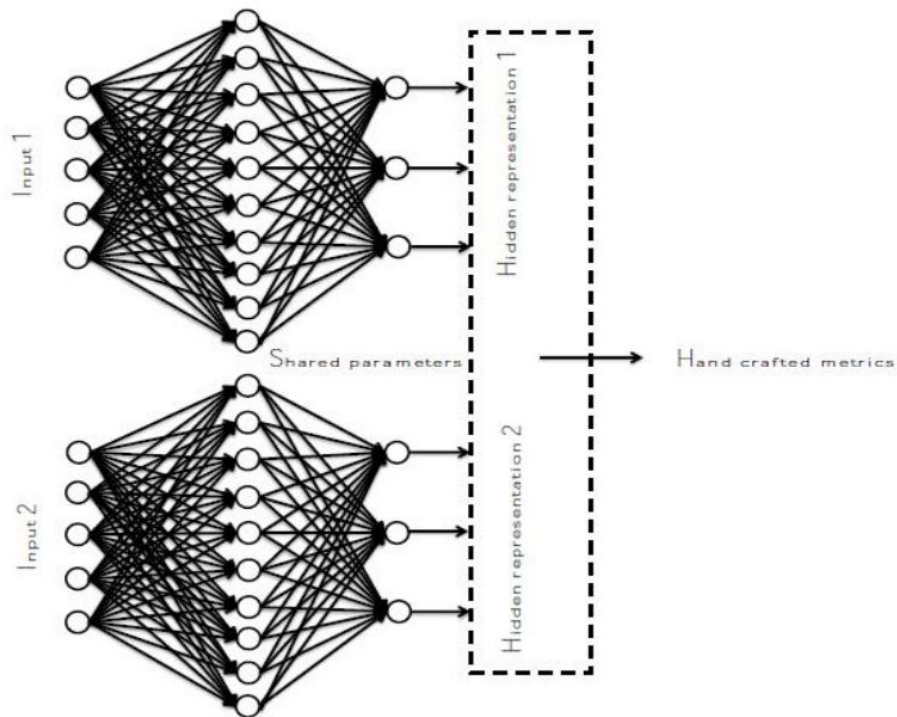
The **President** of the **United States** of America (POTUS) is the elected head of state and head of government of the **United States**. The **president** leads the executive branch of the federal government and is the commander in chief of the **United States Armed Forces**. Barack Hussein Obama II (born August 4, 1961) is an American politician who is the 44th and current **President** of the United States. He is the first African American to hold the office and the first **president** born outside the continental **United States**.

The President of the **United States** of America (POTUS) is the elected head of state and head of government of the **United States**. The **president** leads the executive branch of the federal government and is the commander in chief of the **United States Armed Forces**. **Barack Hussein Obama II** (born August 4, 1961) is an American politician who is the 44th and current **President** of the **United States**. He is the first African American to hold the office and the first **president** born outside the continental **United States**.

- Традиционные модели инфопоиска работают исключительно на лексическом матчинге
- Representation модели могут учитывать как все термины документа соотносятся с запросом
- И те и другие могут быть смоделированы с помощью нейросетей

