



ТЕХНОСФЕРА

Learning to match 1

Владимир Гулин

13 ноября 2019 г.

План

Напоминание

Модели без учителя

Модели основанные на машинном переводе

Модели основанные нейросетях

Задача ранжирования

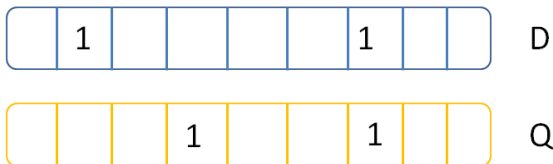
Цель:

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

NDCG

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

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

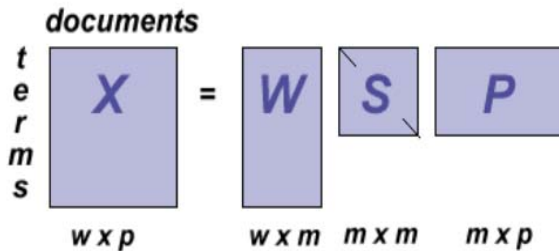
Недостатки?

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)

LSA

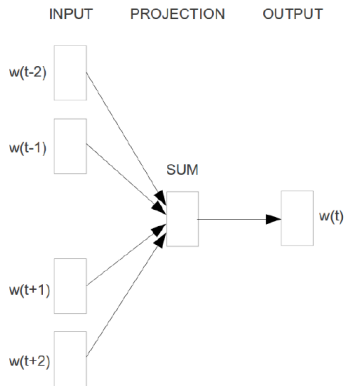


$$\hat{D} = A^T D, \quad \hat{Q} = A^T Q,$$

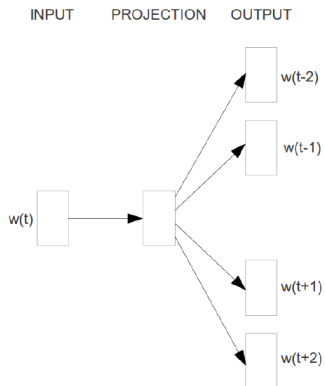
$$A = W_k S^{-1}$$

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

Word2vec



CBOW



Skip-gram

- ▶ Как обучается word2vec?

Word2vec

Недостатки?

Word2vec

Недостатки

- ✗ Качество сильно зависит от обучающих данных, их количества, размера векторов (на вектора большой размерности нужно много вычислительных затрат)
- ✗ Усреднение векторов слов работает плохо уже на текстах среднего размера (не говоря уже о больших)
- ✗ Из word2vec нельзя получить представление фиксированного размера для текстов переменного размера

МОДЕЛИ БЕЗ УЧИТЕЛЯ

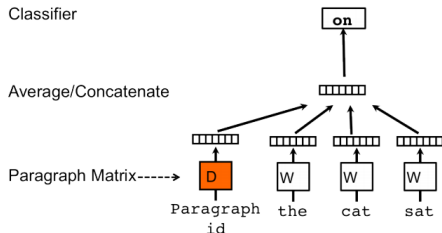
Doc2vec (Mikolov, 2014)

- ▶ Обобщение word2vec модели 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

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

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
- ▶ Но при этом хуже по качеству, чем DM (однако DM + DBOW работают лучше!)

Classifier

the cat sat on



Paragraph Matrix

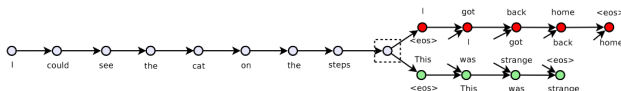


Paragraph
id

Skip-Thought Vectors

► Skip-Thought Vectors

- Conceptually similar to distributional semantics: a unit's representation is a function of its neighbouring units, except units are sentences instead of words.



- Similar to auto-encoding objective: encode sentence, but decode neighboring sentences.
- Pair of LSTM-based seq2seq models with shared encoder.

Skip-Thought Vectors

$x(0)$: Hi, My name is Sanyam

$x(1)$: Today, I went to the zoo.

