Нейросетевые методы машинного перевода

Уржумов Александр Евгеньевич

БПМИ-152

4 декабря 2017

Содержание

- Постановка задачи
- Seq2seq и добавления attention
- Архитектура RNN for MT
- Применения модели
- GNMT
- Преимущества и недостатки

Что требуется

- Дано предложение x (на языке 1)
- $y = argmax_y p(y|x)$ (y – предложение на языке 2)
- Обучить RNN (на вход x, на выходе y)

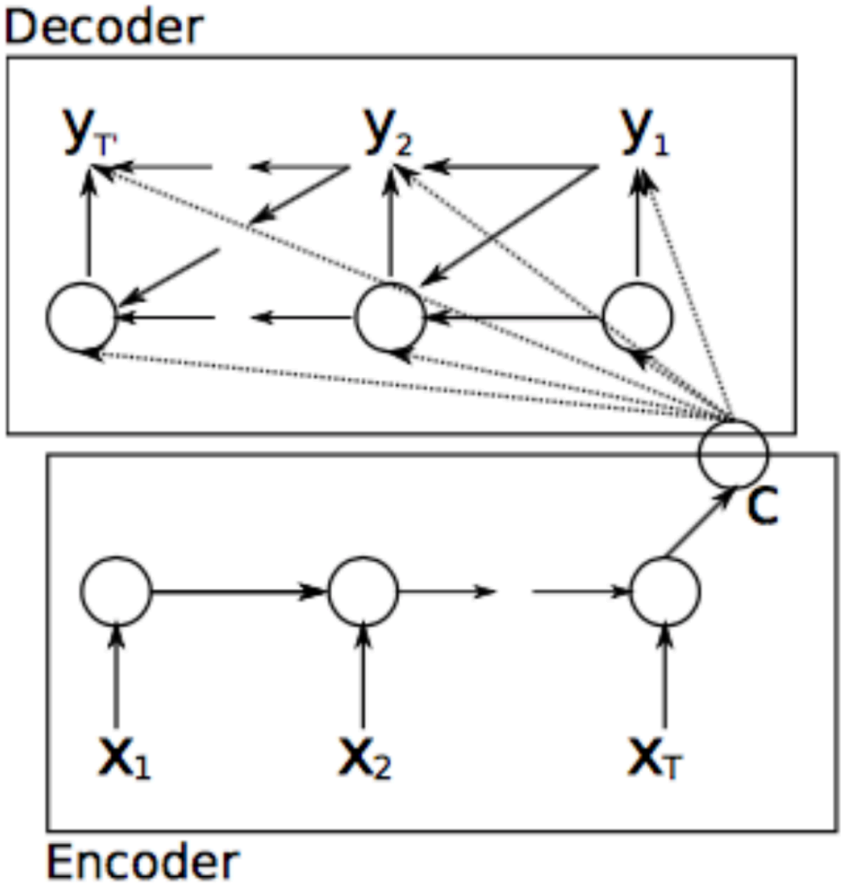
seq2seq

Encoder и decoder совместно обучаются, обученная модель переводит строку x в y за 2 шага:

1) Encoder переводит x в C
2) Decoder переводит C в y

$h_i = f(x_i, h_{i-1})$ – *скрытые состояния Encoder*
$c = q(\{h_1, \dots, h_T\})$
$s_i = f(s_{i-1}, y_{i-1}, c)$ – *скрытые состояния Decoder*
$p(y_i|y_1, \dots, y_{i-1}, x) = g(y_{i-1}, s_i, c)$

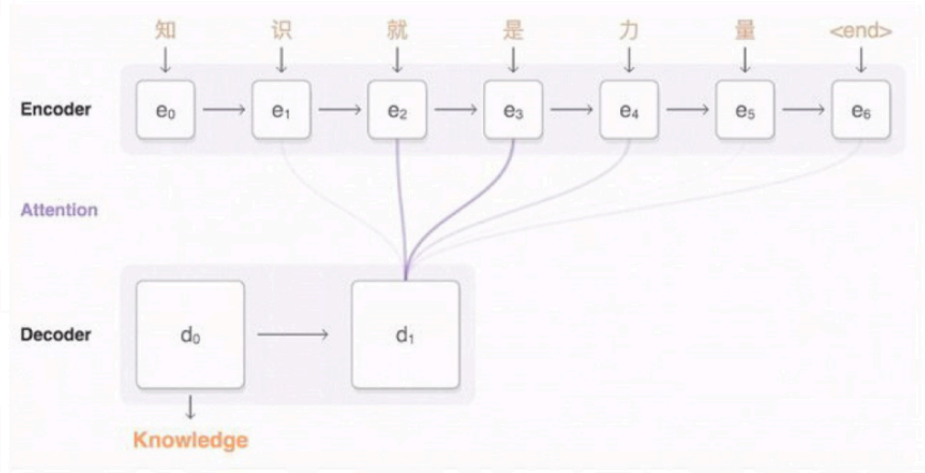
Основная проблема c - вектор фиксированной длины и одинаковый для всех слов



