

# Эмбеддинги NLP

DMIA / Deep Learning / 2019

Савельева Татьяна

@tany\_savelieva

О лекторе

Яндекс Такси



# Цель лекции

- 1) Передать представление о том, как устроена область
- 2) Вложить на концептуальном уровне понятия
- 3) Дать информацию об актуальных моделях, чтобы вы самостоятельно не боялись использовать
- 4) Рассказать, где еще возможно стоит ждать научных прорывов



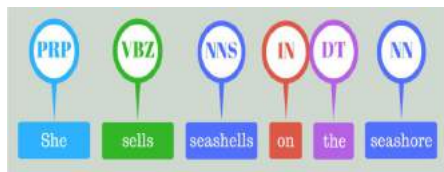
# Какие бывают задачи NLP?

Думаем фундаментально...

# Фундаментальные задачи

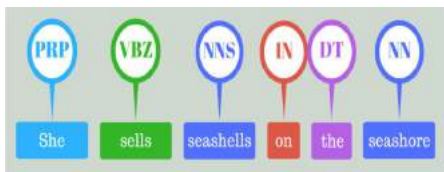


POS



# Фундаментальные задачи

POS



NER

"There was nothing about this storm that was as expected," said **Jeff Masters**, a meteorologist and founder of **Weather Underground**. "**Irma** could have been so much worse. If it had traveled 20 miles north of the coast of **Cuba**, you'd have been looking at a (Category) 5 instead of a (Category) 3."

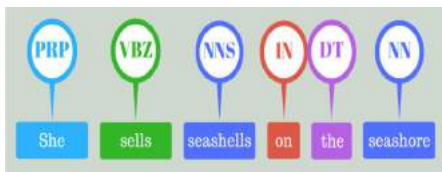
Person

Organization

Location

# Фундаментальные задачи

POS



NER

"There was nothing about this storm that was as expected," said **Jeff Masters**, a meteorologist and founder of **Weather Underground**. "**Irma** could have been so much worse. If it had traveled 20 miles north of the coast of **Cuba**, you'd have been looking at a (Category) 5 instead of a (Category) 3."

Person

Organization

Location

SRL

Semantic Role Labeling

SRL is sometimes also called shallow semantic parsing, is a task in consisting of the detection of the semantic arguments associated with the **predicate** or **verb** of a sentence and their classification into their specific **roles**.

e.g 1. **Mark** **sold** the **car** to **Mary**.  
**agent** **represent** **predicate** **theme** **recipient**

e.g 2. [John]<sub>ARG0</sub> [ate]<sub>REL</sub> [the apple]<sub>ARG1</sub> [in the garden]<sub>ARGM-LOC</sub>



# Фундаментальные задачи

Анализ  
тональности



# Фундаментальные задачи

Анализ  
тональности



Ответы на  
вопросы

## Passage Segment

...The European Parliament and the Council of the European Union have powers of amendment and veto during the legislative process...

## Question

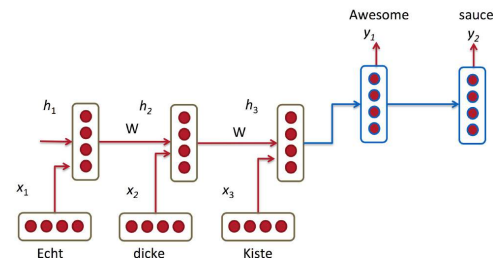
Which governing bodies have veto power?

# Фундаментальные задачи

Анализ  
тональности



Перевод



Ответы на  
вопросы

Passage Segment

...The European Parliament and the Council of the European Union have powers of amendment and veto during the legislative process...

Question

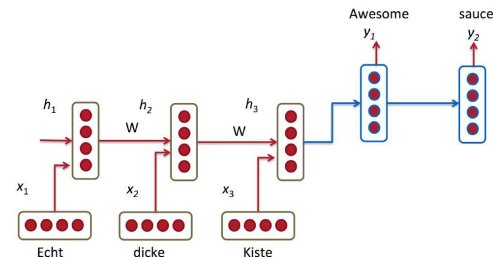
Which governing bodies have veto power?

# Фундаментальные задачи

Анализ  
тональности



Перевод



Ответы на  
вопросы

Passage Segment

...The European Parliament and the Council of the European Union have powers of amendment and veto during the legislative process...

Question

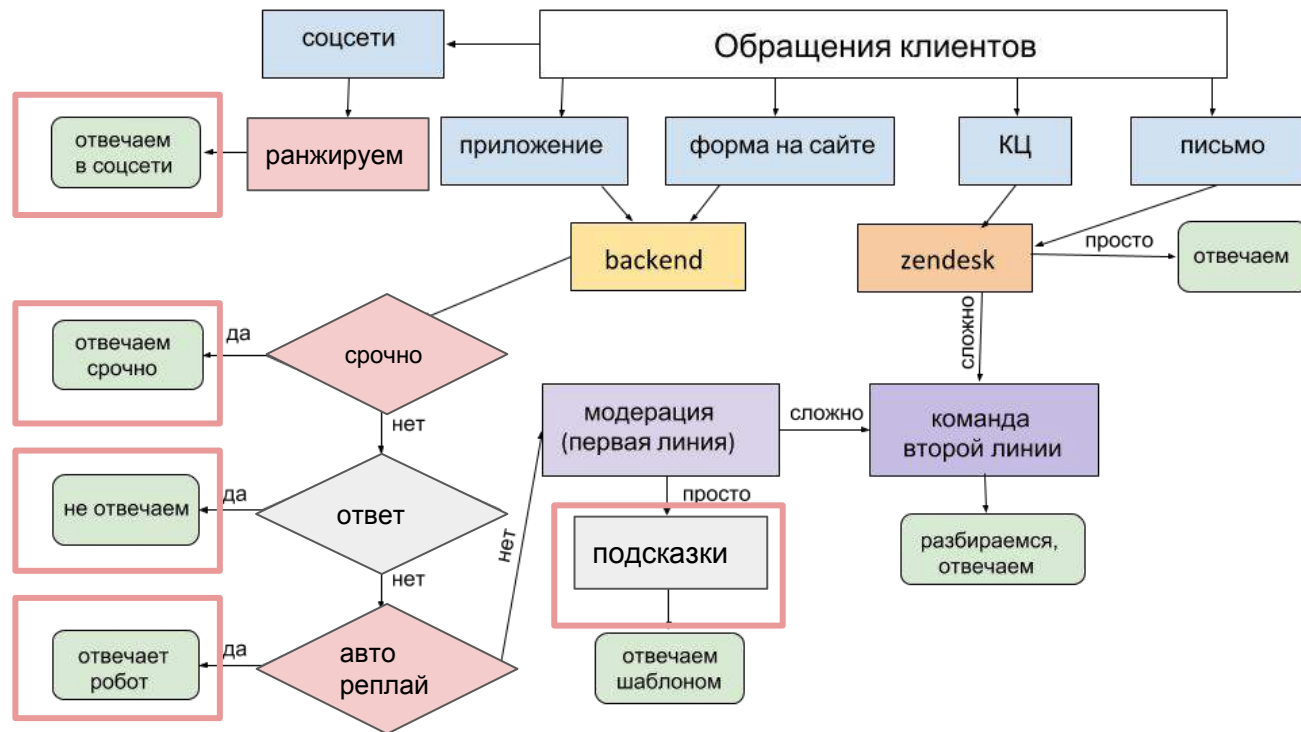
Which governing bodies have veto power?

Где еще используются тексты

# Индустриальные задачи



# Что можно автоматизировать?



# Автореплай

Как это выглядит в реальной жизни

Подъехали с другого входа, хотя в чате писала куда подъехать, ещё и за ожидание взяли. А ждать пришлось, потому что подъехали не туда!!!!

