Нейросетевые подходы в машинном переводе

Элбакян Мовсес

Национальный Исследовательский Университет Высшая Школа Экономики

2018

План

- Метрики в машинном переводе
- Зарождение Neural Machine Translation(NMT)
- seq2seq + attention
- NMT VS SMT
- Проблемы NMT
- Совмещение SMT и NMT

Метрики в машинном переводе

Human evaluation:

- **Automatic Language Processing Advisory Committee (ALPAC)**
  - Intelligibility (how "understandable" the sentence was)
  - Fidelity (how much information the translated sentence retained)
- **Advanced Research Projects Agency (ARPA)**
  - ~~Comprehension evaluation~~ (Lang1 ->[H] Lang2 ->[M] Lang1)
  - ~~Quality panel evaluation~~ (Evaluation of professional translators)
  - Evaluation based on adequacy and fluency (Monolingual evaluation)

Automatic evaluation:

- BLEU (bilingual evaluation understudy)
- NIST
- METEOR (Metric for Evaluation of Translation with Explicit ORdering)
- Word error rate
- …

Метрики в машинном переводе (BLEU)

$$\mathrm{BLEU} = \min\left(1, \frac{\textit{output-length}}{\textit{reference-length}}\right) \left(\prod_{i=1}^{4} \textit{precision}_i\right)^{\frac{1}{4}}$$

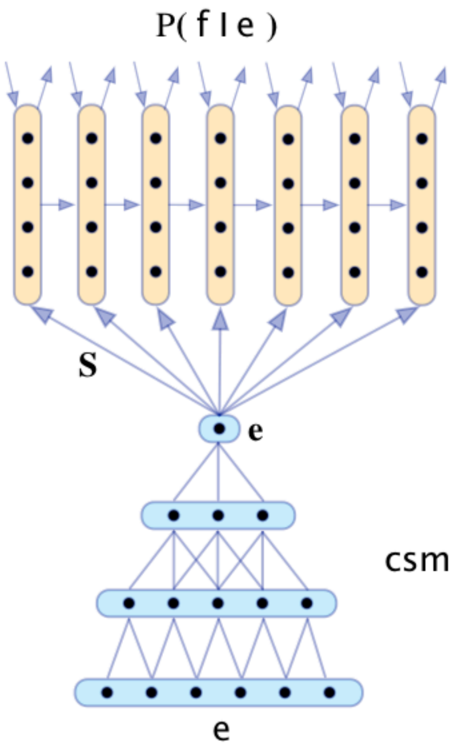
SYSTEM A: Israeli officials responsibility of airport safety
2-GRAM MATCH (Israeli officials) · 1-GRAM MATCH (airport)

REFERENCE: Israeli officials are responsible for airport security

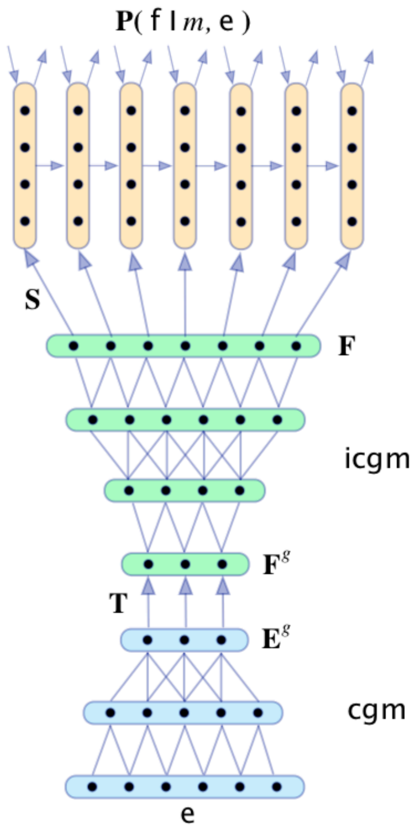
SYSTEM B: airport security Israeli officials are responsible
2-GRAM MATCH (airport security) · 4-GRAM MATCH (Israeli officials are responsible)

Metric	System A	System B
precision (1gram)	3/6	6/6
precision (2gram)	1/5	4/5
precision (3gram)	0/4	2/4
precision (4gram)	0/3	1/3
brevity penalty	6/7	6/7
BLEU	0%	52%

Зарождение Neural Machine Translation(NMT)