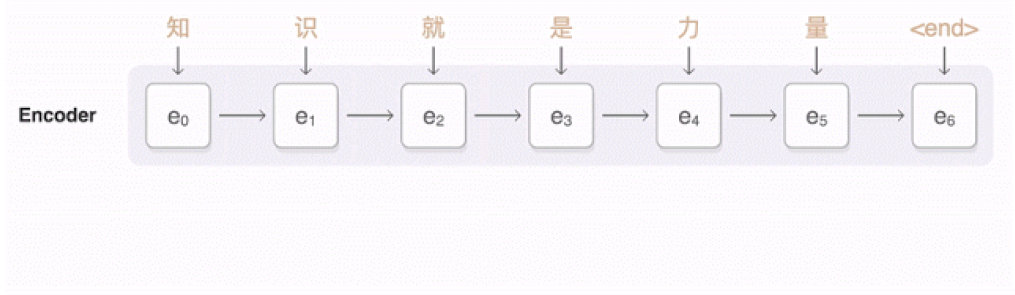
Добавим внимание

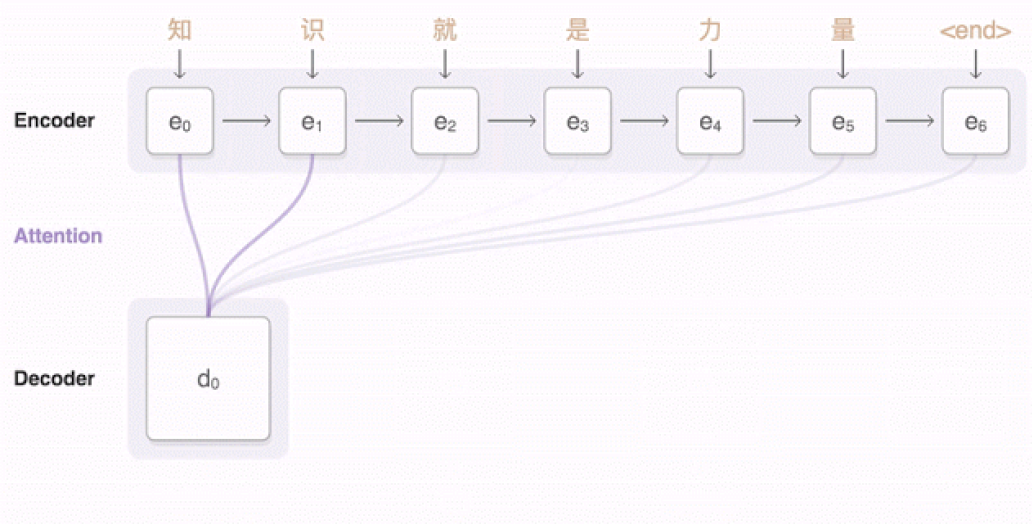
Смотрим не на выход encoder,
а на взвешенную сумму
промежуточных состояний
$c_i = \sum_{j=1}^{T_x} \alpha_{ij} h_j$ - взвешенная
сумма