21 ноя, 14:13

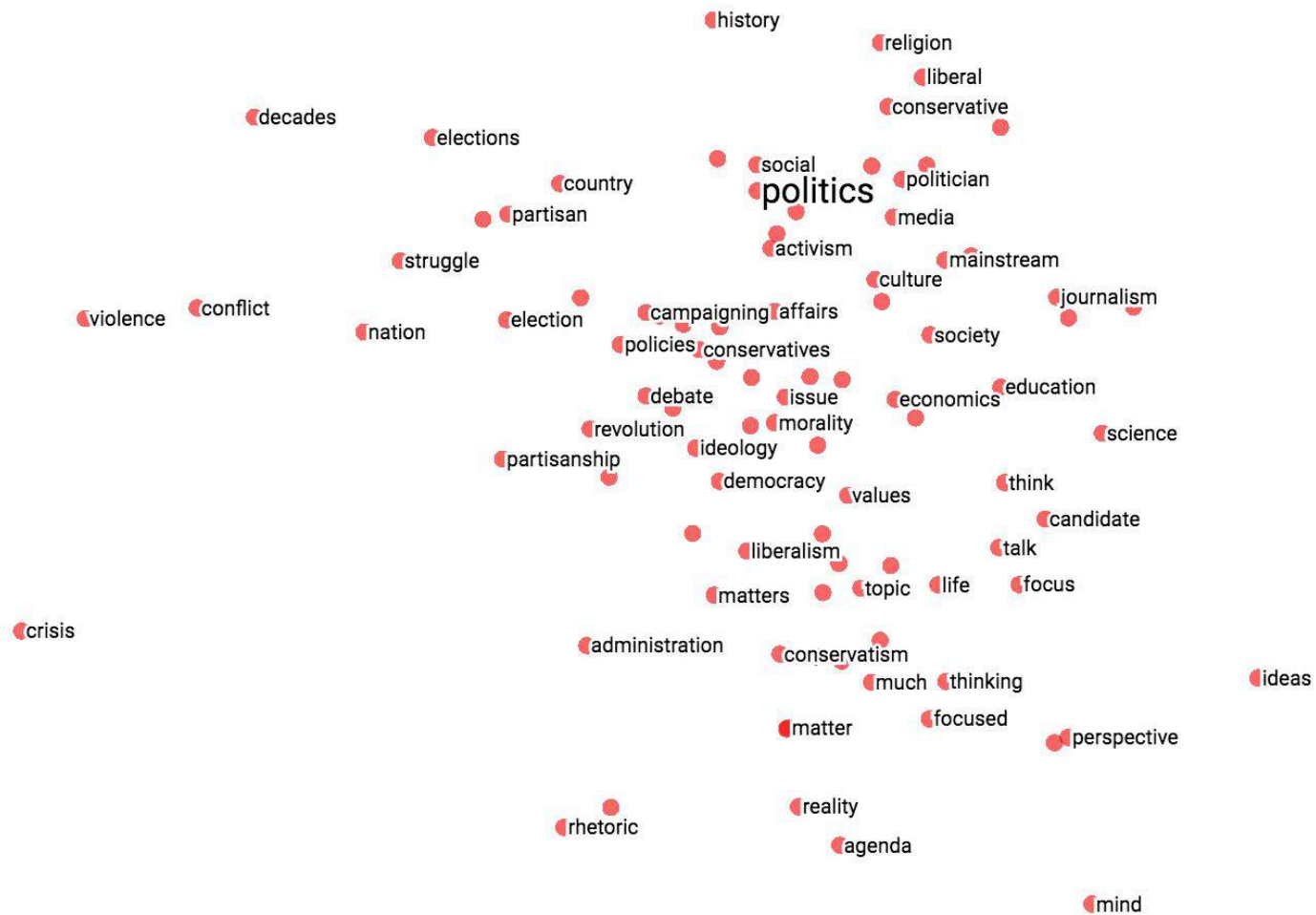
Саппорт

Здравствуйте! Мы закончили проверку по вашему обращению. Пожалуйста, примите наши извинения и промокод regret4749784 со скидкой 50 Р на одну поездку. Его нужно ввести в разделе «Промокод» бокового меню приложения перед оформлением заказа.

21 ноя, 14:23



Как решать задачи NLP?



# Зачем нужно строить вектора для текстов?

- 1) Классификация текстов
- 2) Кластеризация текстов
- 3) Генерация текстов
- 4) На самом деле все задачи NLP...

# Какие вектора можно строить?

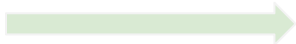
- 1) Вектора символов
- 2) Вектора слогов
- 3) Вектора слов
- 4) Вектора смыслов
- 5) Вектора предложений
- 6) Вектора набора слов
- 7) Вектора текстов

Нужно как-то представить слова в виде векторов ...  
Ваши идеи?

# Нужно как-то представить вектора слов Ваши идеи?

## One hot encoding

Хочу домашку по трендам

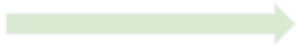


	Хочу	домашку	по	трендам
0	1	0	0	0
1	0	1	0	0
2	0	0	1	0
3	0	0	0	1

# Нужно как-то представить вектора слов Ваши идеи?

## One hot encoding

Хочу домашку по трендам



	Хочу	домашку	по	трендам
0	1	0	0	0
1	0	1	0	0
2	0	0	1	0
3	0	0	0	1

- 1) Не учитывается близость между словами
- 2) Опечатки-ошибки игнорируются
- 3) Слишком много признаков

# Что такое близкие слова?

Скажи мне, кто твой сосед и я скажу тебе, кто ты

*Будем смотреть не только на слова, но и на их контексты!*



# Нужна большая общность!

## Матрица слово-контекст

	и	машинное	обучение	рудн	иностранец
и	1000	40	50	2	3
машинное		100	45	0	0
обучение			150	3	5
рудн				5	10
иностранец					30

# Нужна большая общность!

## Матрица слово-контекст

	и	машинное	обучение	рудн	иностранец
и	1000	40	50	2	3
машинное		100	45	0	0
обучение			150	3	5
рудн				5	10
иностранец					30

Использовать не количество  
а что-то похитрей

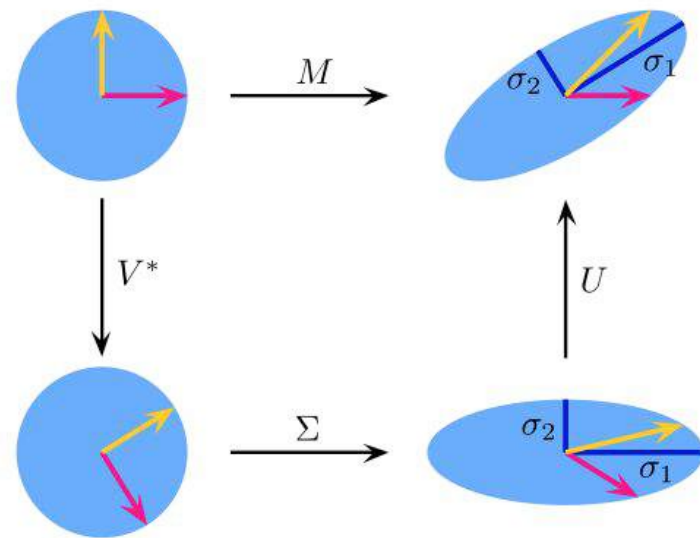
$$\text{PMI}(w_1, w_2) = \log_2 \frac{P(w_1 w_2)}{P(w_1)P(w_2)}$$

# SVD- разложение

$$\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$$

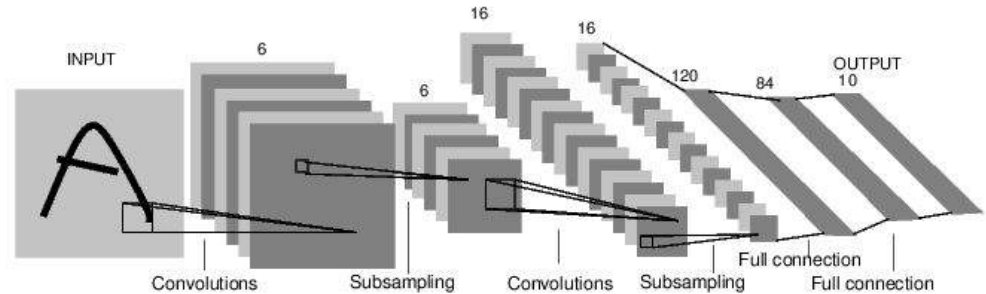
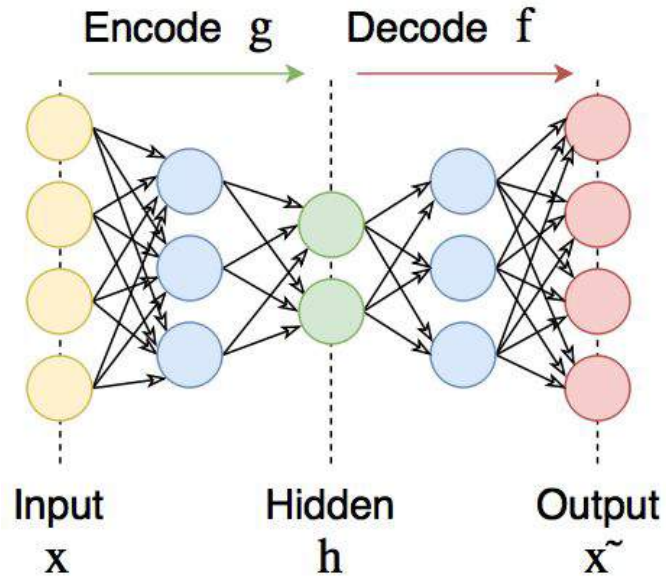
U- матрица слов

V - матрица контекстов



$$M = U \cdot \Sigma \cdot V^*$$

# Следующий уровень - нейронные сети



# Word2vec

Идея - давайте учить сеть предсказывать вероятности встречаемости слова в контексте и надеяться, что признаки, полученные при обучении, окажутся хорошими представлениями для слов

# Прорыв!

---

## Distributed Representations of Words and Phrases and their Compositionality

---

**Tomas Mikolov**  
Google Inc.  
Mountain View  
mikolov@google.com

**Ilya Sutskever**  
Google Inc.  
Mountain View  
ilyasu@google.com

**Kai Chen**  
Google Inc.  
Mountain View  
kai@google.com

**Greg Corrado**  
Google Inc.  
Mountain View  
gcorrado@google.com

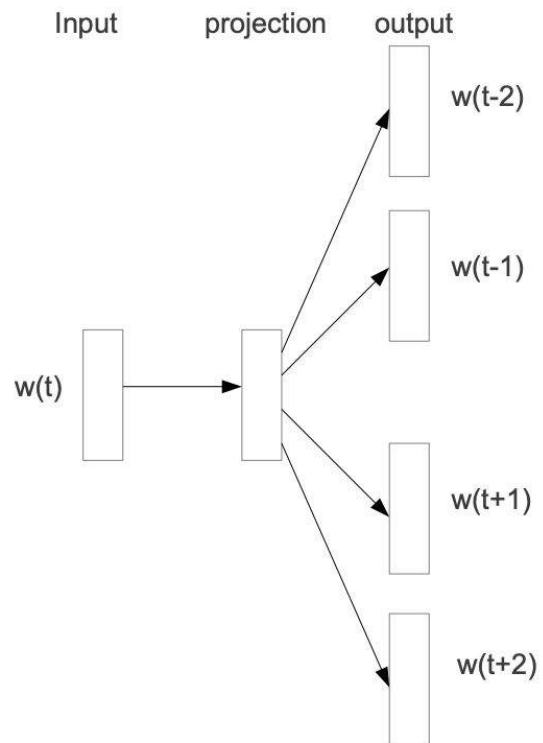
**Jeffrey Dean**  
Google Inc.  
Mountain View  
jeff@google.com

### Abstract

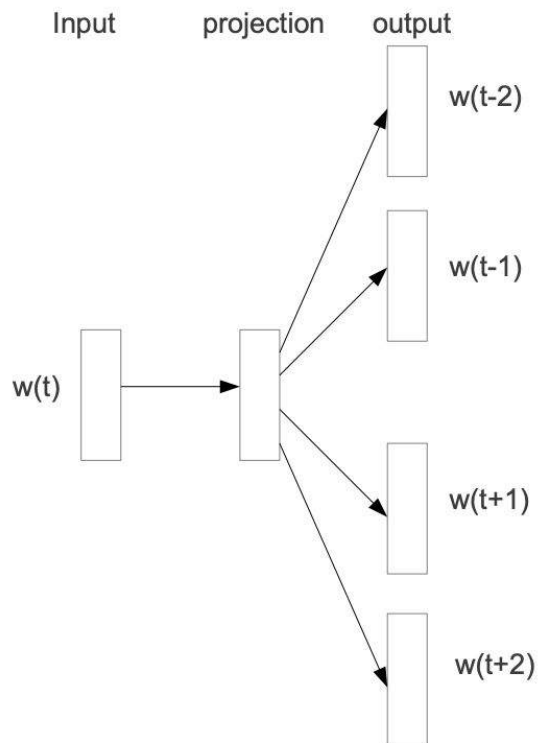
The recently introduced continuous Skip-gram model is an efficient method for learning high-quality distributed vector representations that capture a large number of precise syntactic and semantic word relationships. In this paper we present several extensions that improve both the quality of the vectors and the training speed. By subsampling of the frequent words we obtain significant speedup and also learn more regular word representations. We also describe a simple alternative to the hierarchical softmax called negative sampling.