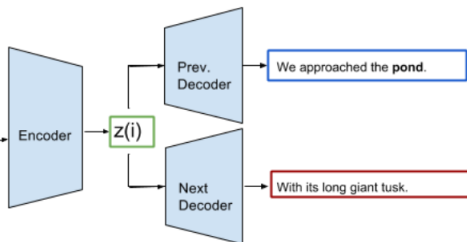
⋮

$x(i-1)$: We approached the tree.

$x(i)$: The elephant was still.

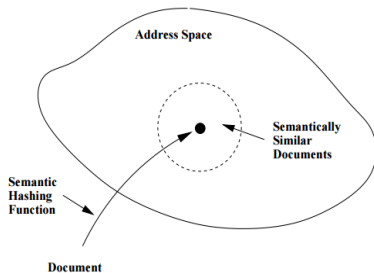
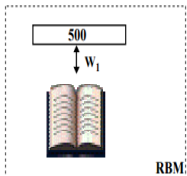
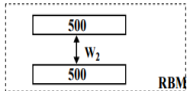
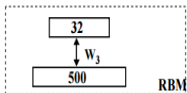
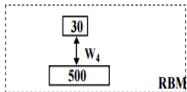
$x(i+1)$: It was taking a nap probably.

⋮



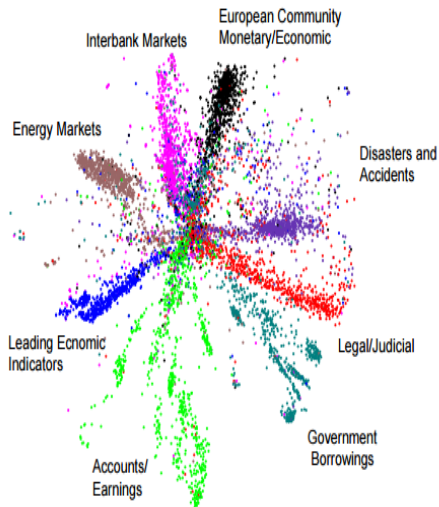
Skip Thoughts model overview

Semantic hashing (Hinton, Salakhutdinov, 2007)

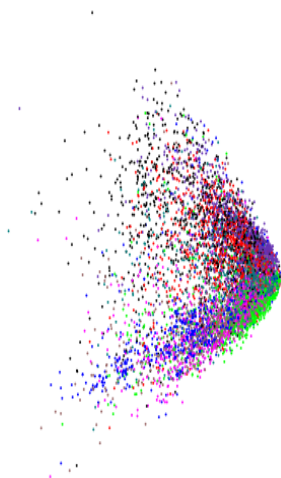


Deep Auto Encoder

Autoencoder 2-D Topic Space



LSA 2-D Topic Space



МОДЕЛИ ОСНОВАННЫЕ НА МАШИННОМ ПЕРЕВОДЕ

О статистическом машинном переводе

Общая схема

- ▶ Находим большой корпус параллельных текстов
- ▶ Выравниваем эти тексты по предложениям
- ▶ Считаем статистику
- ▶ Строим модель перевода

Параллельные тексты

English	German
Diverging opinions about planned tax reform	Unterschiedliche Meinungen zur geplanten Steuerreform
The discussion around the envisaged major tax reform continues .	Die Diskussion um die vorgesehene grosse Steuerreform dauert an .
The FDP economics expert , Graf Lambsdorff , today came out in favor of advancing the enactment of significant parts of the overhaul , currently planned for 1999 .	Der FDP - Wirtschaftsexperte Graf Lambsdorff sprach sich heute dafuer aus , wesentliche Teile der fuer 1999 geplanten Reform vorzuziehen .

Выравнивание предложений

The diagram illustrates the alignment of English and German sentences. It consists of two columns of text, each with a folded top-right corner. Lines connect corresponding words and phrases between the two columns.

English Text (Left Column):

The man looks intently at the window.
He sees a shadow.
It was in the trees.
What was it?
He is alarmed and awake.

He has long lived in the woods.
He likes the isolation and solitude of his house.
It's small, but cozy.
The next village is miles away.
He only goes there once a week.