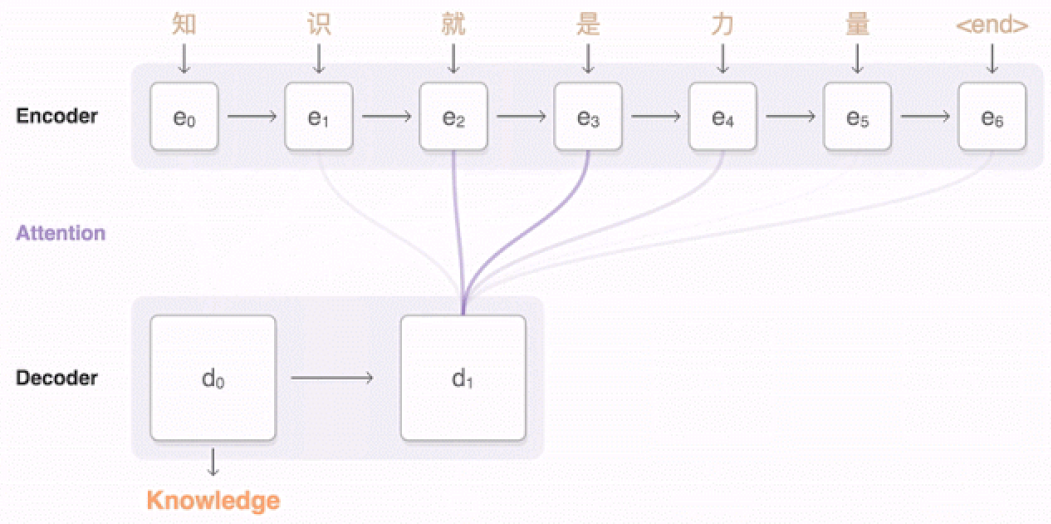
Encoder-Decoder with Attention

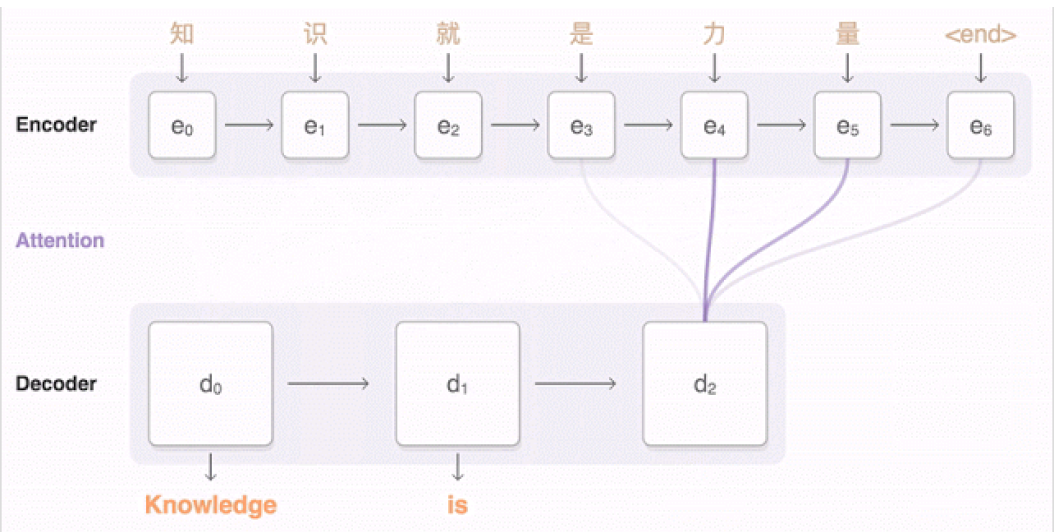


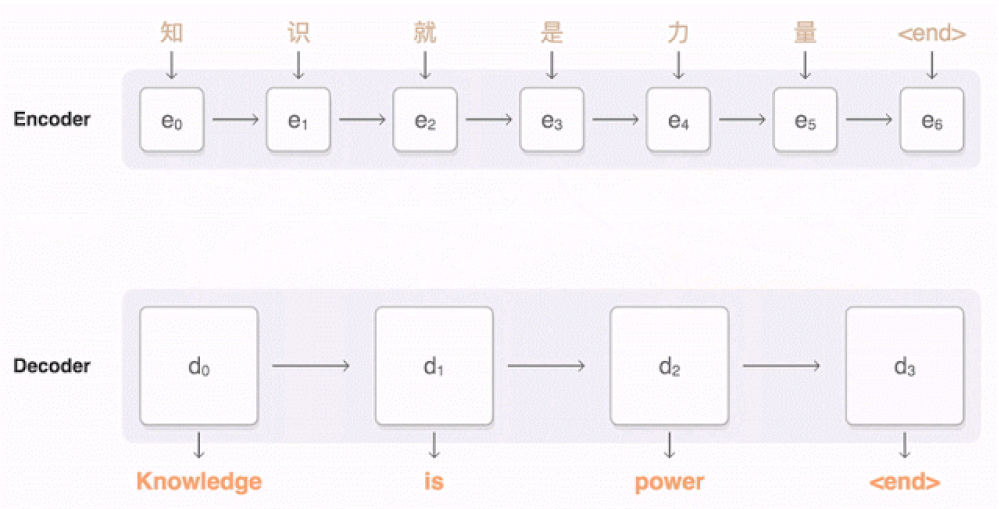

知
识
就
是
力
量
<end>


知
识
就
是
力
量
<end>
Encoder
e0
e1
e2
e3
e4
e5
e6


知
识
就
是
力
量
<end>
Encoder
e0
e1
e2
e3
e4
e5
e6
Attention
Decoder
d0


知
识
就
是
力
量
<end>
Encoder
e0
e1
e2
e3
e4
e5
e6
Attention
Decoder
d0
d1
Knowledge


知
识
就
是
力
量
<end>
Encoder
e0
e1
e2
e3
e4
e5
e6
Attention
Decoder
d0
d1
d2
Knowledge
is