An inherent limitation of word representations is their indifference to word order and their inability to represent idiomatic phrases. For example, the meanings of

# Skip-gram



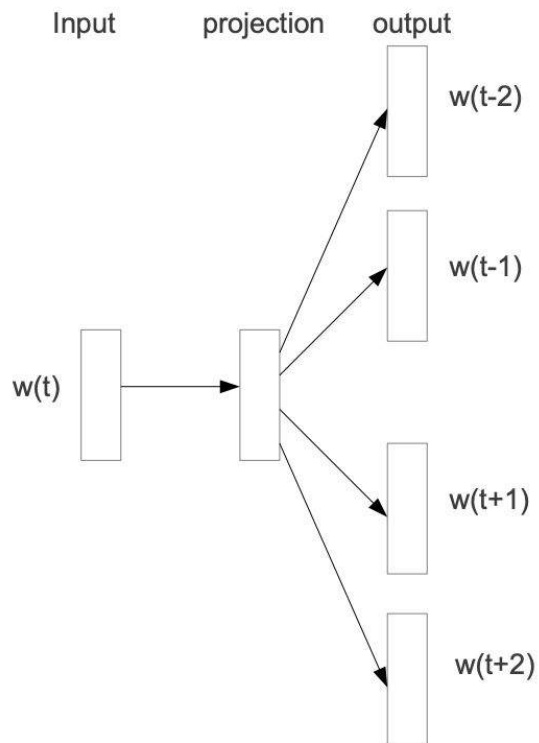
# Skip-gram



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



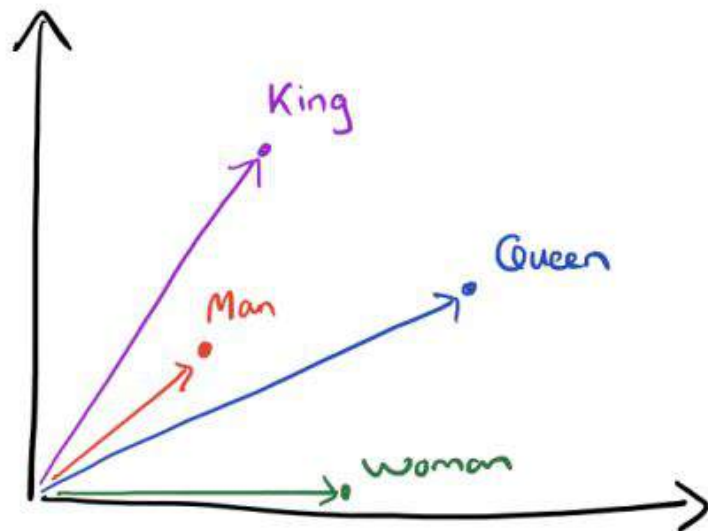
# Skip-gram



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

$$p(w_O | w_I) = \frac{\exp(v'_{w_O}{}^\top v_{w_I})}{\sum_{w=1}^W \exp(v'_w{}^\top v_{w_I})}$$

# Эмбеддинги как признаки



Word  
Vectors

Кажется, слабо для прорыва...

# Проблемы и фишки

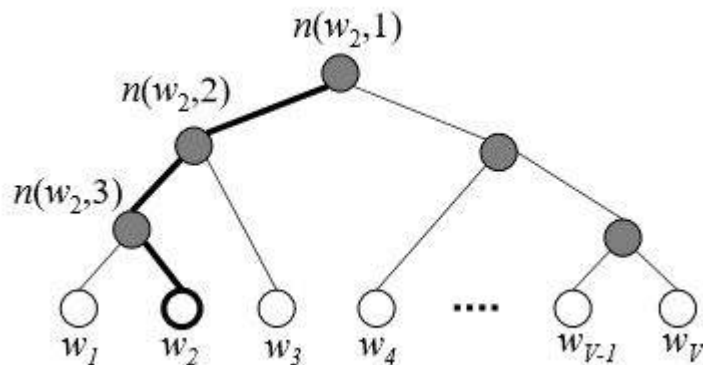
Проблема: Softmax по размеру словаря считать долго

Фишка: Иерархический softmax

# Проблемы и фишки

Проблема: Softmax по размеру словаря считать долго

Фишка: Иерархический softmax



$$p(w|w_I) = \prod_{j=1}^{L(w)-1} \sigma \left( \mathbb{I}[n(w, j+1) = \text{ch}(n(w, j))] \cdot v'_{n(w, j)}{}^\top v_{w_I} \right)$$

# Проблемы и фишки

Проблема: Слишком много “несоседей” для слова в словаре

Фишка: Negative sampling

$$\log \sigma(v'_{w_O}{}^\top v_{w_I}) + \sum_{i=1}^k \mathbb{E}_{w_i \sim P_n(w)} \left[ \log \sigma(-v'_{w_i}{}^\top v_{w_I}) \right]$$

# Проблемы и фишки

Проблема: Постоянно обучаемся на частотных словах в словаре

Фишка: Subsampling of Frequent Words

$$P(w_i) = 1 - \sqrt{\frac{t}{f(w_i)}}$$

# FastText

Идея - давайте рассматривать не слова, а  $n$ - граммы!

В чем плюсы FastText:

- 1) Экономно
- 2) Хорошо работает на редких словах (идеально для опечаток)
- 3) Работает на словах, которых не было в обучении

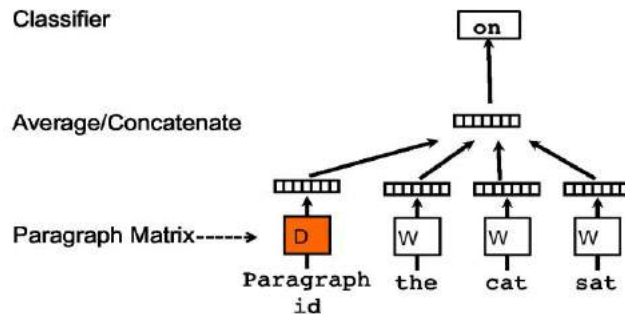
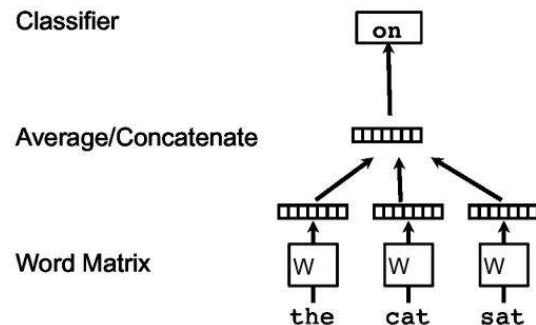


# Проблемы

- 1) Эмбединги только для слов - нужно доучивать на предложения
- 2) Омонемия
- 3) Мало параметров, тяжело выучить сложные зависимости

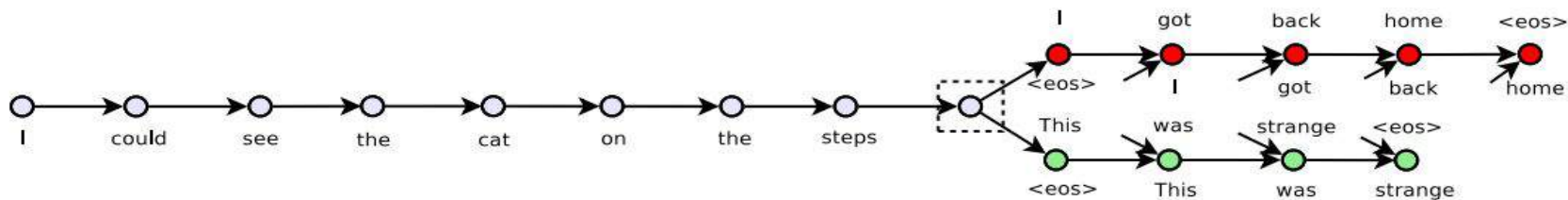
# Эмбединги предложений

ParagraphVector DBOW -  
предсказываем,  
встречаются ли в данном  
предложении слова из  
него



# Эмбединги предложений

SkipThought - учимся для данного предложения предсказывать соседние

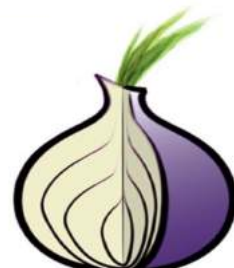


Пройдемся по SOTA ...