It just after dusk.
The hot sun finally set.
The forest was still abuzz in chatter.
Voices of birds and insects fill the air.
A comforting sound.
But the shadow was larger than those animals
Only little creatures live here, not this.
It seemed almost as large as a man.
But why that?
Nobody comes ever here.
So the man's eyes keep looking.

As the minutes passed, nothing happens.
But then, cast against the bright moonlit, it returns.

German Text (Right Column):

Der Mann schaut aus dem Fenster.
Er sieht einen Schatten in den Bäumen.
Was war das?
Er war alarmiert und wach.

Er hat schon lange im Wald gelebt.
Er genießt die Einsamkeit des Hauses
Es ist klein.
Aber es ist gemütlich.
Das nächste Dorf ist meilenweit entfernt.
Er geht dorthin nur einmal in Monat.

Es ist nach der Untergang der heißen Sonne.
Der Wald ist voller Geschwätz.
Stimmen von Vögeln und Insekten dringen herüber.

Aber der Schatten war größer als diese Tiere.
Nur Kleintier lebt hier.
Nicht soetwas Großes.
Es erschien fast so groß wie ein Mensch.
Aber warum, wenn hier niemand jemals herkommt?
Der Mann schaut.
Sein Augen aus dem Fenster gerichtet.

Minuten vergehen, aber nichts passiert.
Dann plötzlich kehrt er im Mondschein zurück.

Word by word

Haus — *house, building, home, household, shell.*

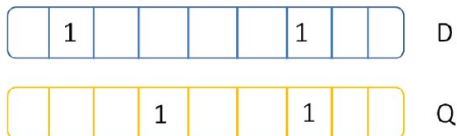
Translation of <i>Haus</i>	Count
<i>house</i>	8000
<i>building</i>	1600
<i>home</i>	200
<i>household</i>	150
<i>shell</i>	50

$$p_f(e) = \begin{cases} 0.8 & \text{if } e = \textit{house} \\ 0.16 & \text{if } e = \textit{building} \\ 0.02 & \text{if } e = \textit{home} \\ 0.015 & \text{if } e = \textit{household} \\ 0.005 & \text{if } e = \textit{shell} \end{cases}$$

Вопрос:

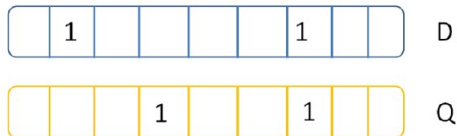
- ▶ Как такое применить к поиску?

Векторная модель



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

Векторная модель

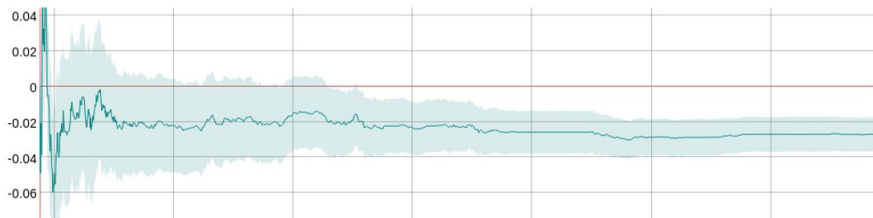


$$\text{Sim}(Q,D) = Q^T D = \sum_{w \in q} tf d_w \textcolor{red}{imp}_w$$

Обучаем важность слов на кликах

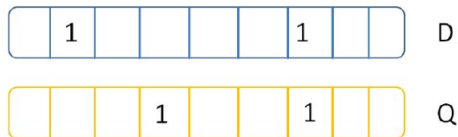
Внедрение

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



+3%


Векторная модель





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


Недостаток: Рассматриваем только пересечение слов запроса и документа


Векторная модель

 **Deep Learning | ВКонтакте**
vk.com/deeplearning
Глубокое обучение (**Deep learning**) - это направление в области Искусственного Интеллекта (Artificial Intelligence) и Машинного Обучения (Machine...

 **Deep Learning Tutorials — DeepLearning 0.1 documentation**
deeplearning.net/tutorial
Deep Learning is a new area of Machine **Learning** research, which has been introduced with the objective of moving Machine **Learning** closer to one of its...