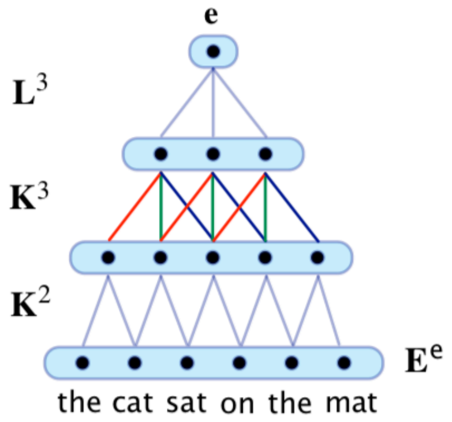
RCTM I

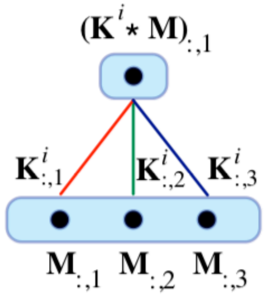


RCTM II

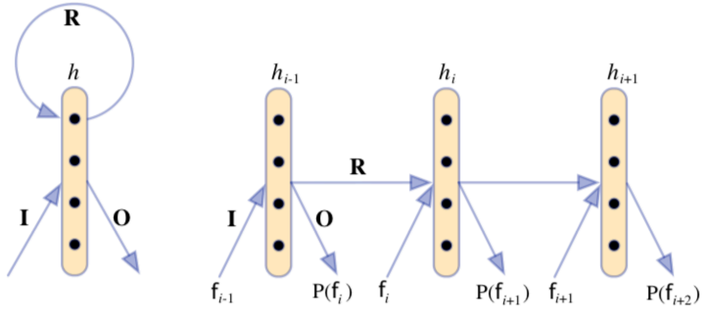
Recurrent Continuous Translation Models,
Nal Kalchbrenner, Phil Blunsom
2013

Зарождение Neural Machine Translation(NMT)





$$
\begin{aligned}
\mathbf{s} &= \mathbf{S} \cdot \mathsf{csm}(\mathsf{e}) \\
h_1 &= \sigma(\mathbf{I} \cdot \mathsf{v}(\mathsf{f}_1) + \mathbf{s}) \\
h_{i+1} &= \sigma(\mathbf{R} \cdot h_i + \mathbf{I} \cdot \mathsf{v}(\mathsf{f}_{i+1}) + \mathbf{s}) \\
o_{i+1} &= \mathbf{O} \cdot h_i
\end{aligned}
$$



$$
P(\mathsf{f}_i = v | \mathsf{f}_{1:i-1}) = \frac{\exp(o_{i,v})}{\sum_{v=1}^{V} \exp(o_{i,v})}
$$

Recurrent Continuous Translation Models,
Nal Kalchbrenner, Phil Blunsom
2013

Зарождение Neural Machine Translation(NMT)

WMT-NT	*2009*	*2010*	*2011*	*2012*
KN-5	218	213	222	225
RLM	178	169	178	181
IBM 1	207	200	188	197
FA-IBM 2	153	146	135	144
RCTM I	143	134	140	142
RCTM II	**86**	**77**	**76**	**77**

Perplexity results

WMT-NT	*2009*	*2010*	*2011*	*2012*
RCTM I + WP	19.7	21.1	22.5	21.5
RCTM II + WP	19.8	21.1	22.5	21.7
cdec (12 features)	19.9	21.2	22.6	21.8

BLEU scores

Минусы:

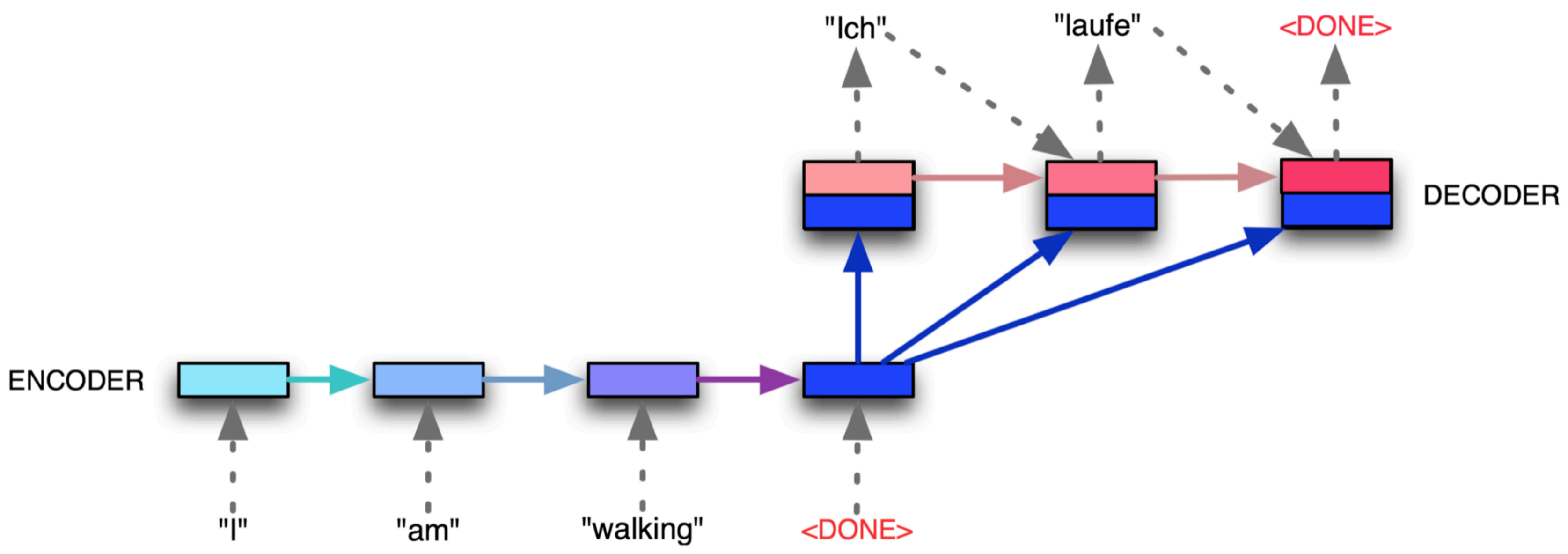
- Проблема взврывающихся/затухающих градиантов
- Не можем запоминать длительные последовательности
- Запоминаем лишнее
- Fixed-vector problem
- RNN плохо учатся