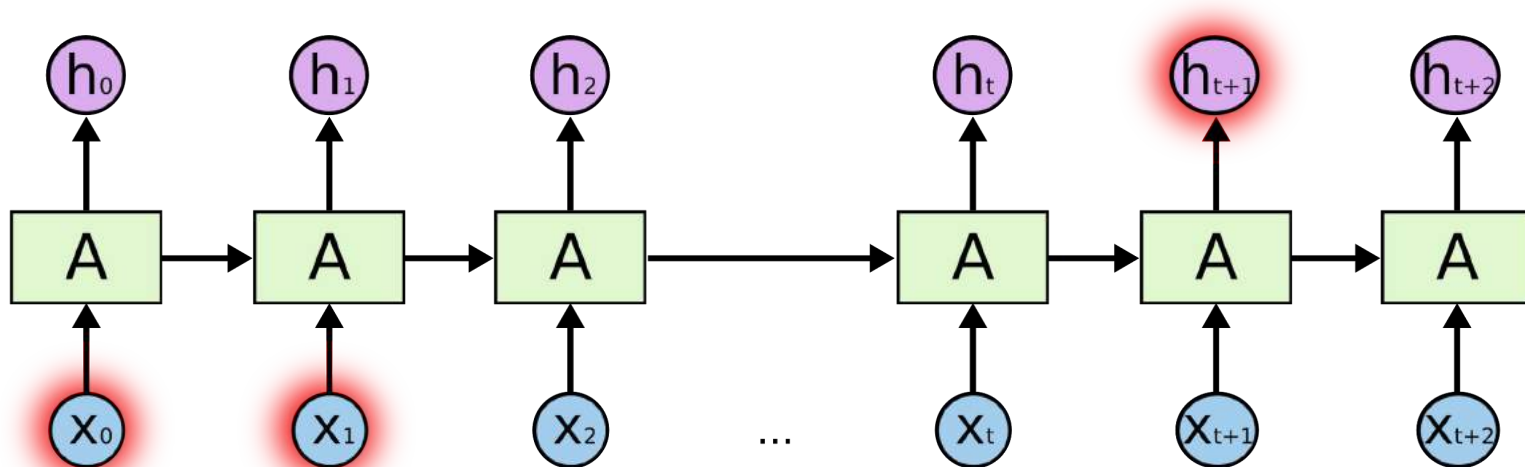
ELMO

# ХОТИМ

- Получать хорошие вектора для задачек nlp
- Учитывать синтаксис, семантику и омонимию
- Использовать нагенерированный человечеством массив информации
- Уметь быстро приспосабливаться под локальные особенности задачи и текста



# Реккурентные нейронные сети



# BiLM(Bidirection language models)

$$(t_1, t_2, \dots, t_N)$$

$$p(t_1, t_2, \dots, t_N) = \prod_{k=1}^N p(t_k \mid t_1, t_2, \dots, t_{k-1})$$

$$p(t_1, t_2, \dots, t_N) = \prod_{k=1}^N p(t_k \mid t_{k+1}, t_{k+2}, \dots, t_N)$$



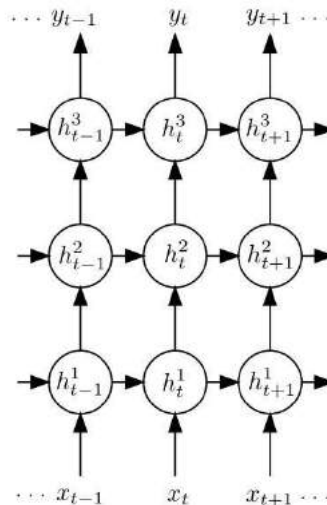
# ELMO

$$\sum_{k=1}^N ( \log p(t_k \mid t_1, \dots, t_{k-1}; \Theta_x, \vec{\Theta}_{LSTM}, \Theta_s) \\ + \log p(t_k \mid t_{k+1}, \dots, t_N; \Theta_x, \overleftarrow{\Theta}_{LSTM}, \Theta_s) )$$

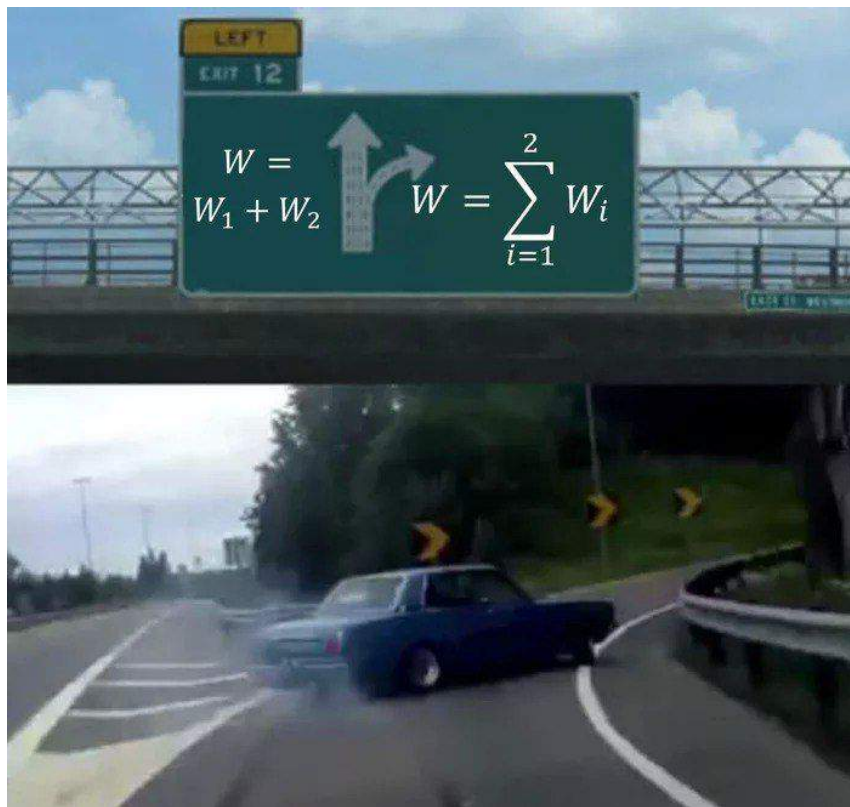
$$\mathbf{ELMo}_k^{task} = E(R_k; \Theta^{task}) = \gamma^{task} \sum_{j=0}^L s_j^{task} \mathbf{h}_{k,j}^{LM}$$

$$R_k = \{ \mathbf{x}_k^{LM}, \vec{\mathbf{h}}_{k,j}^{LM}, \overleftarrow{\mathbf{h}}_{k,j}^{LM} \mid j = 1, \dots, L \} \\ = \{ \mathbf{h}_{k,j}^{LM} \mid j = 0, \dots, L \},$$

$$\lambda \| \mathbf{w} \|_2^2$$



(слоя всего два)



# Побили все SOTA

TASK	PREVIOUS SOTA		OUR BASELINE	ELMo + BASELINE	INCREASE (ABSOLUTE/ RELATIVE)
SQuAD	SAN	84.4	81.1	85.8	4.7 / 24.9%
SNLI	Chen et al. (2017)	88.6	88.0	$88.7 \pm 0.17$	0.7 / 5.8%
SRL	He et al. (2017)	81.7	81.4	84.6	3.2 / 17.2%
Coref	Lee et al. (2017)	67.2	67.2	70.4	3.2 / 9.8%
NER	Peters et al. (2017)	$91.93 \pm 0.19$	90.15	$92.22 \pm 0.10$	2.06 / 21%
SST-5	McCann et al. (2017)	53.7	51.4	$54.7 \pm 0.5$	3.3 / 6.8%

# Интерпретация полученных представлений

Source		Nearest Neighbors
GloVe	play	playing, game, games, played, players, plays, player, Play, football, multiplayer
biLM	Chico Ruiz made a spectacular <u>play</u> on Alusik 's grounder {...}	Kieffer , the only junior in the group , was commended for his ability to hit in the clutch , as well as his all-round excellent <u>play</u> .
	Olivia De Havilland signed to do a Broadway <u>play</u> for Garson {...}	{...} they were actors who had been handed fat roles in a successful <u>play</u> , and had talent enough to fill the roles competently , with nice understatement .

Table 4: Nearest neighbors to “play” using GloVe and the context embeddings from a biLM.

Model	F <sub>1</sub>	Model	Acc.
WordNet 1st Sense Baseline	65.9	<a href="#">Collobert et al. (2011)</a>	97.3
<a href="#">Raganato et al. (2017a)</a>	69.9	<a href="#">Ma and Hovy (2016)</a>	97.6
<a href="#">Iacobacci et al. (2016)</a>	<b>70.1</b>	<a href="#">Ling et al. (2015)</a>	<b>97.8</b>
CoVe, First Layer	59.4	CoVe, First Layer	93.3
CoVe, Second Layer	64.7	CoVe, Second Layer	92.8
biLM, First layer	67.4	biLM, First Layer	97.0
biLM, Second layer	69.0	biLM, Second Layer	95.8

Table 5: All-words fine grained WSD F<sub>1</sub>. For CoVe and the biLM, we report scores for both the first and second layer biLSTMs.

Table 6: Test set POS tagging accuracies for PTB. For CoVe and the biLM, we report scores for both the first and second layer biLSTMs.