# Инструмент

## Siamese Neural Network





# Deep Structured Semantic Model (DSSM) (2013)

Posterior probability  
computed by softmax

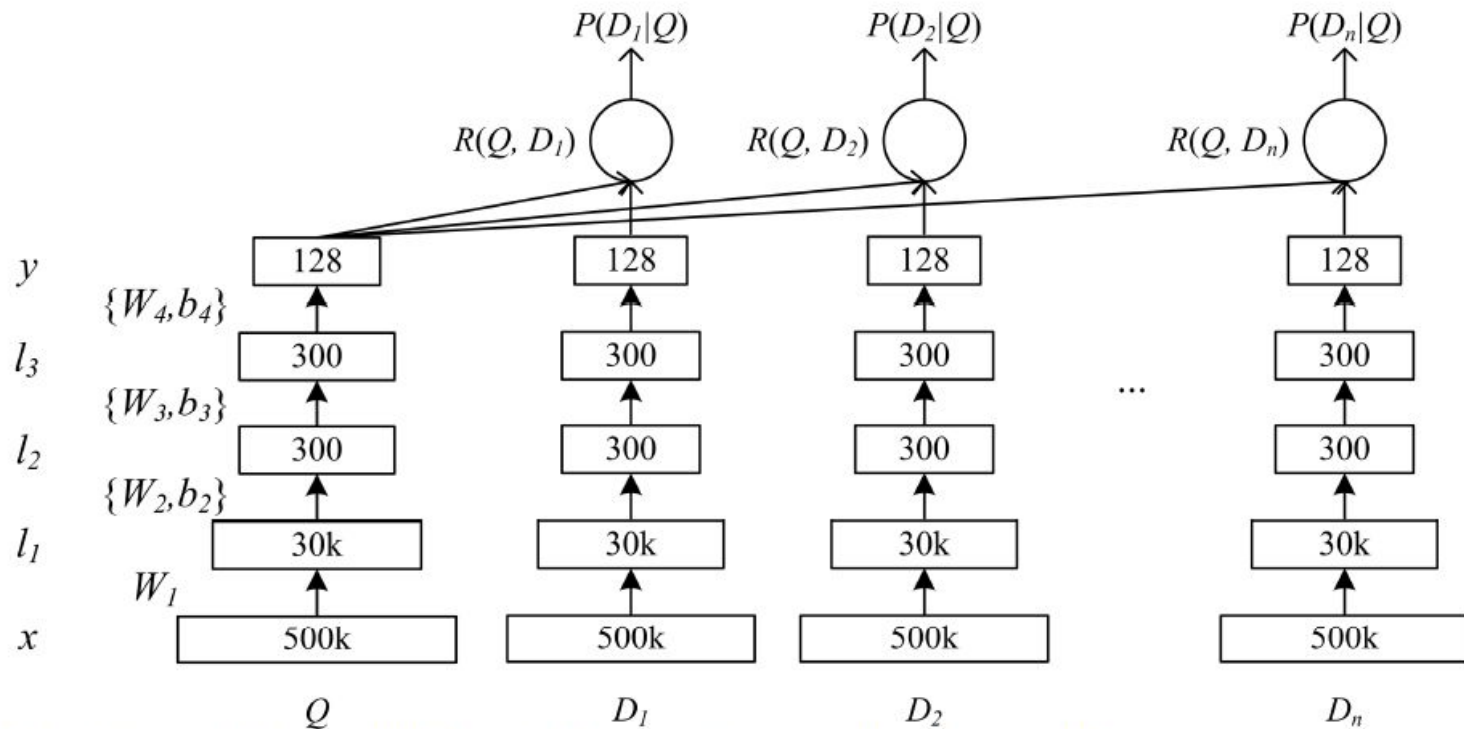
Relevance measured  
by cosine similarity

Semantic feature

Multi-layer non-  
linear projection

Word Hashing

Term Vector



**Figure 1:** Illustration of the DSSM. It uses a DNN to map high-dimensional sparse text features into low-dimensional dense features in a semantic space. The first hidden layer, with 30k units, accomplishes word hashing. The word-hashed features are then projected through multiple layers of non-linear projections. The final layer's neural activities in this DNN form the feature in the semantic space.

# Deep Structured Semantic Model (DSSM) (2013)

## Collisions

Word Size	Letter-Bigram		Letter-Trigram	
	Token Size	Collision	Token Size	Collision
40k	1107	18	10306	2
500k	1607	1192	30621	22

**Table 1:** Word hashing token size and collision numbers as a function of the vocabulary size and the type of letter ngrams.