Nal Kalchbrenner, Phil Blunsom
2013

seq2seq

Sutskever et al. and Cho et al., 2014

seq2seq



Sutskever et al. and Cho et al., 2014

seq2seq

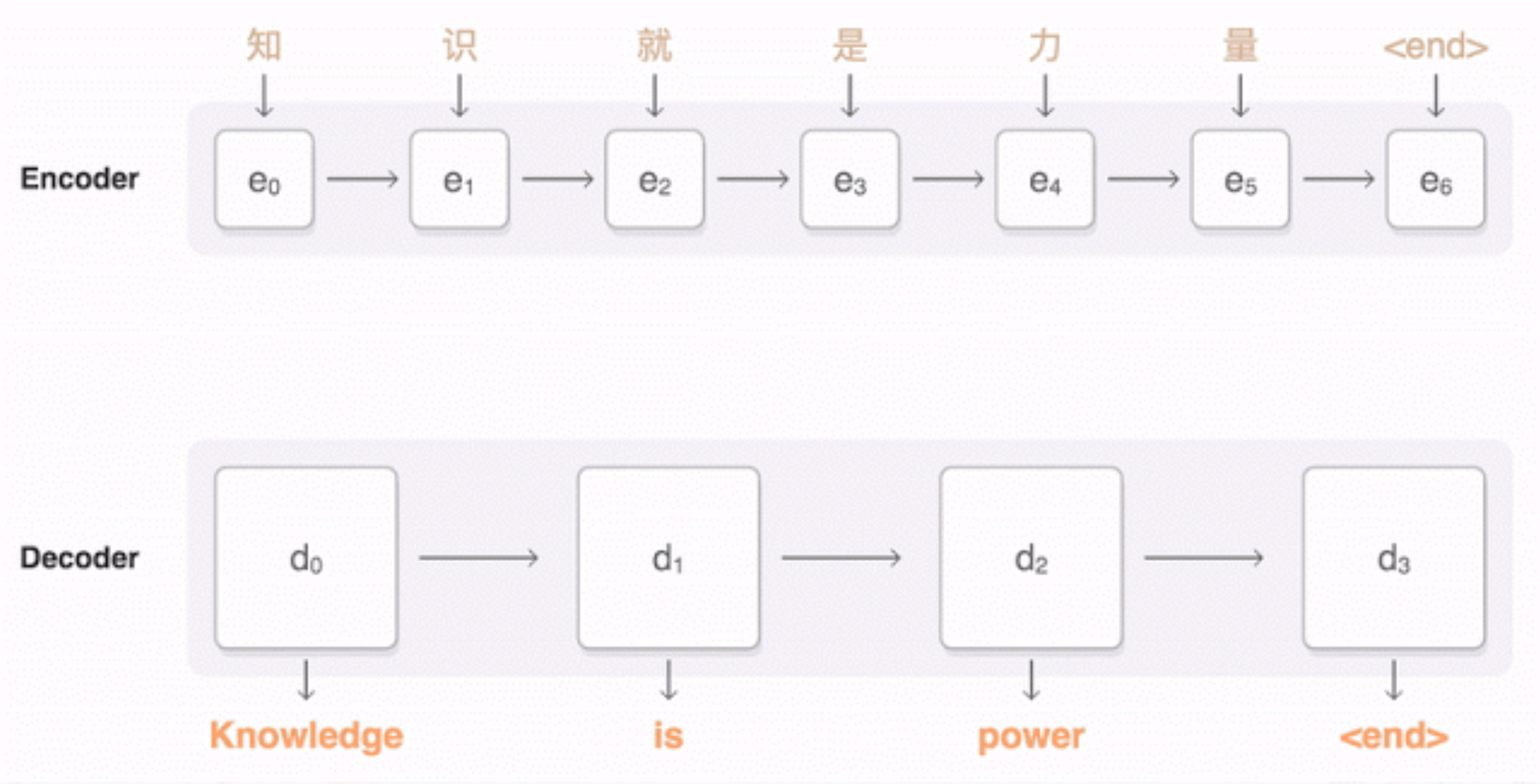
Method	test BLEU score (ntst14)
Bahdanau et al. [2]	28.45
Baseline System [29]	33.30
Single forward LSTM, beam size 12	26.17
Single reversed LSTM, beam size 12	30.59
Ensemble of 5 reversed LSTMs, beam size 1	33.00
Ensemble of 2 reversed LSTMs, beam size 12	33.27
Ensemble of 5 reversed LSTMs, beam size 2	34.50
Ensemble of 5 reversed LSTMs, beam size 12	**34.81**

Минусы:

- Fixed-vector problem
- RNN плохо учатся

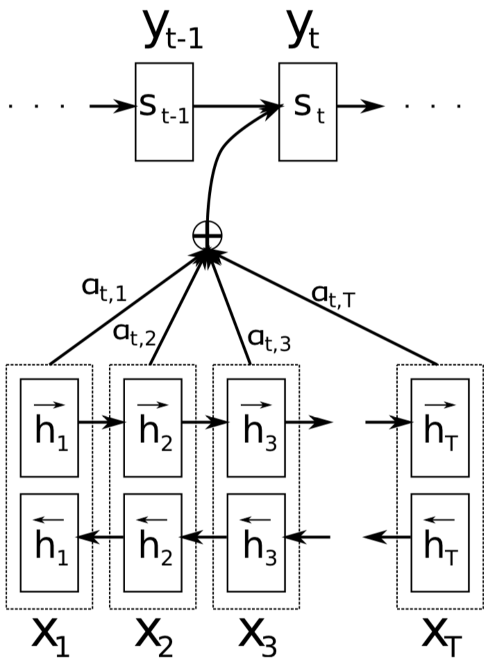
Sutskever et al. and Cho et al.,
2014

Attention



Bahdanau, D., Cho, K., & Bengio, Y.
2014

Attention



$$c_i = \sum_{j=1}^{T_x} \alpha_{ij} h_j .$$

$$\alpha_{ij} = \frac{\exp\left(e_{ij}\right)}{\sum_{k=1}^{T_x} \exp\left(e_{ik}\right)}$$

$$e_{ij} = v_a^\top \tanh\left(W_a s_{i-1} + U_a h_j\right)$$

Bahdanau, D., Cho, K., & Bengio, Y.
2014

Attention

Model	All	No UNK°
RNNencdec-30	13.93	24.19
RNNsearch-30	21.50	31.44
RNNencdec-50	17.82	26.71
RNNsearch-50	26.75	34.16
RNNsearch-50*	28.45	36.15
Moses	33.30	35.63

Минусы:

- RNN плохо учатся

Bahdanau, D., Cho, K., & Bengio, Y.
2014

NMT VS SMT

	Neural Machine Translation	Statistical Machine Translation
Training time	More	Less
Training data	Less	More
Translation (decoding) time	More	Less
CPU usage	More	Less
Space in disk	Less	More
Mechanism	Sentence by sentence	Word by word/ phrase by phrase
	Attentional encoder-decoder networks; optimization	Statistical analysis; probability
	Train multiple features jointly	Feature engineering required
Interpretability		👍
Long distance reordering	👍	
Morphology, syntax, and agreement errors	👍	
Translation style consistency for the same word		👍
Tolerance to noisy data	👍	
Multilingual/ multi-domain translation	👍	
Vocabulary/Rare word Problem		👍

Проблемы NMT

1. Domain mismatch
2. Amount of Training Data
3. Rare Words
4. Long Sentences
5. Word Alignment
6. Beam Search

Philipp Koehn , Rebecca Knowles
2017

Проблемы NMT. Domain mismatch

System ↓	Law		Medical		IT		Koran		Subtitles	
All Data	30.5	32.8	45.1	42.2	35.3	44.7	17.9	17.9	26.4	20.8
Law	31.1	34.4	12.1	18.2	3.5	6.9	1.3	2.2	2.8	6.0
Medical	3.9	10.2	39.4	43.5	2.0	8.5	0.6	2.0	1.4	5.8
IT	1.9	3.7	6.5	5.3	42.1	39.8	1.8	1.6	3.9	4.7
Koran	0.4	1.8	0.0	2.1	0.0	2.3	15.9	18.8	1.0	5.5
Subtitles	7.0	9.9	9.3	17.8	9.2	13.6	9.0	8.4	25.9	22.1