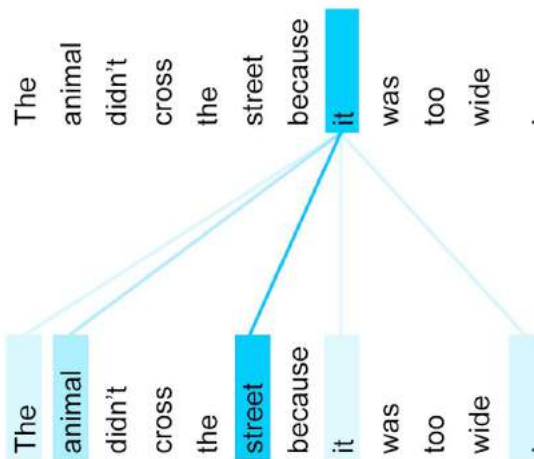
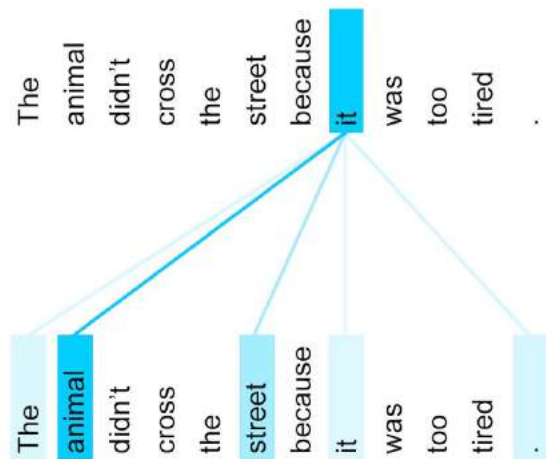
# Transformer

# Attention

$K_{parse}$  key

$V_{parse}$  value

$Q_{parse}$  query

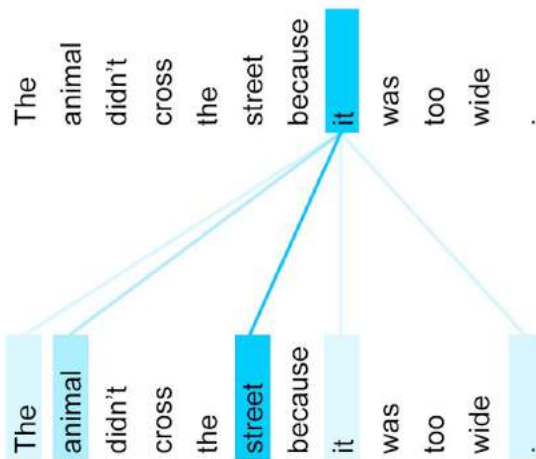
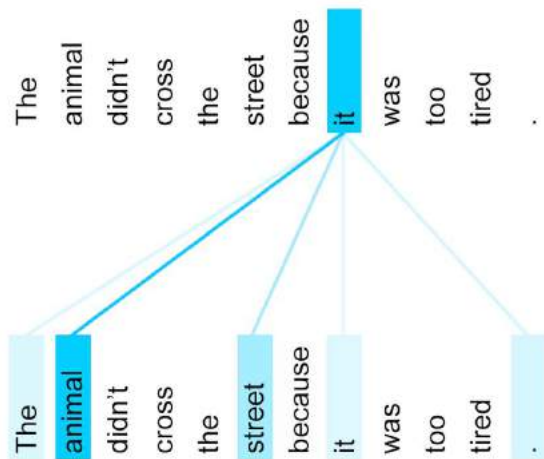


# Attention

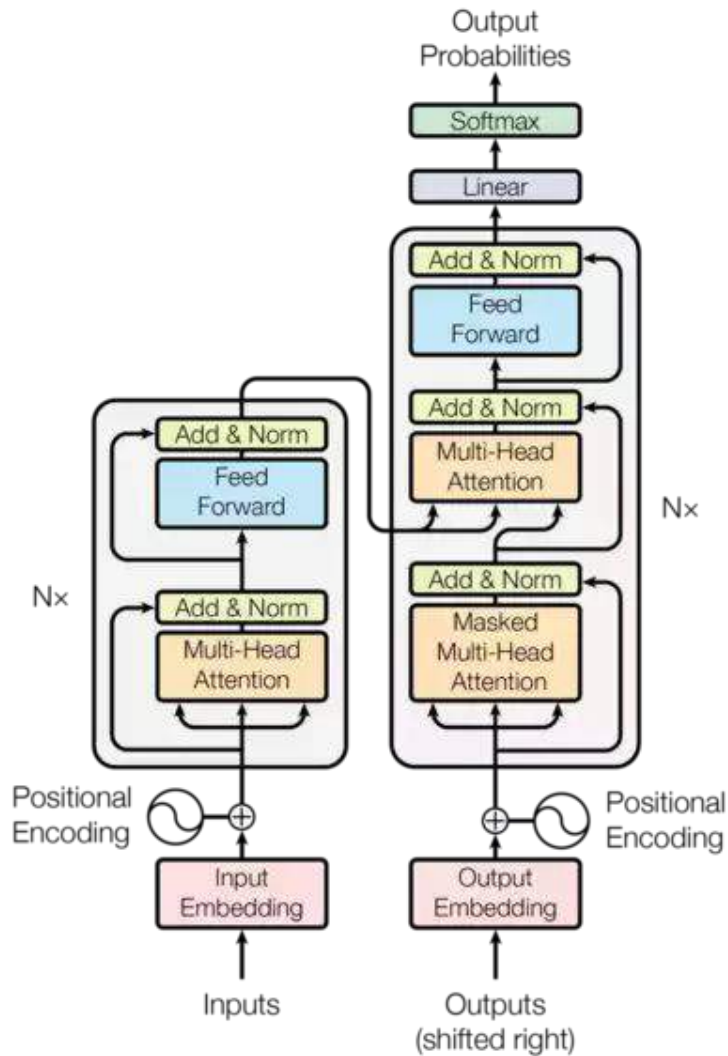
$K_{parse}$  key  
 $V_{parse}$  value  
 $Q_{parse}$  query

$$A_h^{(j)} = \text{softmax}(d_k^{-0.5} Q_h^{(j)} K_h^{(j)T})$$

$$M_h^{(j)} = A_h^{(j)} V_h^{(j)}$$



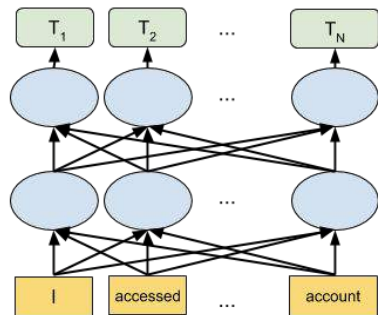
# Transformer



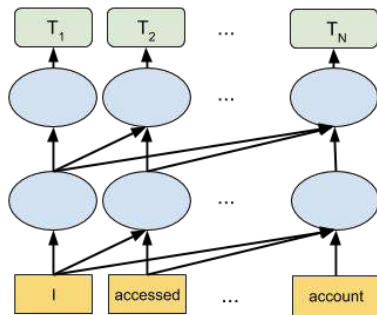


# BERT

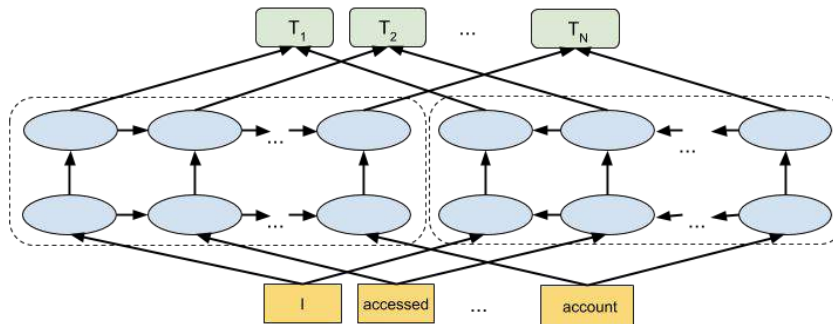
BERT (Ours)



OpenAI GPT



ELMo

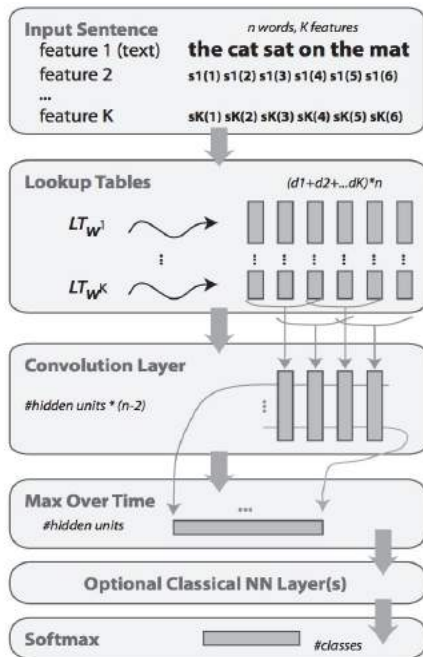


# Как в идеале выглядит решение задач?



Мультитаск- можно попробовать учить задачи  
совместно

# Флэшбек - 2008 - Test of time award ICLR



---

## A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning

---

Ronan Collobert

Jason Weston

NEC Labs America, 4 Independence Way, Princeton, NJ 08540 USA

COLLOBERT@NEC-LABS.COM

JASONW@NEC-LABS.COM

# Предыдущие работы 2008

Виды multitask:

## 1) Cascading Features

Тренируем эмбединги под конкретную задачу, используем их в другой

## 2) Shallow Joint Training

Одна и та же модель предсказывает таргеты для разных задач  
одновременно

# 2018 - Клеим модели с помощью RL

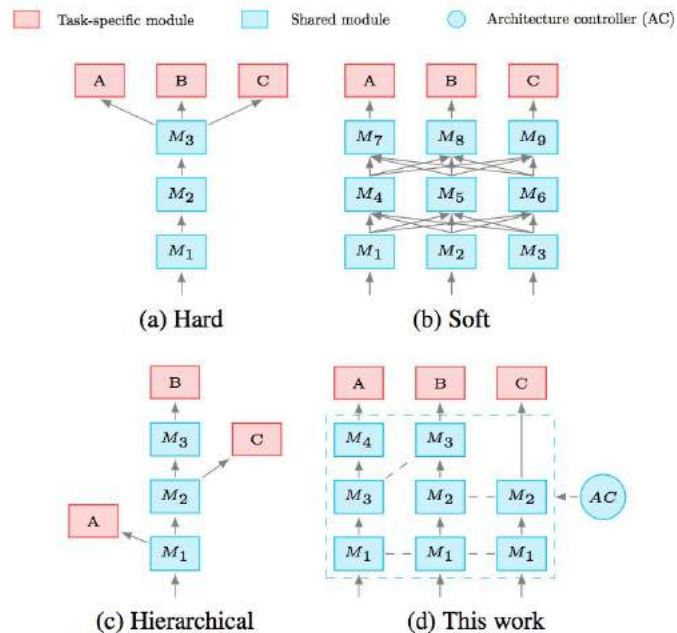


Figure 1: Sharing schemes of multi-task learning. Modules connected by dashed lines in (1d) are identical, selected from a shared module pool.