知
识
就
是
力
量
<end>
Encoder
e0
e1
e2
e3
e4
e5
e6
Decoder
d0
d1
d2
d3
Knowledge
is
power
<end>

RNN Encoder-Decoder with attention

Encoder - BiRNN

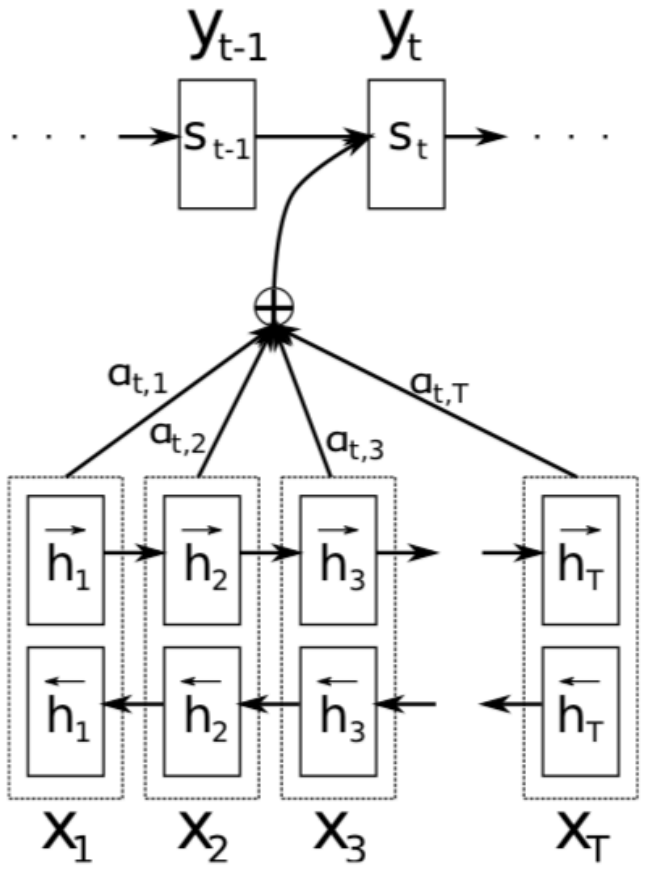
$x = (x_1, \dots, x_{T_x})$

$\overrightarrow{h_j} = f(x_j, h_{j-1}) \in \mathbb{R}^n$-скрытые состояния – forward RNN

$\overleftarrow{h_j} = f(x_j, h_{j+1}) \in \mathbb{R}^n$-скрытые состояния - backward RNN

