Нейросетевые методы машинного перевода

Прокопьева Дарья
НИУ ВШЭ
2019

Содержание

- Метрики качества перевода
- Seq2seq
- Attention
- NMT vs SMT

Метрики качества перевода

- Неавтоматические
- WER
- BLEU
- NIST
- METEOR
- F-measure
- и другие

- $$WER = \frac{S + D + I}{T}$$

S - замена, D – удаление, I – вставка, T - количество слов в распознаваемой фразе

- $$BLEU = P_B \exp\left(\sum_{n=0}^{N} w_n \log p_n\right)$$

$$P_B = \begin{cases} 1, & c > r \\ e^{(1-r/c)}, & c \leq r \end{cases}$$

C – длина кандидата, r - перевода

- NIST
- METEOR

$$P = \frac{m}{w_t} \qquad R = \frac{m}{w_r} \qquad F_{mean} = \frac{10P}{R + 9P}$$

m — количество n-грамм в машинном переводе, которые также были найдены в эталонном переводе, w_t — количество n-грамм в машинном переводе, а w_r - количество n-грамм в эталонном переводе

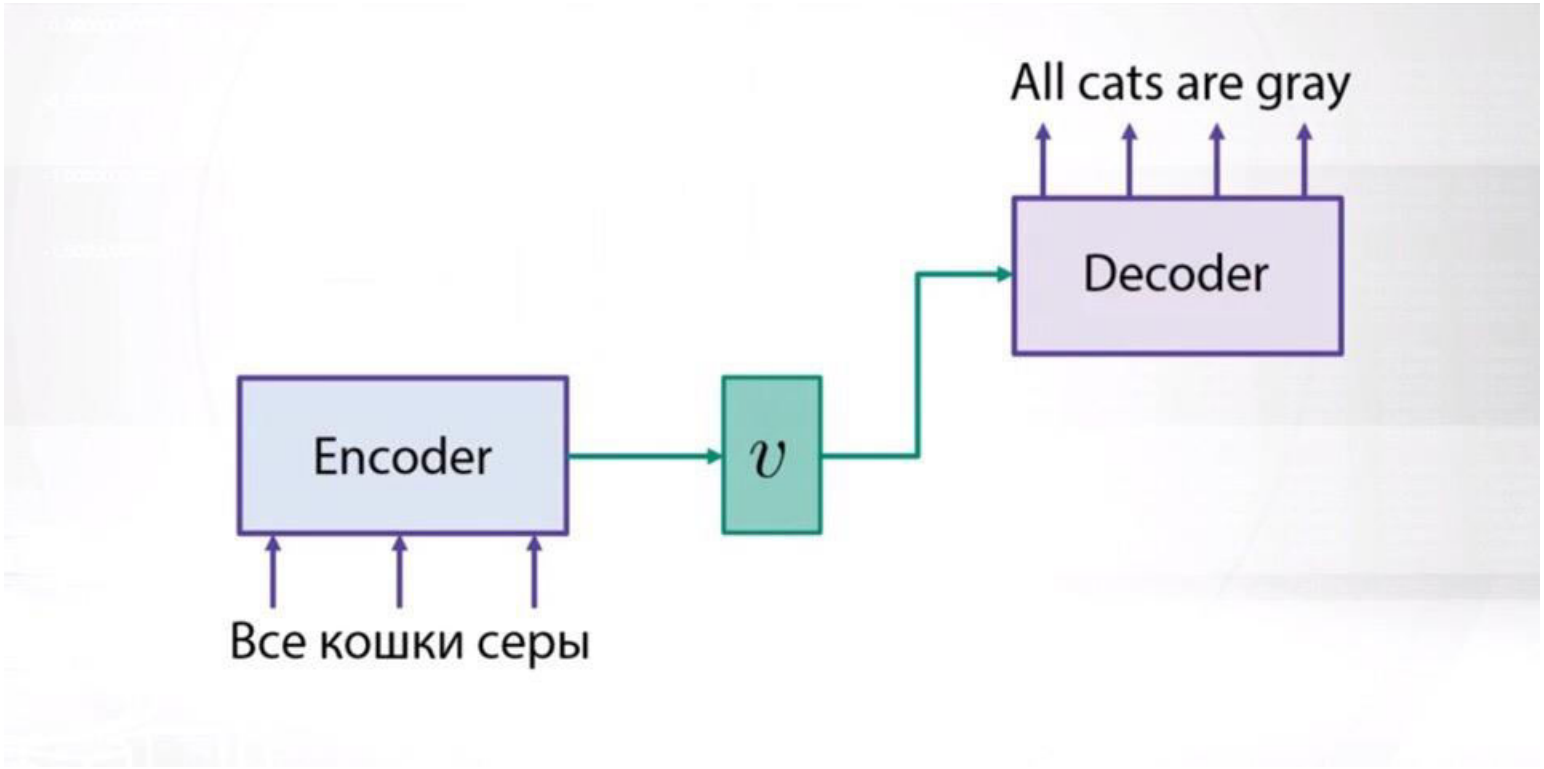
$$p = 0.5\left(\frac{c}{u_m}\right)^3 \qquad M = F_{mean}(1 - p)$$

