ОСНОВЫ ОБРАБОТКИ ТЕКСТОВ

Лекция №7
Машинный перевод

Лектор: н.с. ИСП РАН Сысоев Андрей Анатольевич

31 октября 2018

План лекции

- Постановка задачи
- Наивный подход
- Статистический машинный перевод
- Нейросетевой машинный перевод
- Оценка качества

Задача

Перевести текст с исходного языка на целевой.

Задача

Перевести текст с исходного языка на целевой.

Blankadas velo unusola
En la nebula mara blu'

nuq nej ghaH qaStaHvIS
machchugh Hop Ha'?

¿Qué tiró en el filo de los suyos?

Задача

Перевести текст с исходного языка на целевой.

Blankadas velo unusola
En la nebula mara blu'

(эсперанто)

Белеет парус одинокий
В тумане моря голубом!

nuq nej ghaH qaStaHvIS
machchugh Hop Ha'?

(клингонский)

Что ищет он в стране далекой?

¿Qué tiró en el filo de los suyos?

(испанский)

Что кинул он в краю родном?

Машинный перевод

Первые шаги

Джорджтаунский эксперимент (7 января 1954) - перевод более 60 предложений с русского на английский.

Александр Пиперски
«Машинный перевод: 60 лет в погоне за совершенством»
https://www.youtube.com/watch?v=AFr7KC2-JDM

Машинный перевод

Первые шаги

Джорджтаунский эксперимент (7 января 1954) - перевод более 60 предложений с русского на английский.

Vyelyichyina ugla opryedyelyayetsya otnoshyenyiyem dlyini dugi k radyiusu.

Величина угла определяется отношением длины дуги к радиусу.

Magnitude of angle is determined by the relation of length of arc to radius.

Мы передаем мысли посредством речи.

We transmit thoughts by means of speech.

Александр Пиперски
«Машинный перевод: 60 лет в погоне за совершенством»
https://www.youtube.com/watch?v=AFr7KC2-JDM

Машинный перевод

Наивный подход

- перевод по словарю
- лингвистические правила (время, число, ...)

Машинный перевод

Наивный подход

- перевод по словарю
- лингвистические правила (время, число, ...)

Примеры результатов:

Голый кондуктор бежит под вагоном.

Космический бар прессы продолжит работу.

Машинный перевод

Наивный подход

- перевод по словарю
- лингвистические правила (время, число, ...)

Примеры результатов:

Голый кондуктор бежит под вагоном.

The naked conductor runs under the carriage.

Оголённый провод проходит под тележкой.

Космический бар прессы продолжит работу.

Press space bar to continue operation.

Нажми на „пробел“ чтобы продолжить действие.

Машинный перевод

Наивный подход

- перевод по словарю
- лингвистические правила (время, число, ...)

Примеры результатов:

Голый кондуктор бежит под вагоном.

The naked conductor runs under the carriage.

Оголённый провод проходит под тележкой.

Космический бар прессы продолжит работу.

Press space bar to continue operation.

Нажми на „пробел“ чтобы продолжить действие.

Ясные печеньки. - *Clear cookies.*

Просто казните монтажника. - *Just execute the installer.*

Машинный перевод

Наивный подход

Недостатки:

- нужен предметно-специфичный словарь
- лингвистические правила разрабатываются вручную
- не работает в более широкой предметной области (когда возникают неоднозначности)

Статистический машинный перевод

Источники параллельных корпусов

- законодательство (ООН, Европейский союз, Канада)
- сайты (обычно новостные) на нескольких языках
- переводы художественной литературы