## Learning

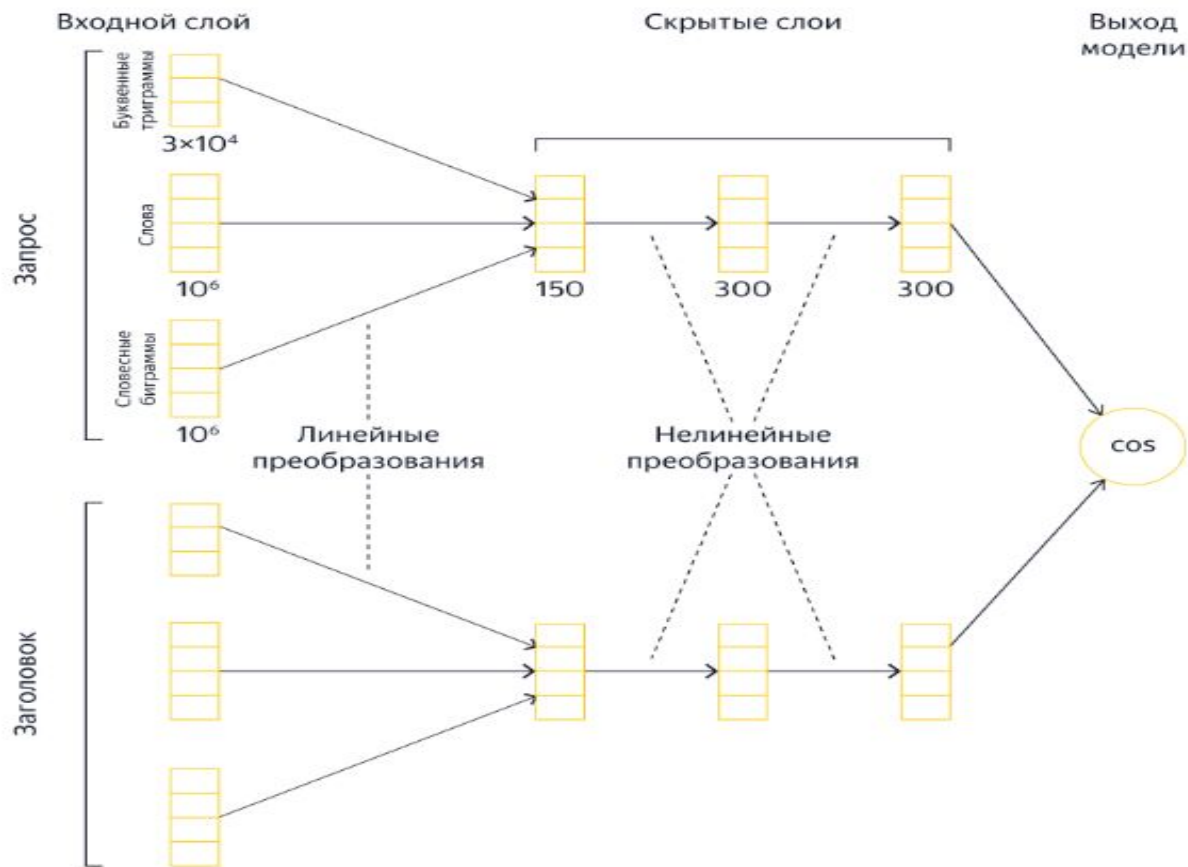
$$R(Q, D) = \frac{y_Q^T y_D}{\|y_Q\| \|y_D\|}$$

$$P(D|Q) = \frac{\exp(R(Q, D))}{\sum_{D' \in \mathbf{D}} \exp(R(Q, D'))}$$

$$L = -\log \prod_{Q, D^+} P(D^+|Q) \rightarrow \min$$



# DSSM в яндексе



# DSSM в яндексе

Запрос: [келлская книга]

Заголовок страницы	BM25	Нейронная модель
келлская книга википедия	0.91	0.92
ученые исследуют келлскую книгу вокруг света	0.88	0.85
book of kells wikipedia	0	0.81
ирландские иллюстрированные евангелия vii viii вв	0	0.58
икеа гипермаркеты товаров для дома и офиса ikea	0	0.09

# DSSM в яндексе

Запрос: [евангелие из келлса]