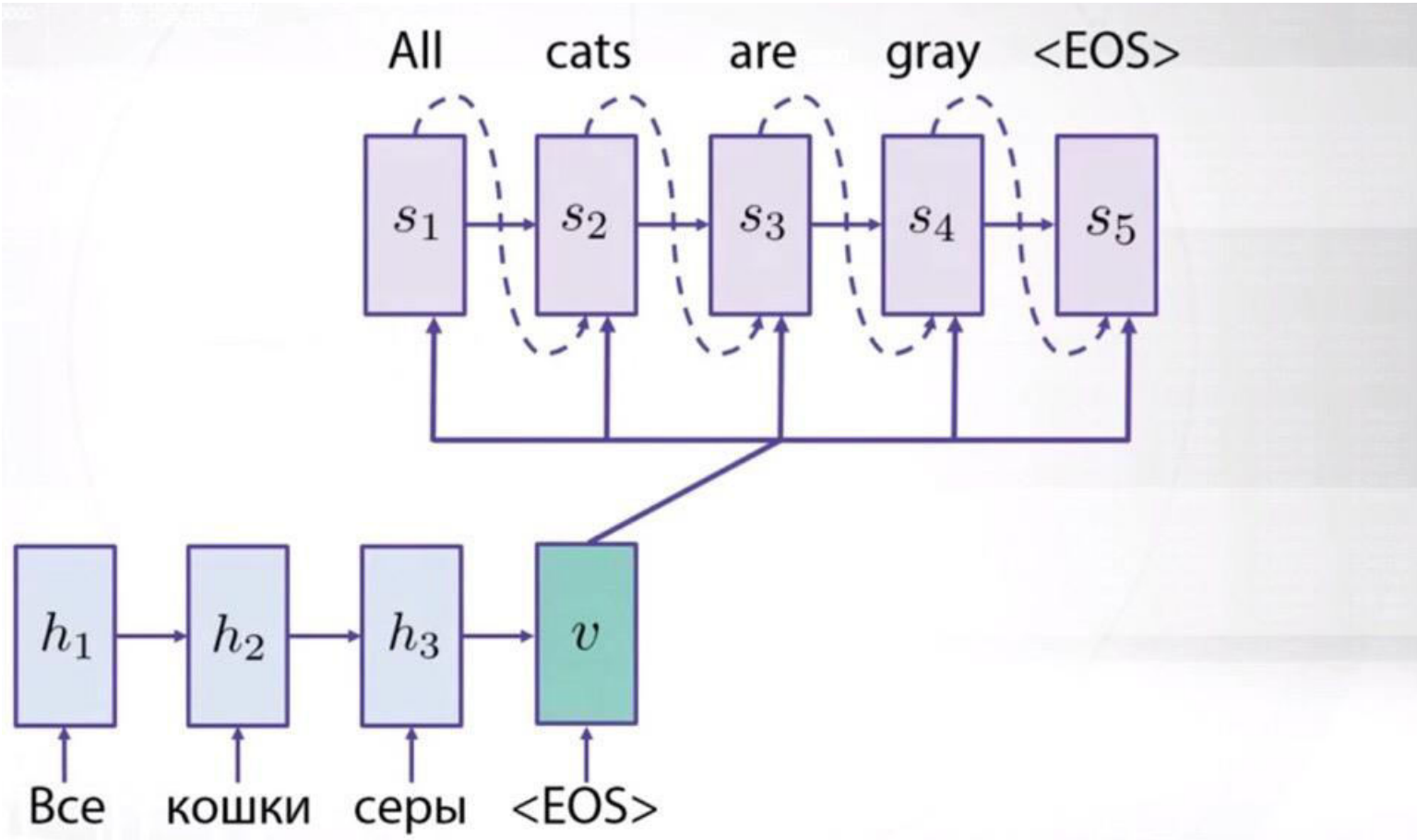
c — число групп n-gram, а u_m — количество n-грамм, которые объединили в группы