Задача статистического машинного перевода

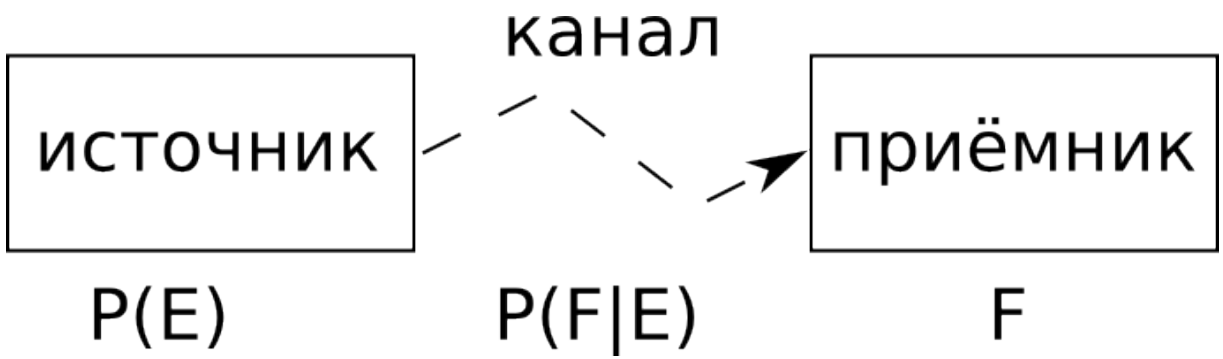
Задача:
перевести предложение F на языке $\mathcal{F}$ в предложение E на языке $\mathcal{E}$

$$\hat{E} = \underset{E}{\operatorname{argmax}}\, P(E|F) = \underset{E}{\operatorname{argmax}}\, \frac{P(F|E)P(E)}{P(F)} = \underset{E}{\operatorname{argmax}}\, P(F|E)P(E)$$

- $P(E)$ - языковая модель
- $P(F|E)$ - модель перевода

Модель зашумленного канала

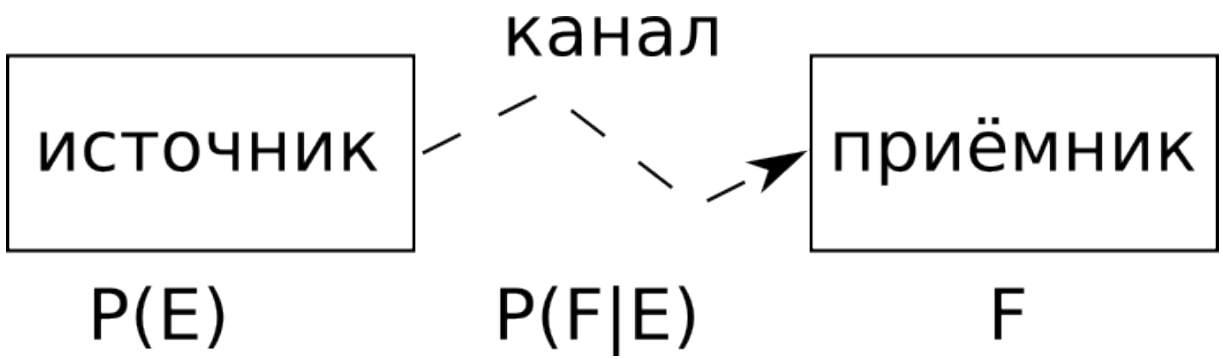
$$\hat{E} = \underset{E}{\operatorname{argmax}}\, P(F|E)P(E)$$



По сообщению F, полученному из зашумлённого канала, нужно восстановить оригинальное сообщение E.

Модель зашумленного канала

$$\hat{E} = \underset{E}{\operatorname{argmax}}\, P(F|E)P(E)$$



По сообщению F, полученному из зашумлённого канала, нужно восстановить оригинальное сообщение E.

Подзадачи:

- $P(E)$ - языковая модель
- $P(F|E)$ - модель перевода
- способ декодирования

Языковая модель

Языковая модель

Предложение $E = (e_1, e_2, ..., e_n)$

$$P(E) = P(e_1, e_2, ..., e_n)$$

Определение условной вероятности

$$P(e_n|e_1, e_2, ...e_{n-1}) = \frac{P(e_1, e_2, ..., e_n)}{P(e_1, e_2, ..., e_{n-1})}$$

Языковая модель

Предложение $E = (e_1, e_2, ..., e_n)$

$$P(E) = P(e_1, e_2, ..., e_n) = P(e_1, e_2, ..., e_{n-1})P(e_n|e_1, e_2, ...e_{n-1}) =$$
$$= ... = P(e_1)P(e_2|e_1)P(e_3|e_1, e_2)...P(e_n|e_1, e_2, ...e_{n-1})$$