## Exploring Shared Structures and Hierarchies for Multiple NLP Tasks

Junkun Chen\*, Kaiyu Chen\*, Xinchu Chen, Xipeng Qiu†, Xuanjing Huang

Shanghai Key Laboratory of Intelligent Information Processing, Fudan University

School of Computer Science, Fudan University

{jkcchen16, kyachen15, xinchuchen13, xpqiu, xjhuang}@fudan.edu.cn

# Чем вдохновились?

## NEURAL ARCHITECTURE SEARCH WITH REINFORCEMENT LEARNING

Barret Zoph\*, Quoc V. Le  
Google Brain  
{barretzoph,qvl}@google.com

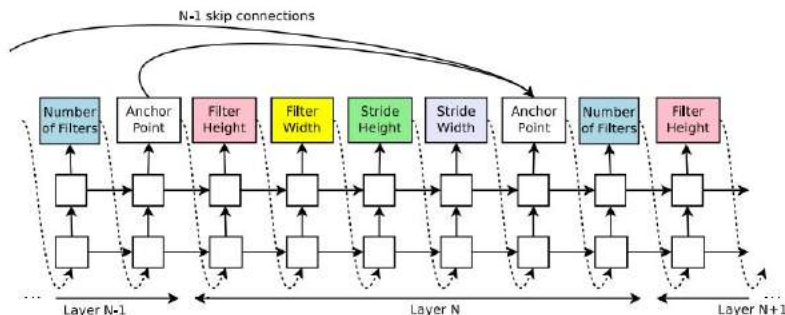


Figure 4: The controller uses anchor points, and set-selection attention to form skip connections.

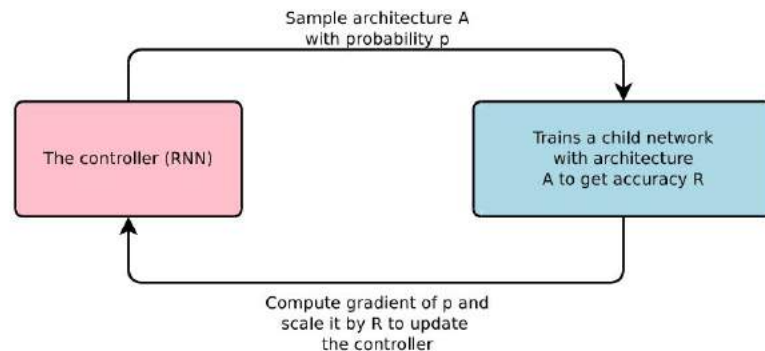


Figure 1: An overview of Neural Architecture Search.



# Обучение

- 1)  $\mathcal{M}$  - набор модулей, которые мы хотим соединять с помощью RL
- 2)  $\mathbf{a}$  - набор действий(соединяем модули)
- 3)  $\mathcal{E}$  - состояние, из какого эмбединга хотим соединять модули
- 4)  $\{x_i^k, y_i^k\}_{i=1}^{N_k}$  - обучающая выборка
- 5)  $R$  - реворд

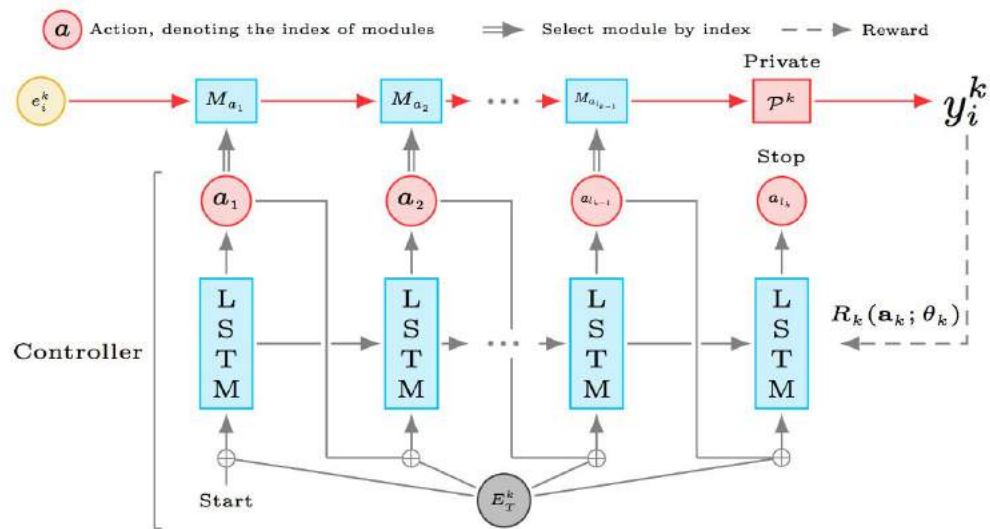
$$\mathbf{e}_i^k \xrightarrow{\mathcal{M}_{\mathbf{a}_k}} \mathbf{s}_i^k \xrightarrow{\mathcal{P}^k} p(y_i^k | x_i^k)$$
$$R_k(\mathbf{a}_k; \theta) = \frac{1}{N_k} \sum_{i=1}^{N_k} \log p(y_i^k | x_i^k).$$

- 6)  $\theta$  - параметры модели

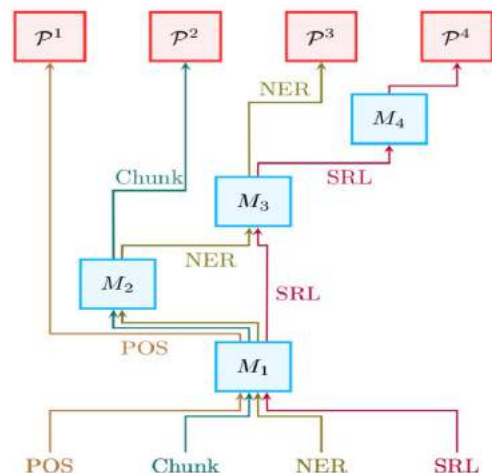
Максимизируем ожидаемый реворд:

$$\mathcal{J}_k(\phi) = \mathbb{E}_{\pi_\phi(E_T(k), \mathbf{a}_k)} [R_k(\mathbf{a}_k; \theta_k)].$$

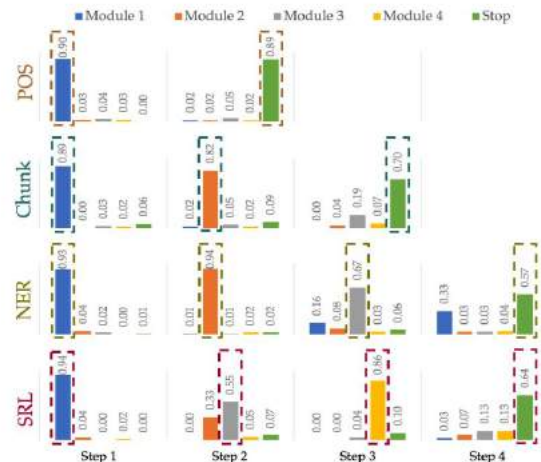
# Обучение



# Анализ выученного



(a) The hierarchical architecture selected by the controller. Different colors represent different tasks. The lines and arrows show the chosen information flow of tasks. For example, given a sample of Chunking task, the controller sequentially chooses  $M_1$ ,  $M_2$ , and  $Stop$ , forming the path in teal of the picture. The selective modules are originally disordered and are renamed  $M_1 \dots M_4$  for more explicit representation.

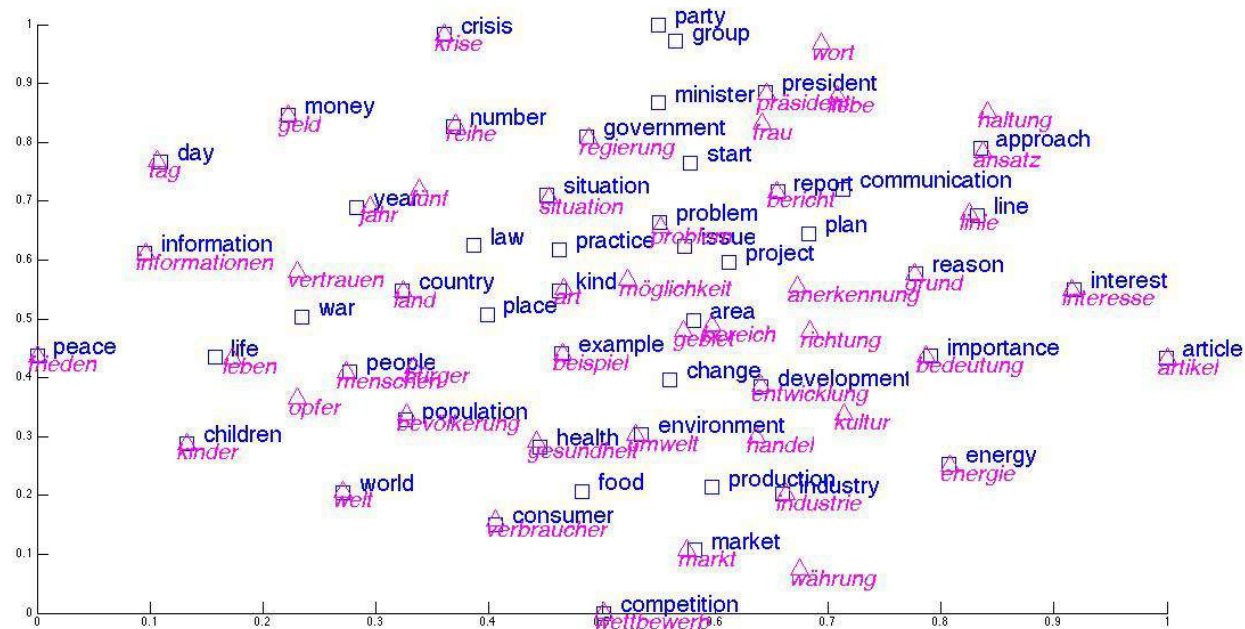


(b) Probabilities of module selection at each step of the decision process. Probabilities are drawn from controller's module decision process after softmax function. The first four columns in every histogram represent the probability of choosing four different modules. The fifth column means the probability of deciding to *Stop*. For chunking, the decision sequence with the biggest probabilities is  $M_1, M_2, Stop$ , which corresponds with the path in 4a.

Figure 4: The selected architecture and the corresponding probabilities at every step of the decision process.