Заголовок страницы	BM25	Нейронная модель
келлская книга википедия	0	0.85
ученые исследуют келлскую книгу вокруг света	0	0.78
book of kells wikipedia	0	0.71
ирландские иллюстрированные евангелия vii viii вв	0.33	0.84
икеа гипермаркеты товаров для дома и офиса икеа	0	0.10

# DSSM в яндексе

Запрос: [рассказ в котором раздавили бабочку]

Заголовок страницы	BM25	Нейронная модель
фильм в котором раздавили бабочку	0.79	0.82
и грянул гром википедия	0	0.43
брэдбери рэй википедия	0	0.27
машина времени роман википедия	0	0.24
домашнее малиновое варенье рецепт заготовки на зиму	0	0.06

# Interaction matrix based approaches

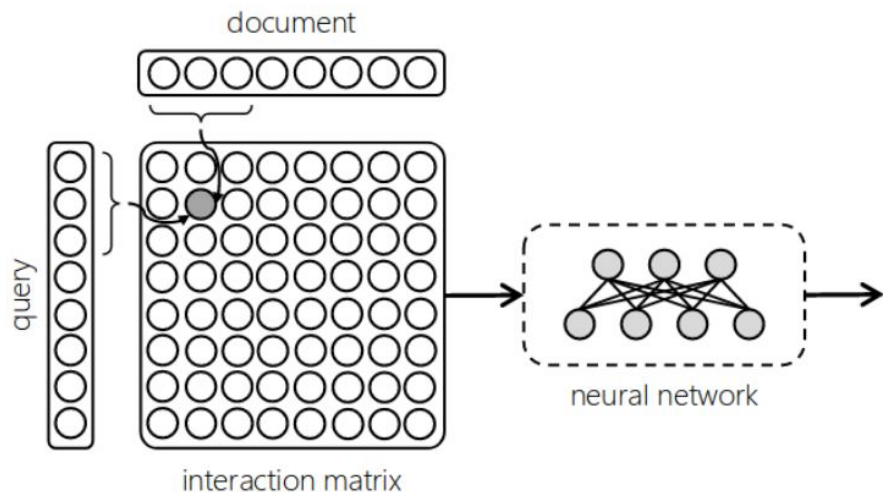
Alternative to Siamese networks

Interaction matrix  $X$ , where  $x_{i,j}$  is obtained by comparing the the  $i^{\text{th}}$  word in source sentence with  $j^{\text{th}}$  word in target sentence

Comparisons can be both lexical or semantic

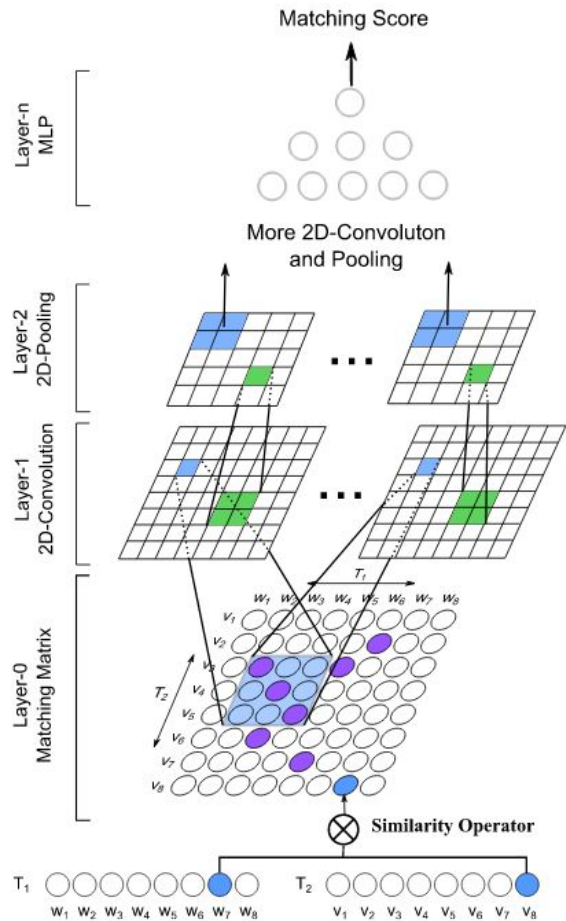
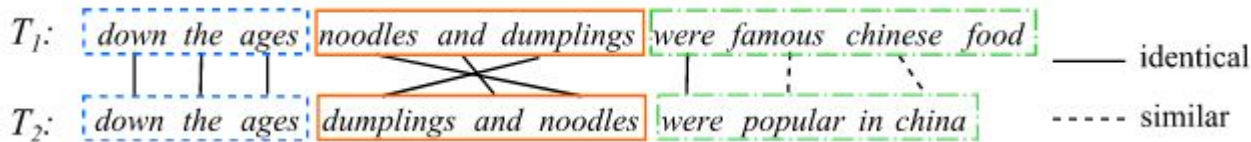
Вопрос:

- ▶ Как такое использовать с линейными моделями?



# MatchPyramid

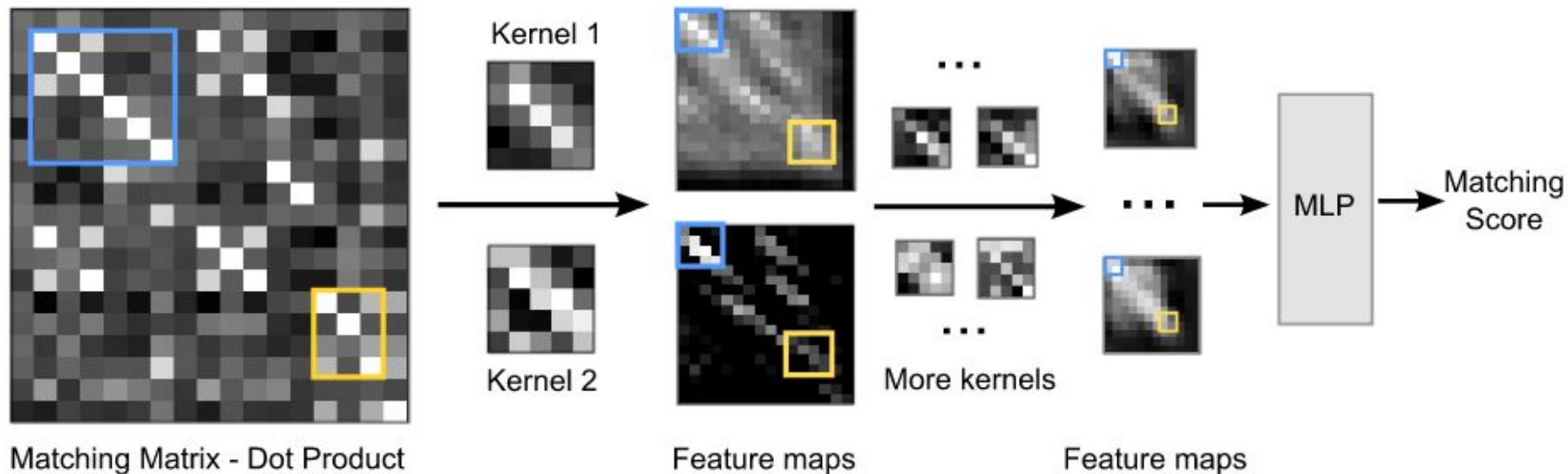
- Сверточные слои поверх interaction matrix между запросом и документом
- После конволюций полносвязные слои



# MatchPyramid

T<sub>1</sub>: PCCW's chief operating officer, Mike Butcher, and Alex Arena, the chief financial officer, will report directly to Mr So.

T<sub>2</sub>: Current Chief Operating Officer Mike Butcher and Group Chief Financial Officer Alex Arena will report to So.