$$h_j = [\overrightarrow{h_j}^T ; \overleftarrow{h_j}^T]^T$$

$h_j -$ учитываются соседи слева и справа



RNN Encoder-Decoder with attention

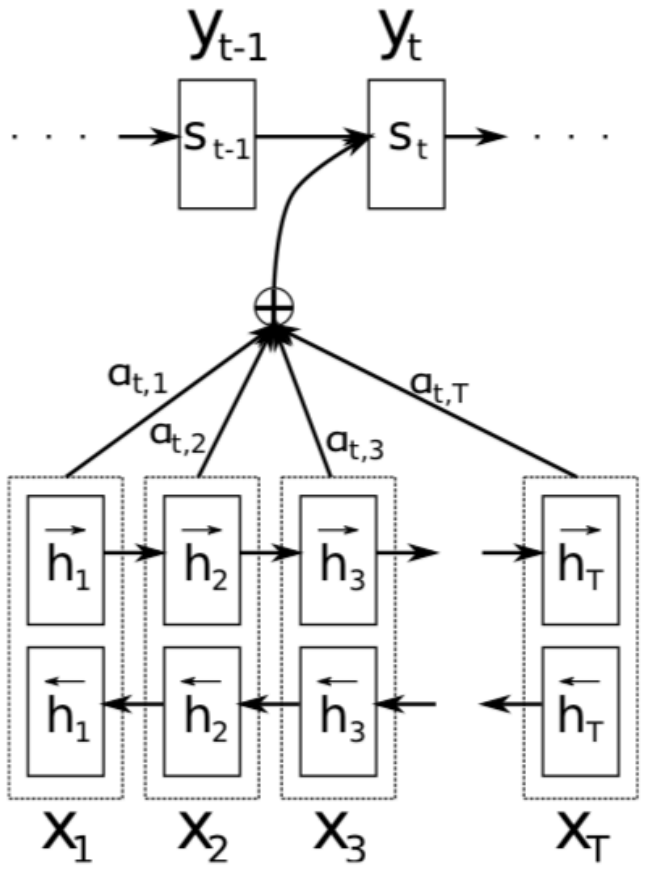
Decoder

$p(y_i|y_1, \dots, y_{i-1}, x) = g(y_{i-1}, s_i, c_i)$
$s_i = f(s_{i-1}, y_{i-1}, c_i)$

$c_i = \sum_{j=1}^{T_x} \alpha_{ij} h_j$ - взвешенная сумма

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

$e_{ij} = a(s_{i-1}, h_j)$ – “энергия похожести” j-ого слова из x и i-ого слова из y



Применение на модели на реальных данных

English-to-French translation task, bilingual parallel corpora ACL WMT ’14 348M слов.

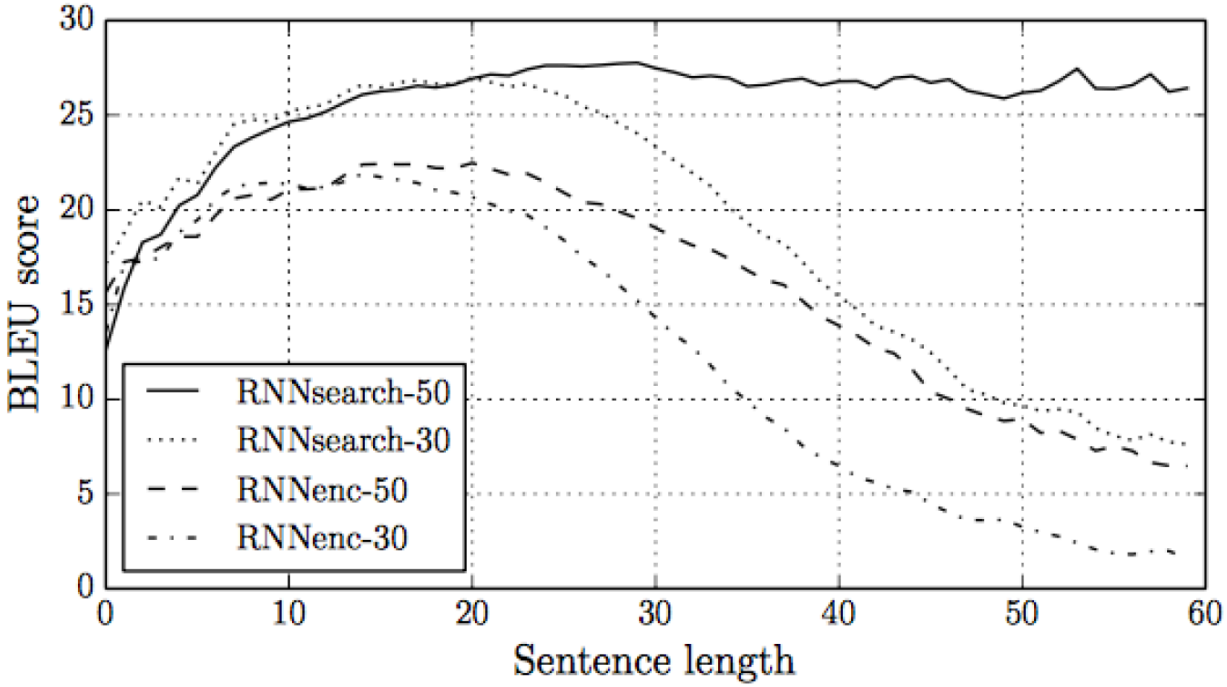
Сравним 2 модели:

1) RNNenc – без attention

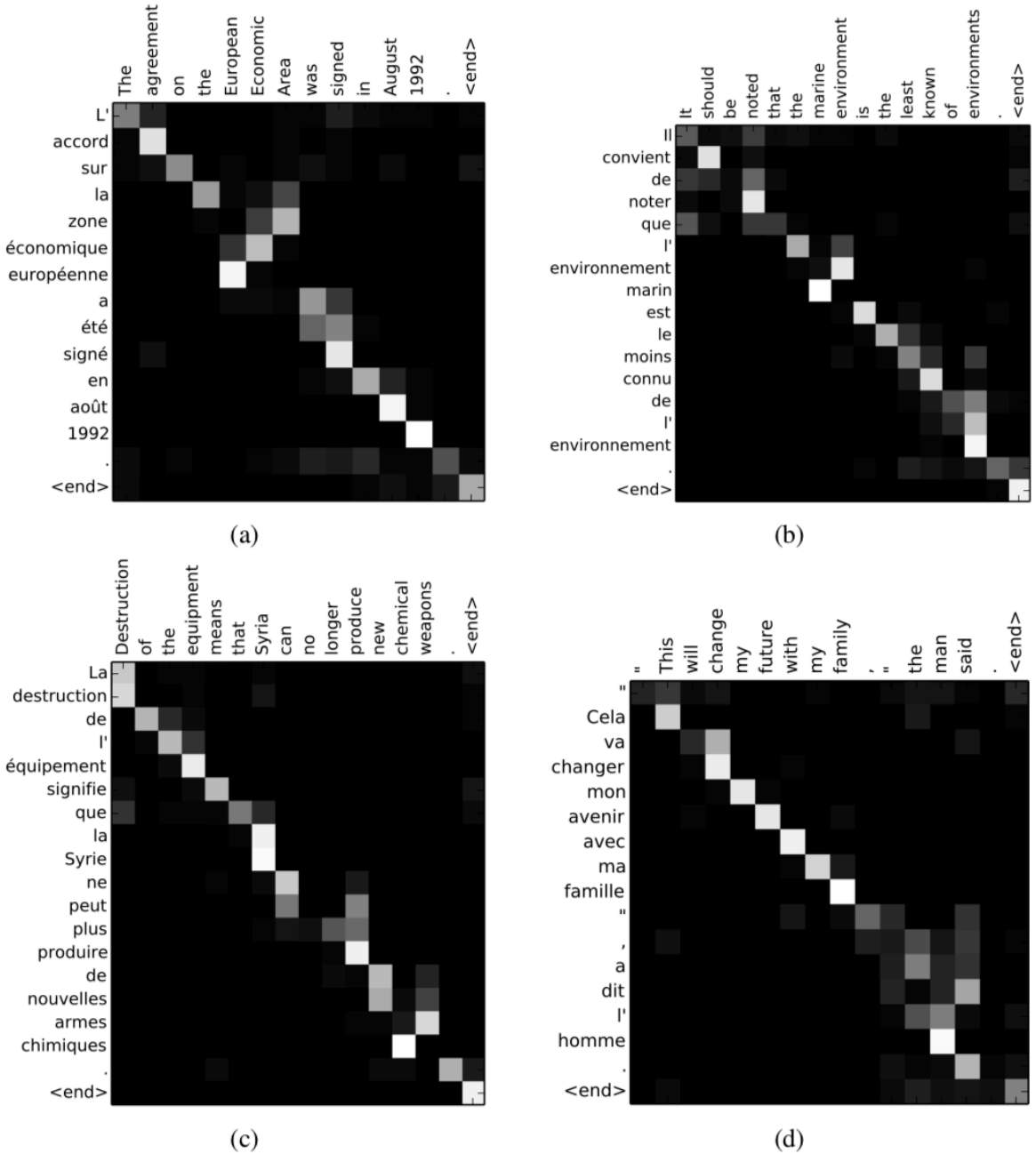
2) RNNsearch – с attention

По 1000 юнитов в каждой

SGD + Adadelta для обучения

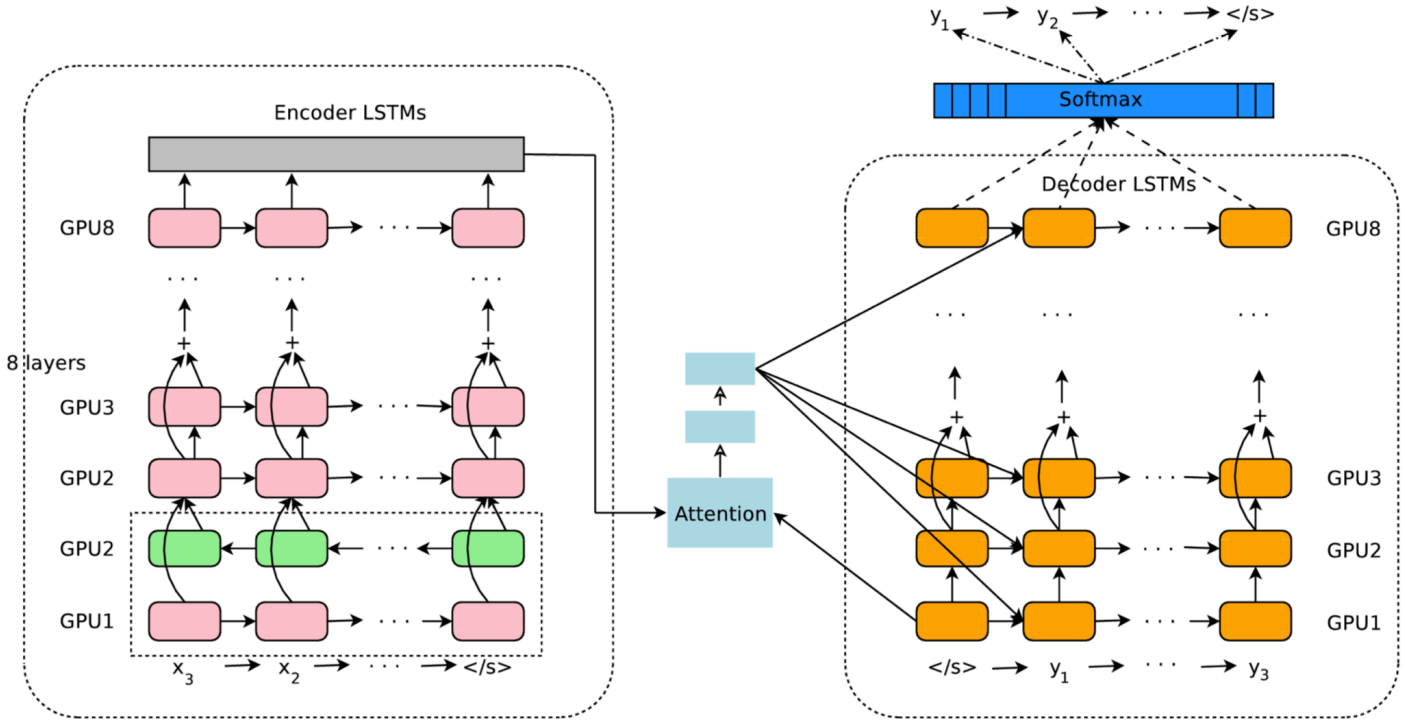


Соответствия между переводом и исходным текстом.
ij-ая ячейка визуализирует α_{ij} в шкале [0,1]



Применения NMT в индустрии (GNMT)

- Более сложная архитектура, работающая с большими объемами данных
- Существенные прогресс на простых предложениях
- В большинстве задач, но не везде, лучше статистических методов