# Кросс-эмбединги

Как сделать перевод без учителя?



# Зачем они нужны?

- 1) Выровнять слова с разных языков в одно пространство
- 2) Тогда мы сможем эффективно решить проблему машинного перевода
- 3) Использовать при информации о языке, полученную при работе с другим языком

# Обучение с учителем

1. Получаем два набора эмбеддингов, обученных независимо на одноязычных данных
2. Находим соответствие между двумя наборами данных так, чтобы переводы были близки в общем пространстве

Проблема - дорого!

Поэтому мы хотим обойтись без обучающей выборки!



# Процесс отображения

$$W^* = \underset{W \in M_d(\mathbb{R})}{\operatorname{argmin}} \|WX - Y\|_F \quad (1)$$

$$W^* = \underset{W \in O_d(\mathbb{R})}{\operatorname{argmin}} \|WX - Y\|_F = UV^T, \text{ with } U\Sigma V^T = \operatorname{SVD}(YX^T). \quad (2)$$

$d$  - размерность эмбедингов

$M_d(\mathbb{R})$  - пространство  $d \times d$  матриц

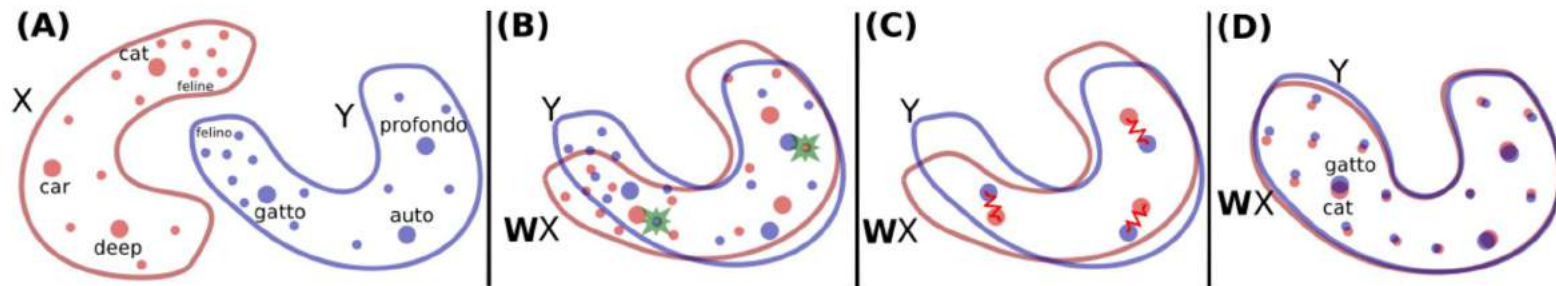
$X$  и  $Y$  - две выровненные матрицы  $d \times n$  содержащие эмбединги слов из параллельного словаря

$W$  - линейное отображение между двумя пространствами

# Что предлагают авторы

- 1) Получите два набора вложений, обученных независимо на одноязычных данных
- 2) Изучите соответствие между двумя наборами так, чтобы переводы были близки в общем пространстве
- 3) Делаем отображение без кросс-язычного надзора!
- 4) Находим соответствующие пары
- 5) Радуемся

# Do mapping without cross-lingual supervision - steps



- 1) Adversarial - B
- 2) Refinement (procrustes) - C
- 3) CSLS - D

$\mathcal{X} = \{x_1, \dots, x_n\}$  and  $\mathcal{Y} = \{y_1, \dots, y_m\}$  - two- sets of n and m word embeddings

$\theta_D$  - discriminator parameters

$W$  - mapping

## Discriminator loss

$$\mathcal{L}_D(\theta_D|W) = -\frac{1}{n} \sum_{i=1}^n \log P_{\theta_D}(\text{source} = 1|Wx_i) - \frac{1}{m} \sum_{i=1}^m \log P_{\theta_D}(\text{source} = 0|y_i). \quad (3)$$

## Mapper loss

$$\mathcal{L}_W(W|\theta_D) = -\frac{1}{n} \sum_{i=1}^n \log P_{\theta_D}(\text{source} = 0|Wx_i) - \frac{1}{m} \sum_{i=1}^m \log P_{\theta_D}(\text{source} = 1|y_i). \quad (4)$$

# Find matching pairs (with CSLS)

**Хотим-** найти переводы из одного пространства в другое

**Проблема-** ближайшие соседи работают плохо в пространствах высокой размерности

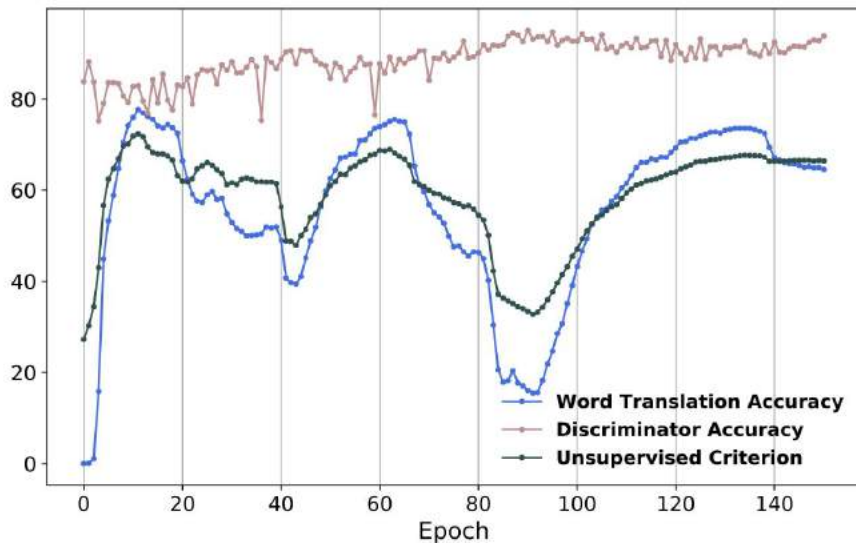
**Решение-**

$$r_T(Wx_s) = \frac{1}{K} \sum_{y_t \in \mathcal{N}_T(Wx_s)} \cos(Wx_s, y_t),$$

$$\text{CSLS}(Wx_s, y_t) = 2 \cos(Wx_s, y_t) - r_T(Wx_s) - r_S(y_t).$$

# Выбор модели

Используем обучение без учителя, количественно определяем близость исходных и целевых пространств эмбедингов.



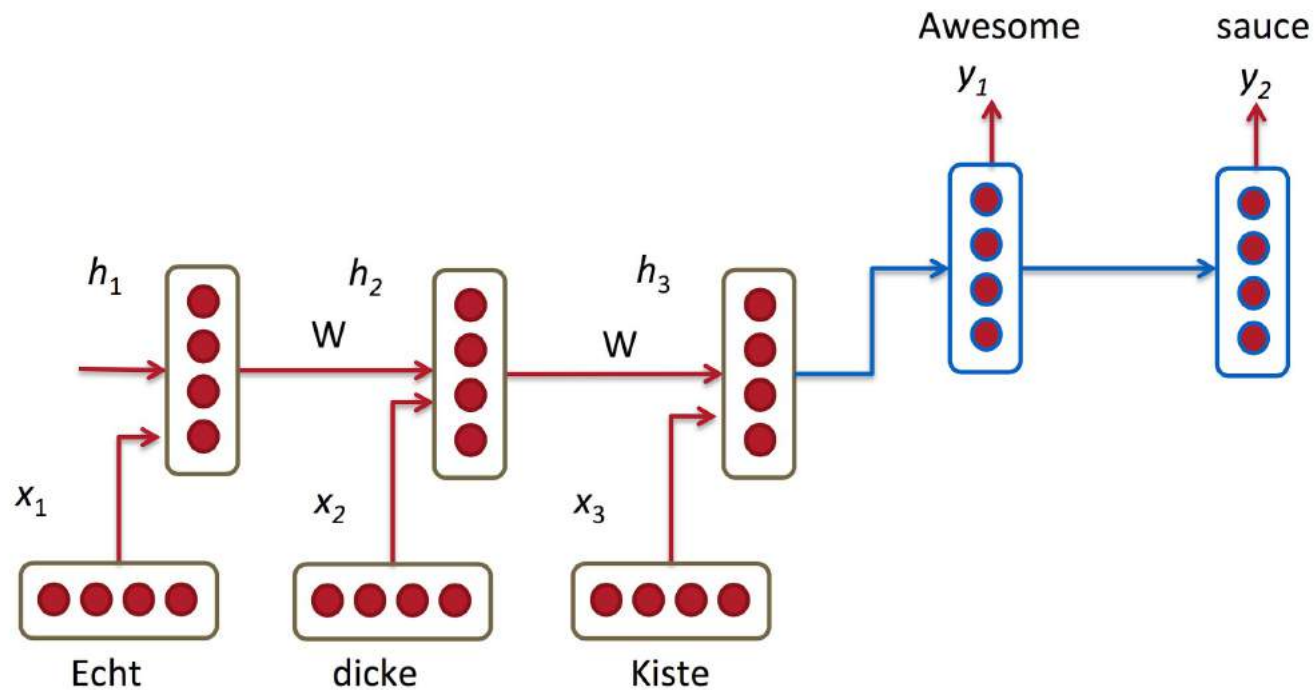
**Figure 2: Unsupervised model selection.** Correlation between our unsupervised validation criterion (black line) and actual word translation accuracy (blue line). In this particular experiment, the selected model is at epoch 10. Observe how our criterion is well correlated with translation accuracy.