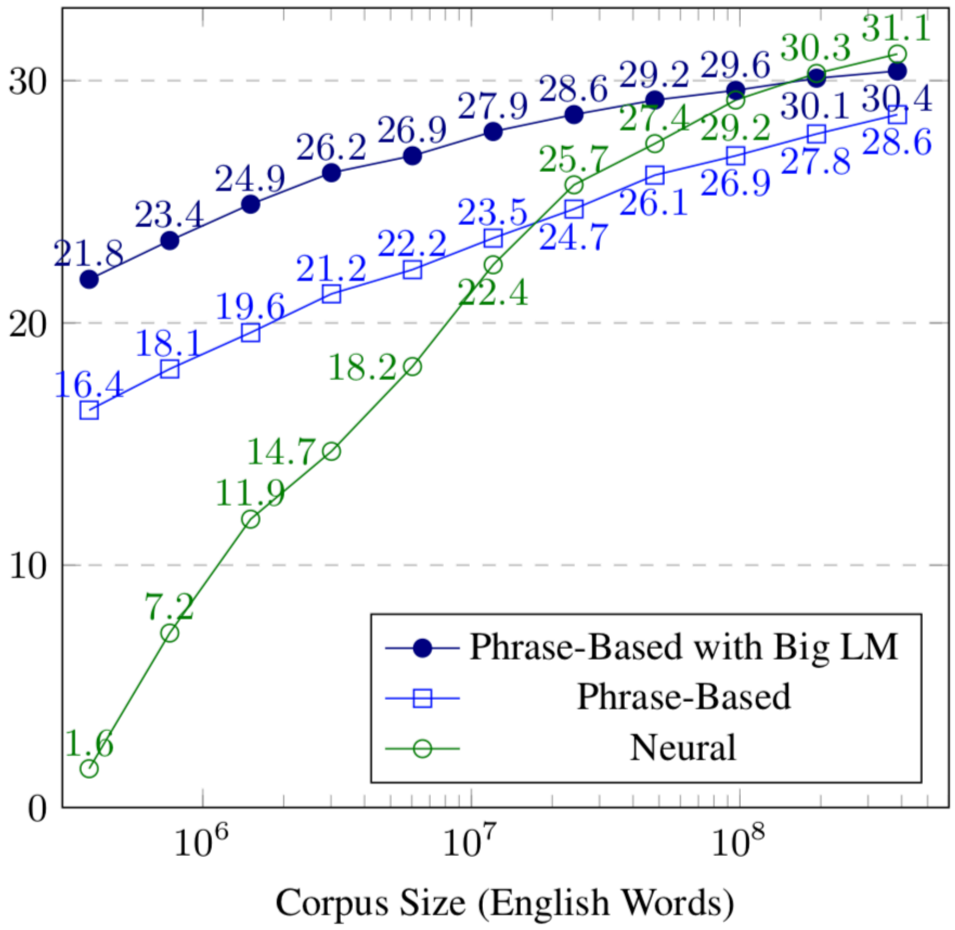
Philipp Koehn , Rebecca Knowles
2017

Проблемы NMT. Domain mismatch

Source	Schaue um dich herum.
Ref.	Look around you.
All	NMT: Look around you. SMT: Look around you.
Law	NMT: Sughum gravecorn. SMT: In order to implement dich Schaue .
Medical	NMT: EMEA / MB / 049 / 01-EN-Final Work progamme for 2002 SMT: Schaue by dich around .
IT	NMT: Switches to paused. SMT: To Schaue by itself . \t \t
Koran	NMT: Take heed of your own souls. SMT: And you see.
Subtitles	NMT: Look around you. SMT: Look around you .