 **Глубокое обучение — Википедия**
ru.wikipedia.org/wiki/Глубокое_...
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](#)

 **Что такое deep learning и как эти компьютерные алгоритмы перевернут...**
ideanomics.ru/?p=1849
«**Deep learning** сейчас очень высоко ценится, — говорит Йошуа Бенгио, профессор факультета компьютерных наук Монреальского университета, одного из...

 **Deep Learning in Python**
datacamp.com/courses/deep-...
He has contributed to the Keras and Tensorflow libraries for **deep learning**, finished 2nd (out of 1353 teams) in \$3million Heritage Health Prize data...

28 / 53

Векторная модель

deep learning

[w Deep learning — Wikipedia](#)
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

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](#)

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

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

deep learning



глубокое обучение - википедия

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

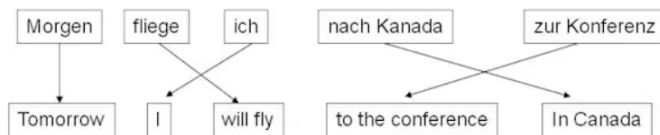
Word-based translation model

$$P(Q|D) = \prod_{q \in Q} \sum_{w \in D} P(q|w)P(w|D)$$

$P(w|D)$ - unigram probability of word w in D

$P(q|w)$ - probability of translating w into query term q

Phrase Based Machine Translation



- Foreign input segmented in to phrases
 - “phrase” is any sequence of words
- Each phrase is probabilistically translated into English
 - $P(\text{to the conference} \mid \text{zur Konferenz})$
 - $P(\text{into the meeting} \mid \text{zur Konferenz})$
- Phrases are probabilistically re-ordered

Результаты

#	Models	NDCG@1	NDCG@3	NDCG@10
1	UM	0.308	0.373	0.454
2	PLSA ($\lambda_2 = 0$)	0.295	0.371	0.456
3	PLSA	0.325	0.391	0.470
4	BLTM ($\lambda_2 = 0$)	0.330	0.399	0.476
5	BLTM	0.338	0.404	0.479
6	BLTM-PR ($\lambda_2 = 0$)	0.334	0.403	0.479
7	BLTM-PR	0.342	0.406	0.482
8	BLTM-PR-1V ($\lambda_2 = 0$)	0.337	0.403	0.480
9	BLTM-PR-1V	0.344	0.407	0.483
10	WTM M1 ($\lambda_2 = 0$)	0.332	0.400	0.478
11	WTM M1	0.338	0.404	0.480

Table 1: Web document ranking results using different topic models, tested on the evaluation data set, where only the title field of each document is used.

МОДЕЛИ ОСНОВАННЫЕ НА НЕЙРОСЕТЯХ

Чего мы хотели добиться, когда учили embedding?

МОДЕЛИ ОСНОВАННЫЕ НА НЕЙРОСЕТЯХ

Чего мы хотели добиться, когда учили embedding?

- ▶ Хотели научиться вытаскивать семантику

В каких терминах мы это хотели сделать?

МОДЕЛИ ОСНОВАННЫЕ НА НЕЙРОСЕТЯХ

Чего мы хотели добиться, когда учили embedding?

- ▶ Хотели научиться вытаскивать семантику

В каких терминах мы это хотели сделать?

- ▶ В терминах косинусного расстояния

Идея:

- ▶ А давайте учить такой embedding нейросетью)

Lexical and Semantic matching

Query: united states president

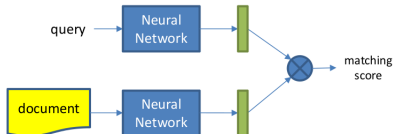
The **President** of the **United States** of America (POTUS) is the elected head of state and head of government of the **United States**. The **president** leads the executive branch of the federal government and is the commander in chief of the United States Armed Forces. Barack Hussein Obama II (born 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.