- Качество зависит от исходного языка и языка на который переводят



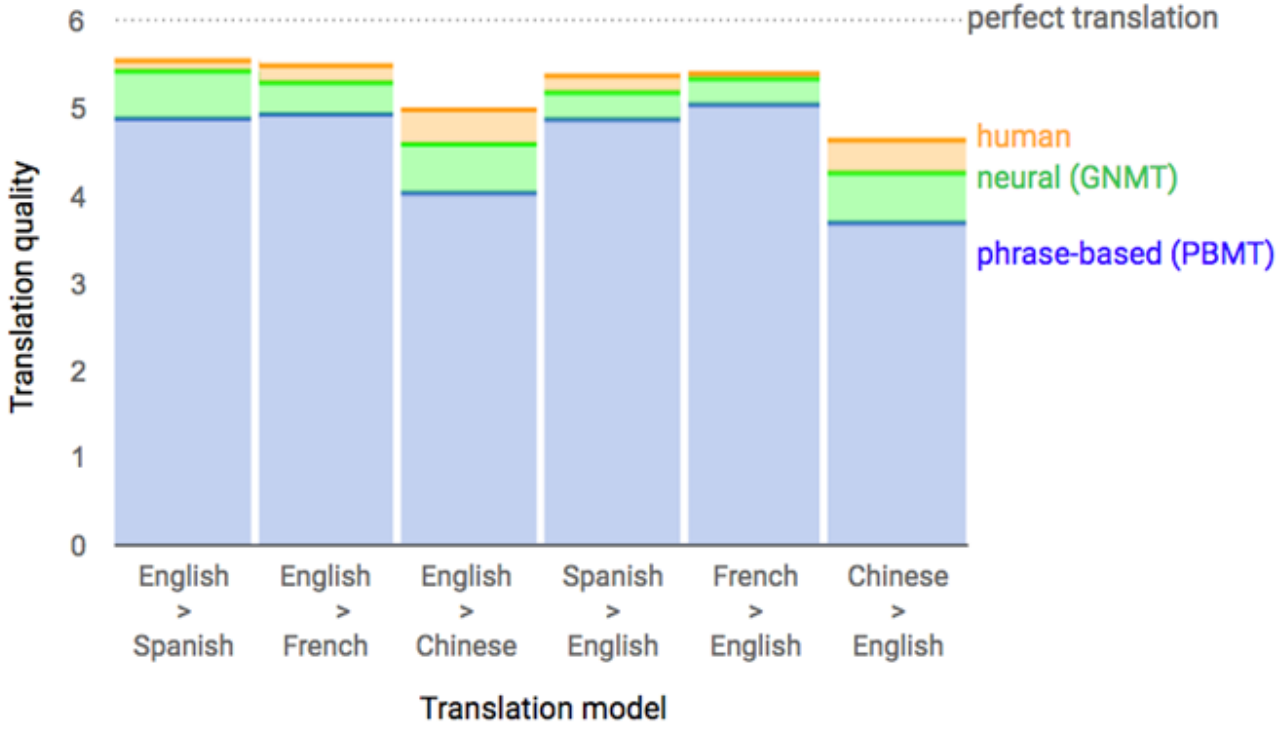
Time	Company	NMT Implementation	Framework	Characteristic
2016.09	Google	GNMT	Tensorflow	State of the art industrial implementation of the "attentional encoder-decoder networks" model
2016.11	Microsoft	No technical details disclosed		
2017.05	Facebook	Fairseq	Torch	Used CNN to replace RNN
2017.06	Google	Transformer	Tensorflow/ Tensor2Tensor	Solely attention based NMT
2017.07	Amazon	Sockeye	MXNet	