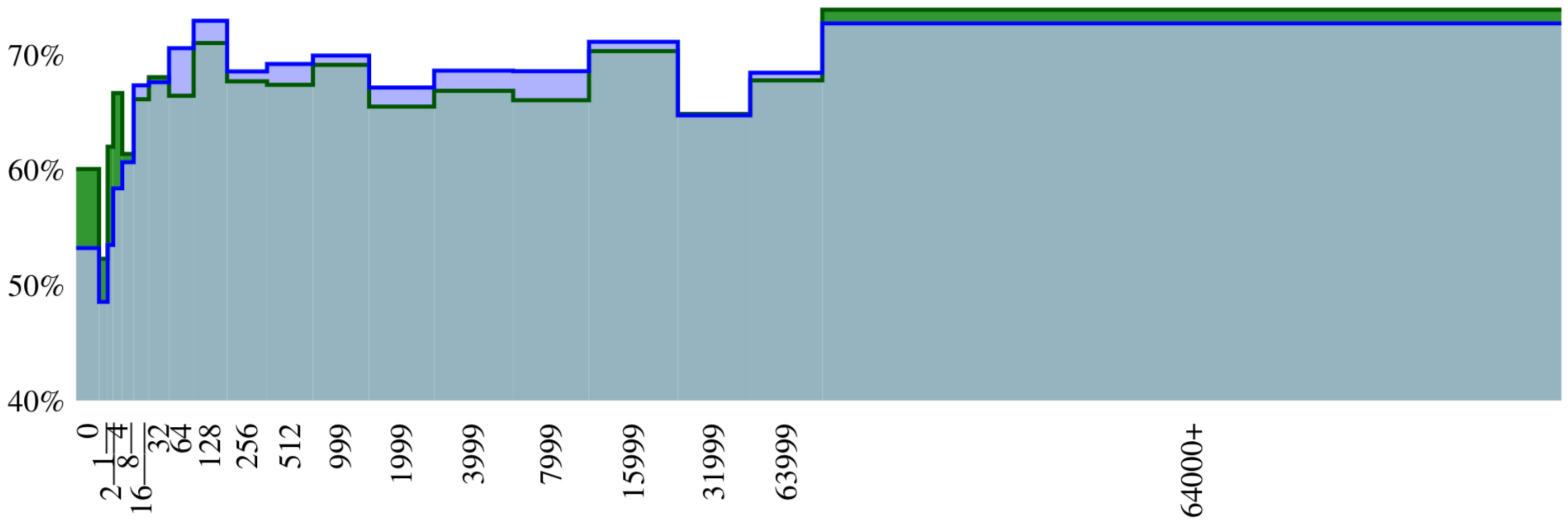
Philipp Koehn , Rebecca Knowles
2017

Проблемы NMT. Amount of training data



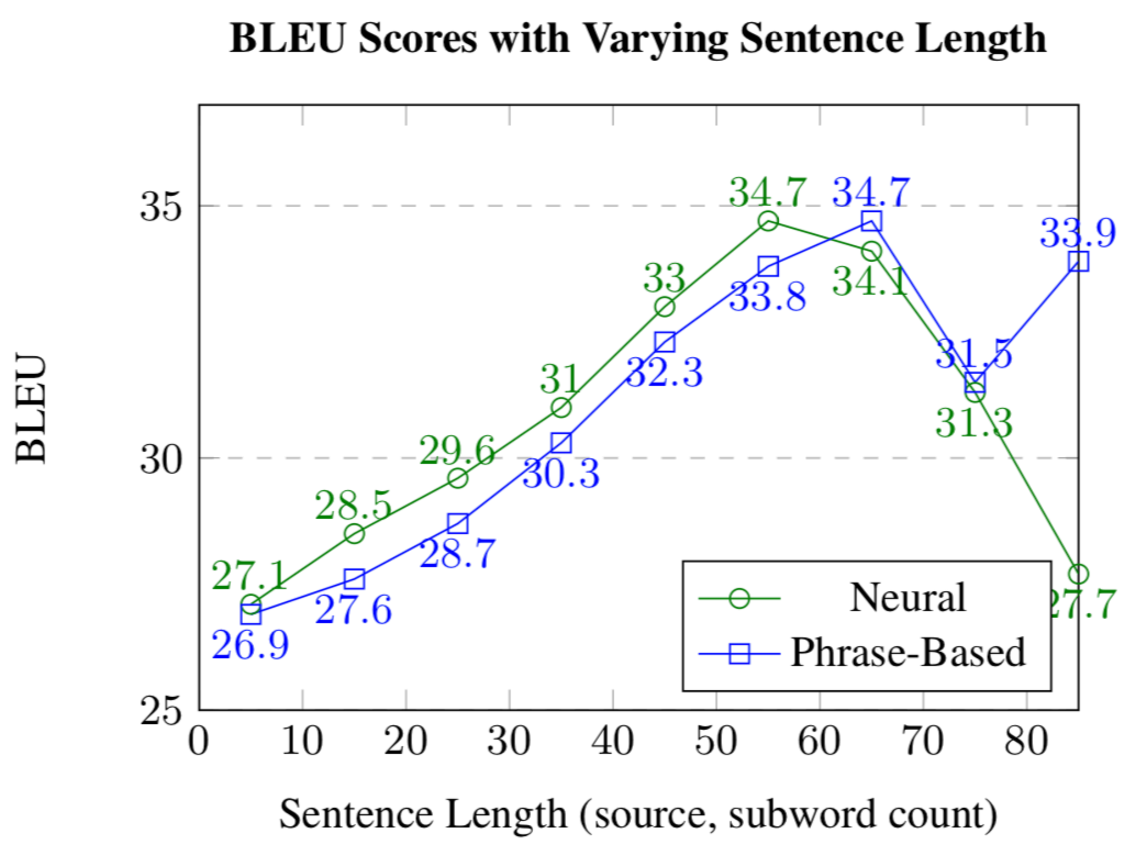
Philipp Koehn , Rebecca Knowles
2017

Проблемы NMT. Rare words



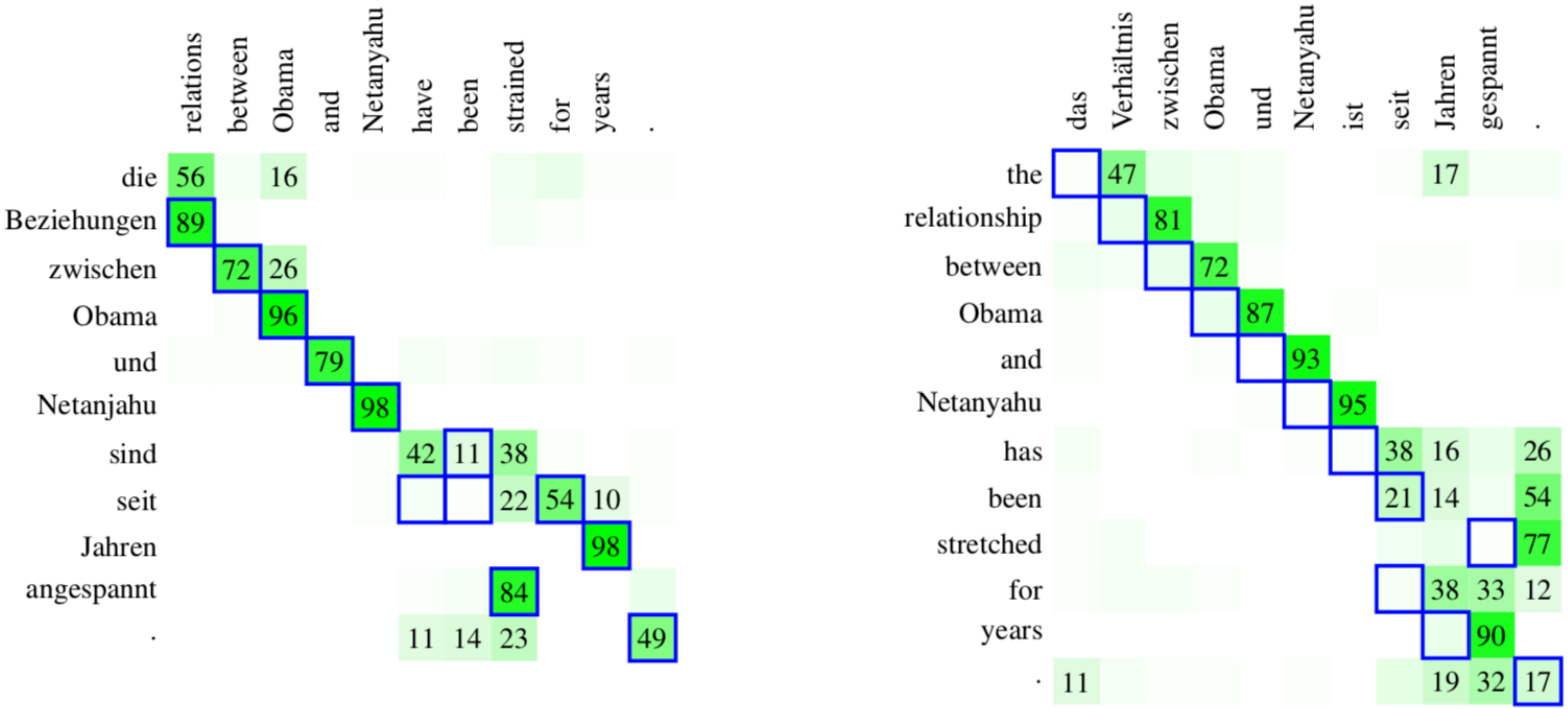
Philipp Koehn , Rebecca Knowles
2017

Проблемы NMT. Long Sentences



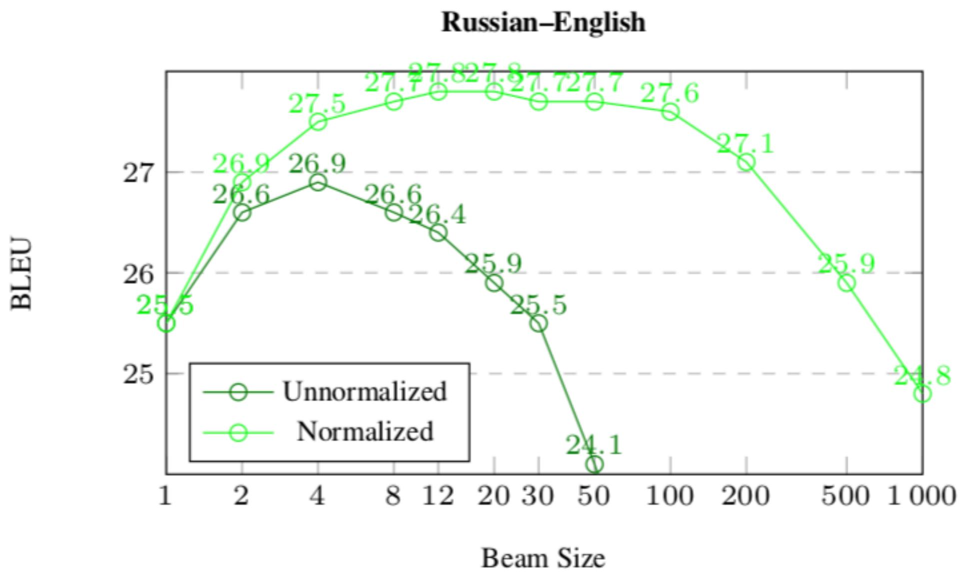
Philipp Koehn , Rebecca Knowles
2017

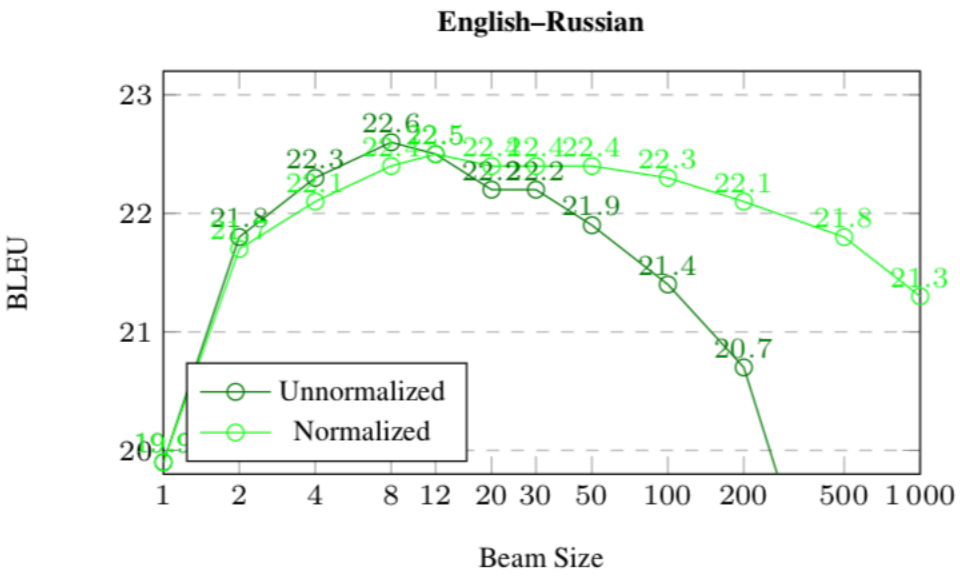