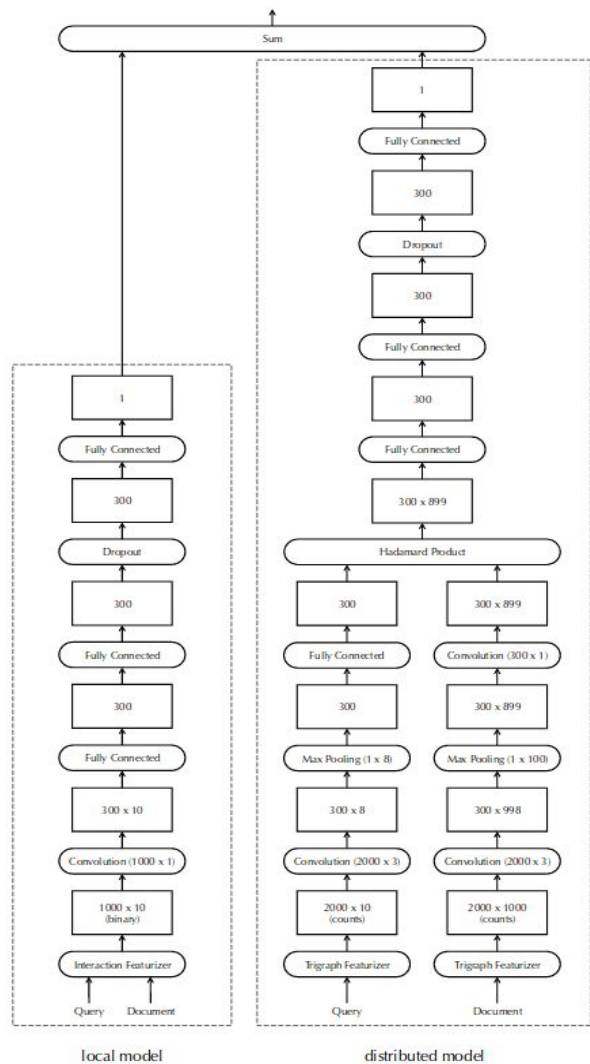
# Duet Model

- Совместно тренируем 2 подсетки, сфокусированные на лексическом и семантическом матчинге

Training sample:

$q, d^+, d_1, d_2, d_3, d_4$

$$p(d^+ | q) = \frac{e^{nr\text{dm}(q, d^+)}}{\sum_{d \in D^-} e^{nr\text{dm}(q, d)}}$$





# Duet Model

Term  
importance

Local model

Only query terms have an impact

Earlier occurrences have bigger impact

Query: united states president

The **President** of the **United States** of America (POTUS) is the elected head of state and head of government of the **United States**. The **president** leads the executive branch of the federal government and is the commander in chief of the **United States** Armed Forces. Barack Hussein Obama II (born August 4, 1961) is an American politician who is the 44th and current **President** of the United States. He is the first African American to hold the office and the first **president** born outside the continental **United States**.

Visualizing impact of dropping terms on model score

# Duet Model

## Term importance

### Distributed model

Non-query terms (e.g., *Obama* and *federal*) has positive impact on score

Common words like 'the' and 'of' probably good indicators of well-formedness of content

### Query: united states president

The President of the United States of America (POTUS) is the elected head of state and head of government of the United States. The president leads the executive branch of the federal government and is the commander in chief of the United States Armed Forces. Barack Hussein Obama II (born August 4, 1961) is an American politician who is the 44th and current President of the United States. He is the first African American to hold the office and the first president born outside the continental United States.

Visualizing impact of dropping terms on model score

# Duet Model (Результаты)

	NDCG@1	NDCG@10
<b>Non-neural baselines</b>		
LSA	22.4	44.2
BM25	24.2	45.5
DM	24.7	46.2
QL	24.6	46.3
<b>Neural baselines</b>		
DRMM	24.3	45.2
DSSM	25.8	48.2
CDSSM	27.3	48.2
DESM	25.4	48.3
<b>Our models</b>		
Local model	24.6	45.1
Distributed model	28.6	50.5
<b>Duet model</b>	<b>32.2</b>	<b>53.0</b>