Other companies took part in the NMT R&D include: IBM, NVIDIA, SYSTRAN

In China: Baidu, NetEase-Youdao, Tencent, Sogou, iFlytek, Alibaba

Сравнение качества

	PBMT	GNMT	Human	Relative Improvement
English→Spanish	4.885	5.428	5.55	87%
English→French	4.932	5.295	5.496	64%
English→Chinese	4.035	4.594	4.987	58%
Spanish →English	4.872	5.187	5.372	63%
French →English	5.046	5.343	5.404	83%
Chinese → English	3.694	4.263	4.636	60%



Плюсы и минусы

+) высокий blue

+) качество не падает для длинных предложений

+) в общем случае работает лучше

-) для обучения требуется большой объем данных

-) время обучения

-) плохая поддержка специфической терминологии(научной, бизнес и т.д.)

-) подходит не для всех задач

Input sentence:	***Translation (PBMT):***	***Translation (GNMT):***	***Translation (human):***
李克強此行將啟動中加總理年度對話機制，與加拿大總理杜魯多舉行兩國總理首次年度對話。	Li Keqiang premier added this line to start the annual dialogue mechanism with the Canadian Prime Minister Trudeau two prime ministers held its first annual session.	Li Keqiang will start the annual dialogue mechanism with Prime Minister Trudeau of Canada and hold the first annual dialogue between the two premiers.	Li Keqiang will initiate the annual dialogue mechanism between premiers of China and Canada during this visit, and hold the first annual dialogue with Premier Trudeau of Canada.

NMT хорошо работает

- Быстрый перевод слов и коротких текстов для различных целей;
- Автоматический перевод в процессе общения на форумах, в социальных сетях, мессенджерах;
- Автоматический перевод при чтении новостей, статей Wikipedia;
- Переводчик в путешествиях (mobile).

NMT плохо работает

- Перевод деловой переписки с клиентами, партнерами, инвесторами, иностранными сотрудниками;
- Локализация сайтов, интернет-магазинов, описаний продуктов, инструкций;
- Перевод пользовательского контента (отзывы, форумы, блоги);
- Возможность интеграции перевода в бизнес-процессы и программные продукты и сервисы;
- Точность перевода с соблюдением терминологии, конфиденциальность и безопасность.

Выводы

• В общем случае нейронный автоматический перевод дает результат более высокого качества, чем «чисто» статистический подход;

• Автоматический перевод через нейронную сеть – лучше подходит для решения задачи «универсального перевода»;

• Ни один из подходов к МП сам по себе не является идеальным универсальным инструментом для решения любой задачи перевода;

Источники

- https://arxiv.org/pdf/1409.0473.pdf
- https://arxiv.org/pdf/1609.08144.pdf
- https://research.googleblog.com/2016/09/a-neural-network-for-machine.html
- https://habrahabr.ru/company/oleg-bunin/blog/340184/
- https://habrahabr.ru/post/330654/
- https://medium.com/@Synced/history-and-frontier-of-the-neural-machine-translation-dc981d25422d