Проблемы NMT. Word alignment



Philipp Koehn , Rebecca Knowles
2017

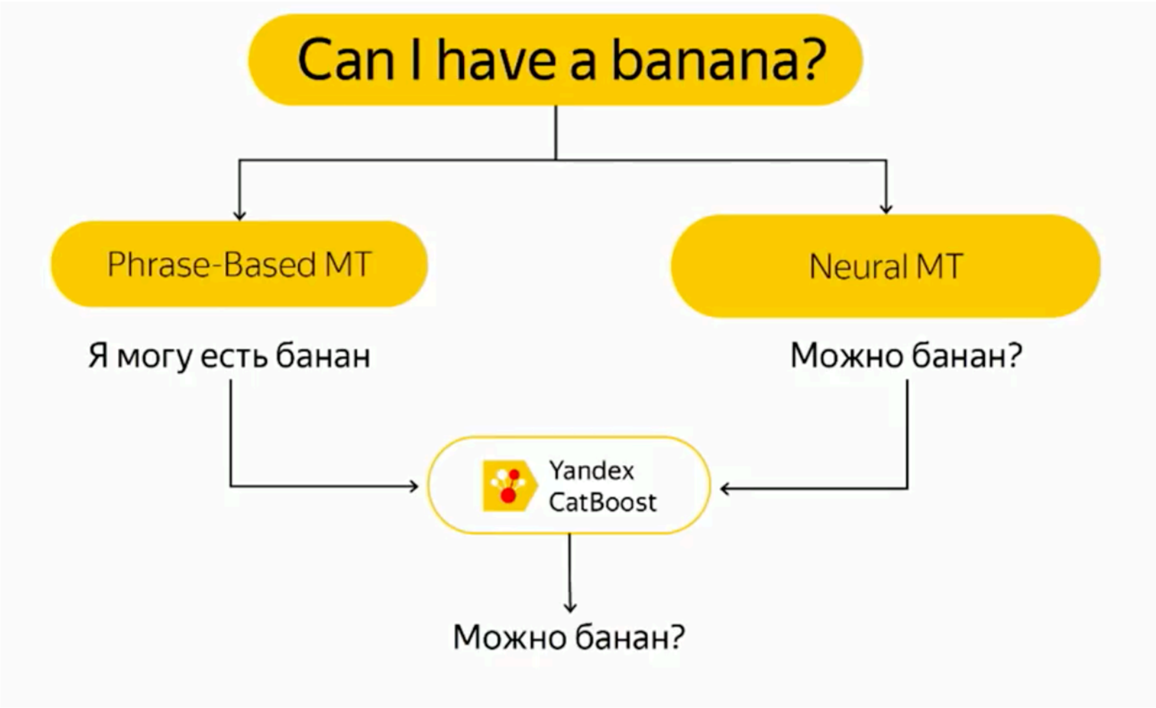
Проблемы NMT. Beam Search

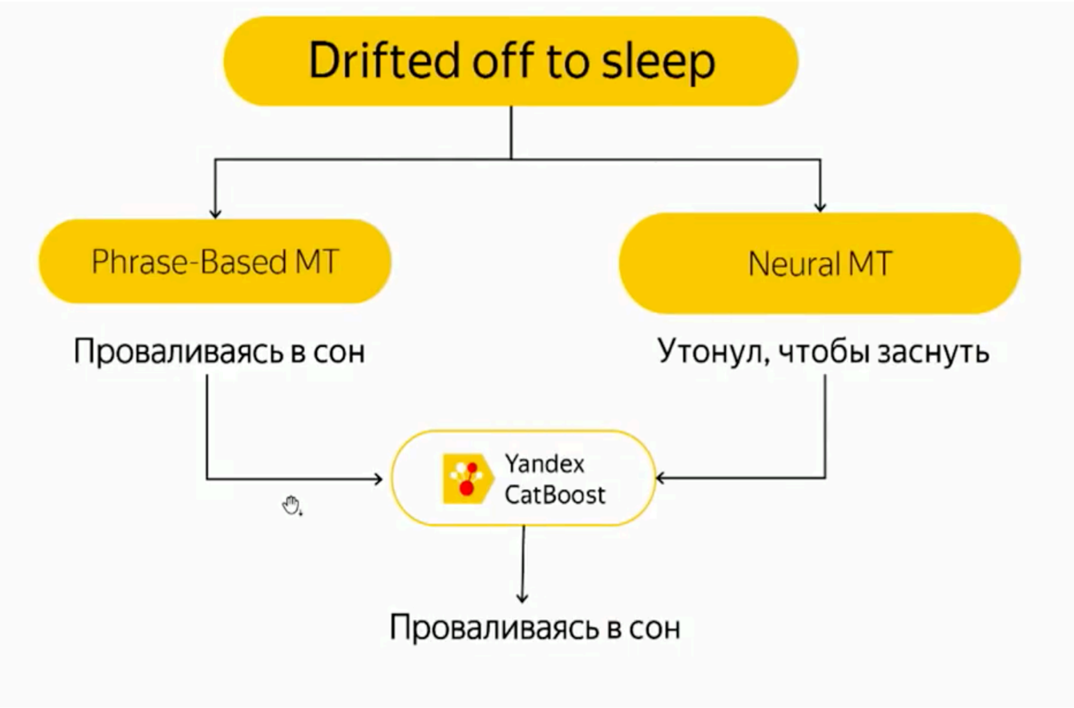




Philipp Koehn , Rebecca Knowles
2017

Совмещение подходов SMT и NMT





Выводы

- NMT >> SMT
- У NMT все еще много нерешенных проблем
- Метрики, подсчитываемые автоматически неидеальны
- Гибридные подходы очень хорошо себя показывают

Источники

- https://www.aclweb.org/anthology/D13-1176
- https://papers.nips.cc/paper/5346-sequence-to-sequence-learning-with-neural-networks.pdf
- https://medium.com/@devnag/seq2seq-the-clown-car-of-deep-learning-f88e1204dac3
- https://arxiv.org/pdf/1409.0473.pdf
- http://www.statmt.org/book/slides/08-evaluation.pdf
- https://syncedreview.com/2017/08/17/history-and-frontier-of-the-neural-machine-translation/
- http://www.aclweb.org/anthology/W17-3204