# Результаты

	English to italian			Italian to english		
	P@1	P@5	P@10	P@1	P@5	P@10
<i>Methods with cross-lingual supervision</i>						
Mikolov et al. (2013b) <sup>†</sup>	10.5	18.7	22.8	12.0	22.1	26.7
Dinu et al. (2015) <sup>†</sup>	45.3	72.4	80.7	48.9	71.3	78.3
Smith et al. (2017) <sup>†</sup>	54.6	72.7	78.2	42.9	62.2	69.2
Procrustes - NN	42.6	54.7	59.0	53.5	65.5	69.5
Procrustes - CSLS	<b>66.1</b>	77.1	80.7	<b>69.5</b>	<b>79.6</b>	<b>83.5</b>
<i>Methods without cross-lingual supervision</i>						
Adv - CSLS	42.5	57.6	63.6	47.0	62.1	67.8
Adv - Refine - CSLS	65.9	<b>79.7</b>	<b>83.1</b>	69.0	<b>79.7</b>	83.1

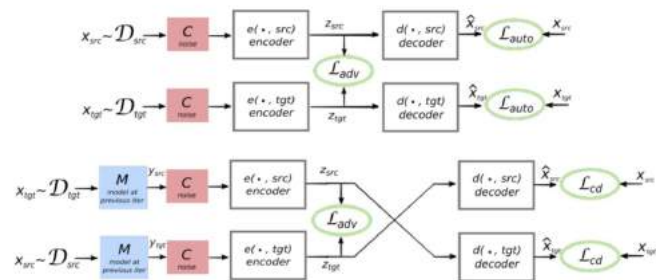
**Table 3: English-Italian sentence translation retrieval.** We report the average P@k from 2,000 source queries using 200,000 target sentences. We use the same embeddings as in Smith et al. (2017). Their results are marked with the symbol <sup>†</sup>.

# Перевод с учителем





# Перевод без учителя



## Unsupervised Machine Translation Using Monolingual Corpora Only- Давайте похожую идею запишем в машинный перевод

	Multi30k-Task1				WMT			
	en-fr	fr-en	de-en	en-de	en-fr	fr-en	de-en	en-de
Supervised	56.83	50.77	38.38	35.16	27.97	26.13	25.61	21.33
word-by-word	8.54	16.77	15.72	5.39	6.28	10.09	10.77	7.06
word reordering	-	-	-	-	6.68	11.69	10.84	6.70
oracle word reordering	11.62	24.88	18.27	6.79	10.12	20.64	19.42	11.57
Our model: 1st iteration	27.48	28.07	23.69	19.32	12.10	11.79	11.10	8.86
Our model: 2nd iteration	31.72	30.49	24.73	21.16	14.42	13.49	13.25	9.75
Our model: 3rd iteration	32.76	32.07	26.26	22.74	15.05	14.31	13.33	9.64

Table 1: BLEU score on the WMT and Multi30k-Task1 datasets using greedy decoding.

И еще похожую идею используем в переводах стиля?



Кросс-эмбединги картинка-текст

# Лирическое отступление 1 - может в обобщенных пространствах есть смысл?

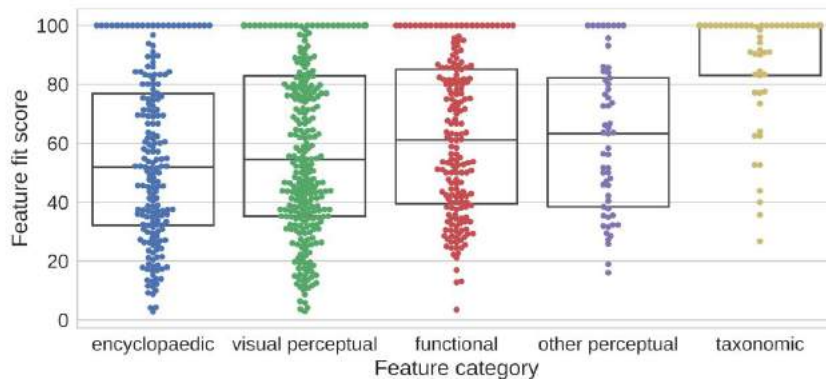
Как научить вектора слов передавать физический опыт?

Идея - в мире мы получаем достаточно много опыта - не только читая текста, но и осязая мир. Что если нам недостаточно текстов, чтобы выучиться?

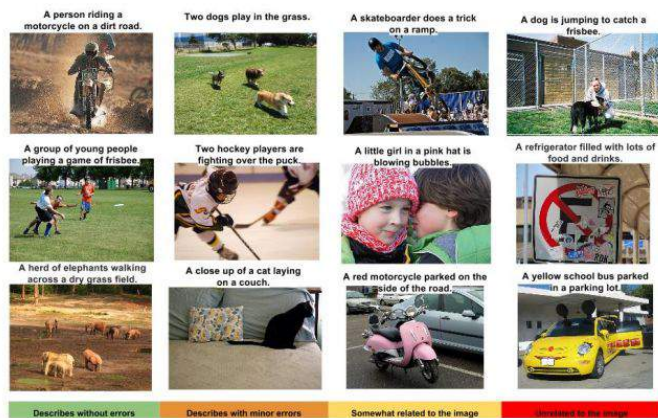
A	B
Concept	Feature
accordion	a_musical_instrument
accordion	has_keys
accordion	requires_air
accordion	associated_with_polkas
accordion	has_buttons
accordion	used_by_moving_bellows
accordion	inbeh-_produces_music
accordion	is_loud
accordion	worn_on_chest

Таксономические и функциональные concepts в среднем выше, чем воспринимаемые concepts!

Идея - давайте возьмем и для каждой feature будем предсказывать, к какому concept она относится.



# Кросс-эмбединги картинки и текст



Learning Deep Structure-Preserving Image-Text Embeddings

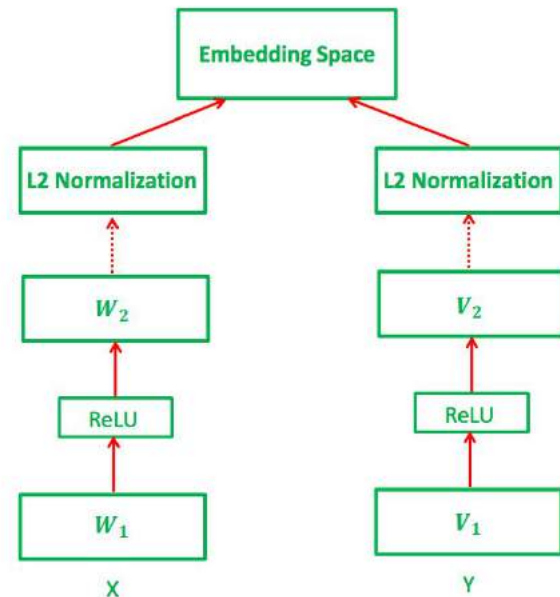


Figure 1. Our model structure: there are two branches in the network, one for images ( $X$ ) and the other for text ( $Y$ ). Each branch consists of fully connected layers with ReLU nonlinearities between them, followed by L2 normalization at the end.

Машины сами создают язык

# Переговоры

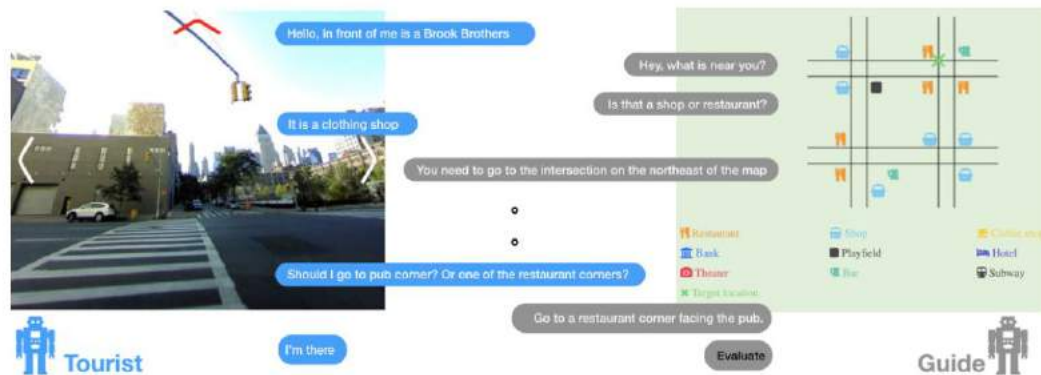


Figure 1: Example of the Talk The Walk task: two agents, a “tourist” and a “guide”, interact with each other via natural language in order to have the tourist navigate towards the correct location. The guide has access to a map and knows the target location but not the tourist location, while the tourist does not know the way but can navigate in a 360-degree street view environment.

Алгоритм, позволяющий коммуницировать двум агентам - туристу, который видит мир вокруг себя и должен добраться до финальной точки, но где она, турист не знает, и гида, который не видит мир вокруг туриста и по описанию должен рассказать туристу, что сделать дальше, чтобы дойти до цели. С рядом предположений алгоритм бьет качество решения этой задачи людьми.

de Vries, Harm, et al. "Talk the walk: Navigating new york city through grounded dialogue." *arXiv preprint arXiv:1807.03367*(2018).

# Переговоры

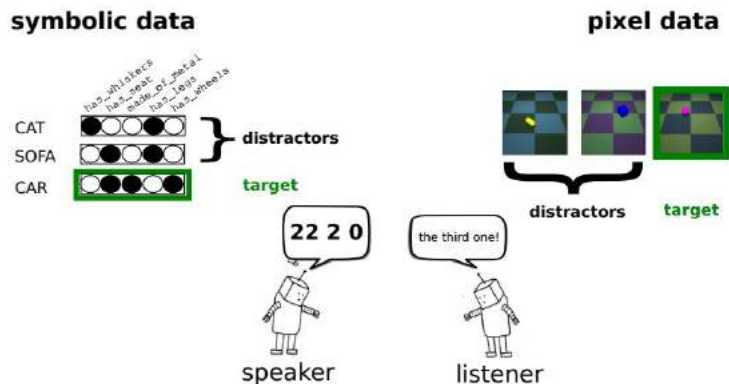


Figure 1: High-level overview of the referential game.

Когда мы заставляем агентов объяснять друг другу различные понятия, у них получается создать язык с проблемами и преимуществами, с которыми сталкиваемся и мы при создании языка - композиционная структура, омонимия, специфика, связанная с важностью распознавания понятий, похожая структура модели для говорящего и слушающего. У агентов получается добиться хорошего понимания друг друга даже на картинках



# Спасибо за внимание!

@tany\_savelieva

[savelievatanyya@gmail.com](mailto:savelievatanyya@gmail.com)

@tldr\_arxiv