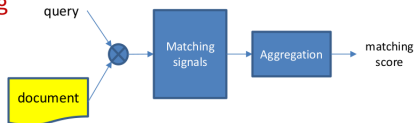
- ▶ Традиционные модели инфопоиска работают исключительно на лексическом матчинге
- ▶ Representation модели могут учитывать как все термины документа соотносятся с запросом
- ▶ И те и другие могут быть смоделированы с помощью нейросетей

Основные подходы

- Representation learning:
representing queries and
document in
semantic space

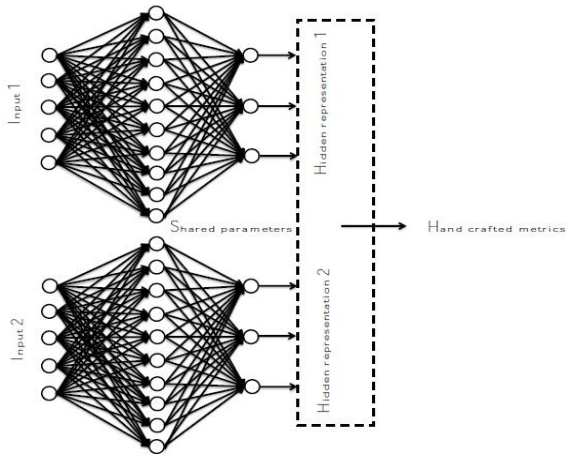


- Matching function learning:
discovering and aggregating
the query-document
matching patterns



Инструмент

Siamese Neural Network



Discriminative projection model (DPM) (GAO 2010)

Идея:

- ▶ У нас есть кликовые данные. Будем считать клики хорошими примерами
- ▶ Неклики будем считать негативными примерами
- ▶ И будем учить линейный embedding на pairwise loss

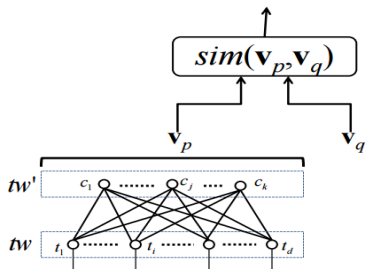
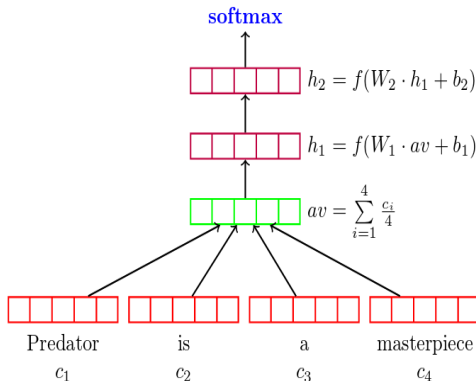


Figure 1: Learning concept vectors. The output layer consists of a small number of concept nodes, where the weight of each node is a linear combination of all the original term weights.

Deep Averaging Network (DAN) 2014

Идея:

- ▶ У нас есть обученные эмбединги
- ▶ Давайте научимся их усреднять



Deep Structured Semantic Model (DSSM) (GAO 2013)

Posterior probability
computed by softmax

Relevance measured
by cosine similarity

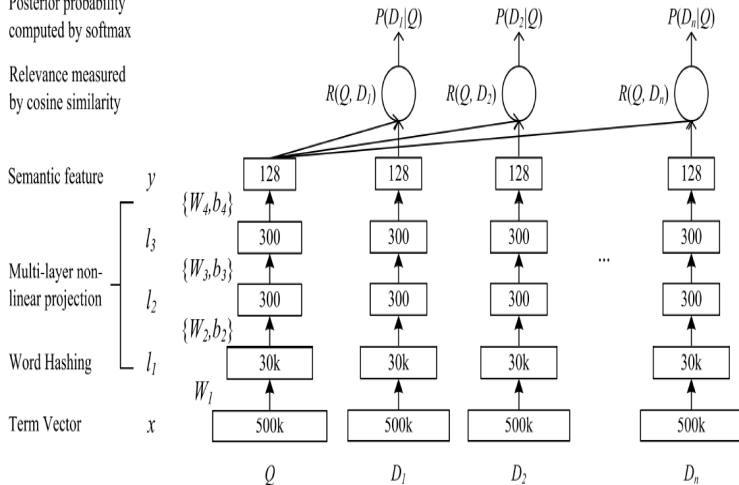


Figure 1: Illustration of the DSSM. It uses a DNN to map high-dimensional sparse text features into low-dimensional dense features in a semantic space. The first hidden layer, with 30k units, accomplishes word hashing. The word-hashed features are then projected through multiple layers of non-linear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

DSSM

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.