- F-measure

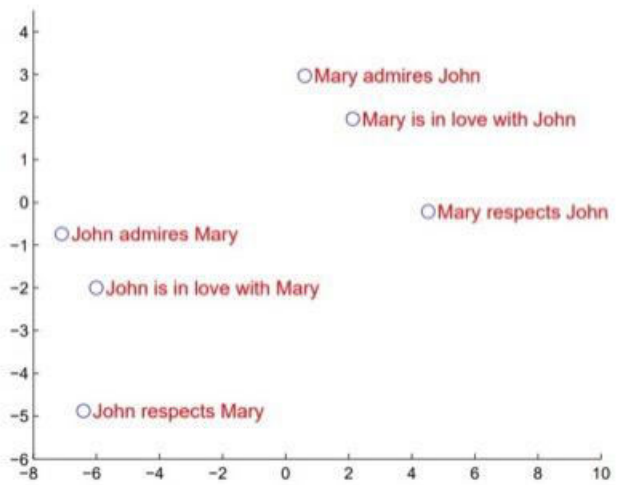
- Precision $\frac{\textit{correct}}{\textit{output-length}} = \frac{3}{6} = 50\%$
- Recall $\frac{\textit{correct}}{\textit{reference-length}} = \frac{3}{7} = 43\%$
- F-measure $\frac{\textit{precision} \times \textit{recall}}{(\textit{precision} + \textit{recall})/2} = \frac{.5 \times .43}{(.5 + .43)/2} = 46\%$

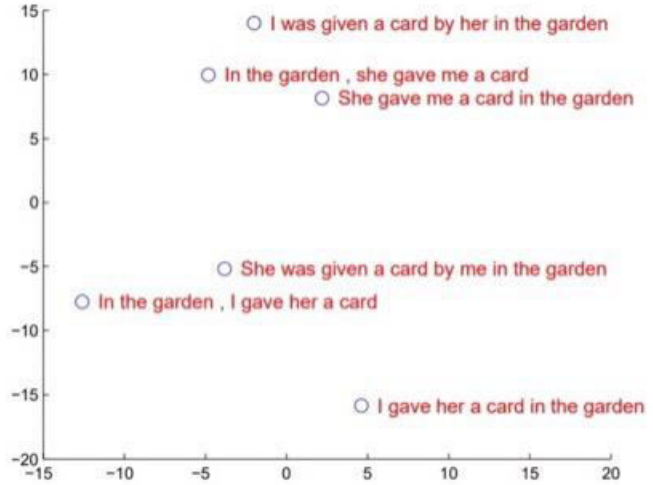
Sequence to sequence



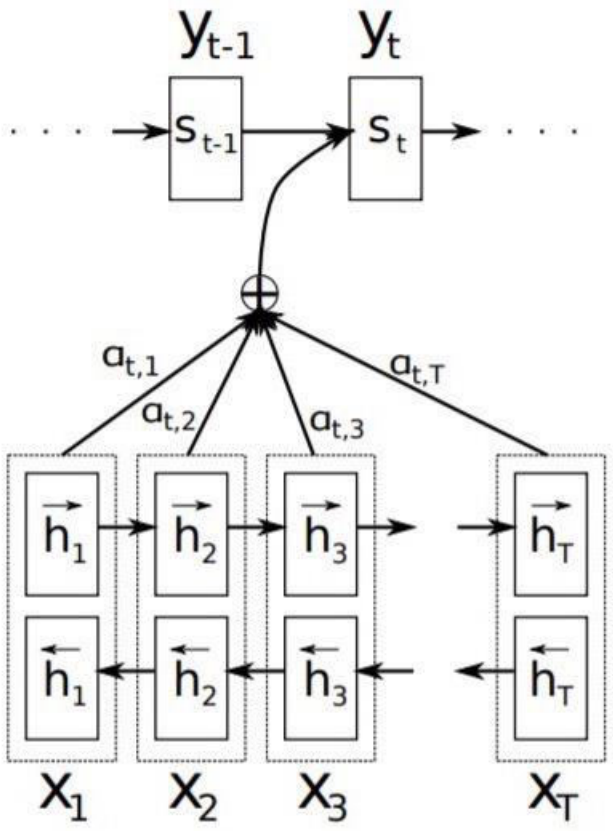


$$s_j = g(s_{j-1}, y_{j-1}, v)$$


Mary admires John
Mary is in love with John
Mary respects John
John admires Mary
John is in love with Mary
John respects Mary


I was given a card by her in the garden
In the garden , she gave me a card
She gave me a card in the garden
She was given a card by me in the garden
In the garden , I gave her a card
I gave her a card in the garden

Attention



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

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

$$e_{ij} = a(s_{i-1}, h_j)$$

$$a(s,h) = s^T h$$

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		👍

Источники

- https://lib.ugent.be/fulltxt/RUG01/002/162/240/RUG01-002162240_2014_0001_AC.pdf
- https://en.wikipedia.org/wiki/Evaluation_of_machine_translation
- https://papers.nips.cc/paper/5346-sequence-to-sequence-learning-with-neural-networks.pdf
- https://www.aclweb.org/anthology/D14-1179
- https://arxiv.org/pdf/1409.0473.pdf
- https://medium.com/syncedreview/history-and-frontier-of-the-neural-machine-translation-dc981d25422d