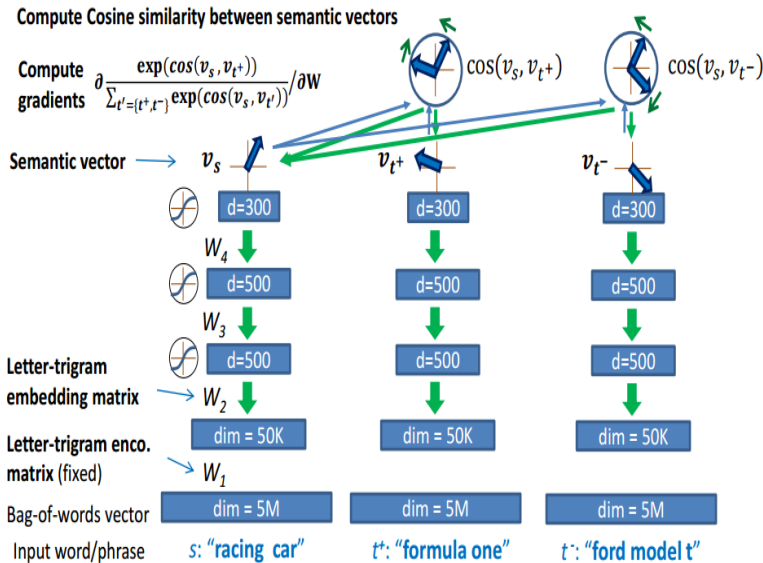
Learning

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

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

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

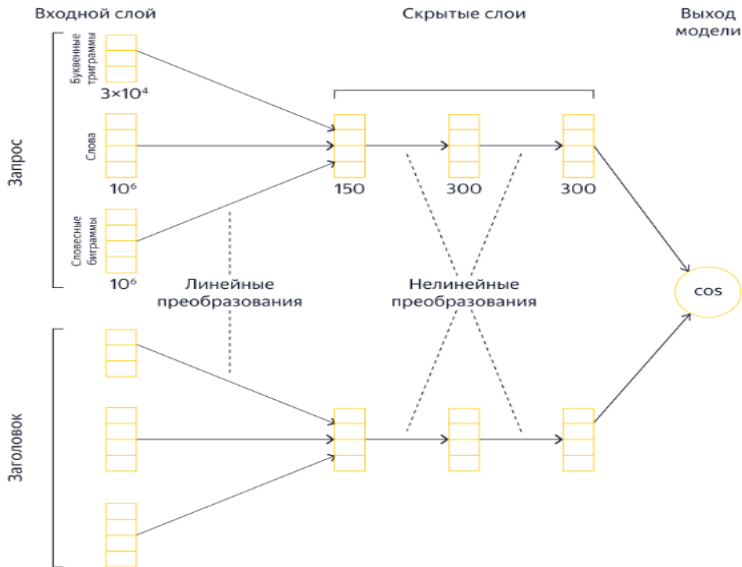
DSSM



#	Models	NDCG@1	NDCG@3	NDCG@10
1	TF-IDF	0.319	0.382	0.462
2	BM25	0.308	0.373	0.455
3	WTM	0.332	0.400	0.478
4	LSA	0.298	0.372	0.455
5	PLSA	0.295	0.371	0.456
6	DAE	0.310	0.377	0.459
7	BLTM-PR	0.337	0.403	0.480
8	DPM	0.329	0.401	0.479
9	DNN	0.342	0.410	0.486
10	L-WH linear	0.357	0.422	0.495
11	L-WH non-linear	0.357	0.421	0.494
12	L-WH DNN	0.362	0.425	0.498

Table 2: Comparative results with the previous state of the art approaches and various settings of DSSM.

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 ikea	0	0.10

DSSM в яндексе

Запрос: [рассказ в котором раздавили бабочку]

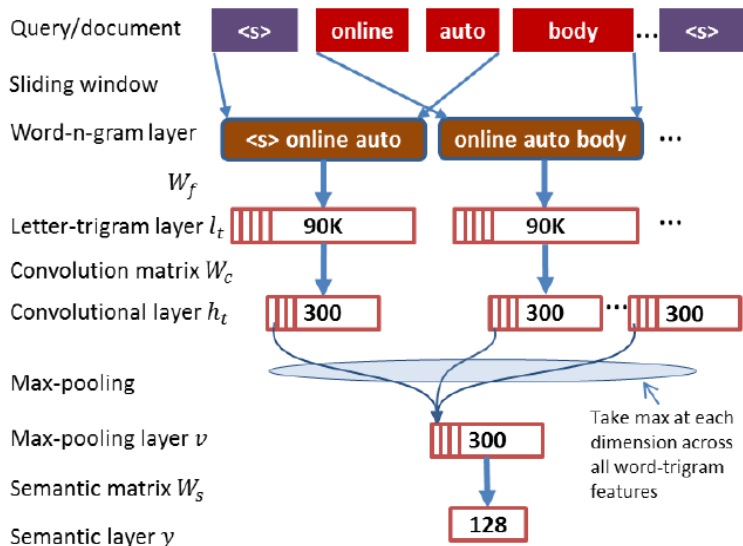
Заголовок страницы	BM25	Нейронная модель
фильм в котором раздавили бабочку	0.79	0.82
и грянул гром википедия	0	0.43
брэдбери рэй википедия	0	0.27
машина времени роман википедия	0	0.24
домашнее малиновое варенье рецепт заготовки на зиму	0	0.06

DSSM недостатки

microsoft *office* excel could allow remote code execution
welcome to the apartment *office*
online *body* fat percentage calculator
online auto *body* repair estimates

Table 1: Sample document titles. The text is lower-cased and punctuation removed. The same word, e.g., “*office*”, has different meanings depending on its contexts.

CLSM



#	Models	NDCG@1	NDCG@3	NDCG@10
1	BM25	0.305	0.328	0.388
2	ULM	0.304	0.327	0.385
3	PLSA (T=100)	0.305	0.335 ^{α}	0.402 ^{α}
4	PLSA (T=500)	0.308	0.337 ^{α}	0.402 ^{α}
5	LDA (T=100)	0.308	0.339 ^{α}	0.403 ^{α}
6	LDA (T=500)	0.310 ^{α}	0.339 ^{α}	0.405 ^{α}
7	BLTM	0.316 ^{α}	0.344 ^{α}	0.410 ^{α}
8	MRF	0.315 ^{α}	0.341 ^{α}	0.409 ^{α}
9	LCE	0.312 ^{α}	0.337 ^{α}	0.407 ^{α}
10	WTM	0.315 ^{α}	0.342 ^{α}	0.411 ^{α}
11	PTM (maxlen = 3)	0.319 ^{α}	0.347 ^{α}	0.413 ^{α}
12	DSSM ($J = 4$)	0.320 ^{α}	0.355 ^{$\alpha\beta$}	0.431 ^{$\alpha\beta$}
13	DSSM ($J = 50$)	0.327 ^{$\alpha\beta$}	0.363 ^{$\alpha\beta$}	0.438 ^{$\alpha\beta$}
14	CLSM ($J = 4$)	0.342 ^{$\alpha\beta\gamma$}	0.374 ^{$\alpha\beta\gamma$}	0.447 ^{$\alpha\beta\gamma$}
15	CLSM ($J = 50$)	0.348^{$\alpha\beta\gamma$}	0.379^{$\alpha\beta\gamma$}	0.449^{$\alpha\beta\gamma$}

Table 5: Comparative results with the previous state of the art approaches. BLTM, WTM, PTM, DSSM, and CLSM use the same clickthrough data described in section 5.1 for learning. Superscripts α , β , and γ indicate statistically significant improvements ($p < 0.05$) over **BM25**, **PTM**, and **DSSM ($J = 50$)**, respectively.

Вопросы