Определение условной вероятности

$$P(e_n|e_1, e_2, ...e_{n-1}) = \frac{P(e_1, e_2, ..., e_n)}{P(e_1, e_2, ..., e_{n-1})}$$

Языковая модель

N-граммы

$$P(E) = P(e_1)P(e_2|e_1)P(e_3|e_1, e_2)...P(e_n|e_1, e_2, ..., e_{n-1})$$

Языковая модель

N-граммы

$$P(E) = P(e_1)P(e_2|e_1)P(e_3|e_1, e_2)...P(e_n|e_1, e_2, ..., e_{n-1})$$

Марковское предположение: e_i зависит только от k предыдущих слов: $e_{i-k}, ..., e_{i-1}$

$$P(E) = \prod_i P(e_i|e_{i-k}, ..., e_{i-1})$$

Языковая модель

N-граммы

$$P(E) = P(e_1)P(e_2|e_1)P(e_3|e_1, e_2)...P(e_n|e_1, e_2, ..., e_{n-1})$$

Марковское предположение: e_i зависит только от k предыдущих слов: $e_{i-k}, ..., e_{i-1}$

$$P(E) = \prod_i P(e_i|e_{i-k}, ..., e_{i-1})$$

Это и есть N-граммная языковая модель ($N = k + 1$).

Языковая модель

Как оценить $P(e_i|e_{i-k}, ..., e_{i-1})$?

Языковая модель

Как оценить $P(e_i|e_{i-k}, ..., e_{i-1})$?

- Оценка по частоте в корпусе

$$P(e_i|e_{i-k}, ..., e_{i-1}) = \frac{count(e_{i-k}, ..., e_{i-1}, e_i)}{count(e_{i-k}, ..., e_{i-1})}$$

Языковая модель

Как оценить $P(e_i|e_{i-k}, ..., e_{i-1})$?

- Оценка по частоте в корпусе

$$P(e_i|e_{i-k}, ..., e_{i-1}) = \frac{count(e_{i-k}, ..., e_{i-1}, e_i)}{count(e_{i-k}, ..., e_{i-1})}$$

- Аддитивное сглаживание (Лапласа, Лидстоуна);
  V - количество слов в словаре

$$P(e_i|e_{i-k}, ..., e_{i-1}) = \frac{\delta + count(e_{i-k}, ..., e_{i-1}, e_i)}{\delta V + count(e_{i-k}, ..., e_{i-1})}$$

Языковая модель

Как оценить $P(e_i|e_{i-k}, ..., e_{i-1})$?

- Оценка по частоте в корпусе

$$P(e_i|e_{i-k}, ..., e_{i-1}) = \frac{count(e_{i-k}, ..., e_{i-1}, e_i)}{count(e_{i-k}, ..., e_{i-1})}$$

- Аддитивное сглаживание (Лапласа, Лидстоуна);
  V - количество слов в словаре

$$P(e_i|e_{i-k}, ..., e_{i-1}) = \frac{\delta + count(e_{i-k}, ..., e_{i-1}, e_i)}{\delta V + count(e_{i-k}, ..., e_{i-1})}$$

- Линейная интерполяция

$$P'(e_i|e_{i-k}, ..., e_{i-1}) = \alpha_{k+1} P(e_i|e_{i-k}, ..., e_{i-1}) + ... + \\ + \alpha_2 P(e_i|e_{i-1}) + \alpha_1 P(e_i) + \alpha_0$$

$$\sum_j \alpha_j = 1 \qquad \alpha_{j+1} > \alpha_j$$

Модель перевода

Модель перевода

Исходная задача:
перевести предложение F на языке $\mathcal{F}$ в предложение E на языке $\mathcal{E}$.

В модели перевода все наоборот: нужно вычислить вероятность $P(F|E)$ получения F из оригинала E.

Модель перевода

Может быть,

$$P(F|E) = \frac{count(F, E)}{count(E)}$$

Модель перевода

Может быть,

$$P(F|E) = \frac{count(F, E)}{count(E)}$$

Но это не будет работать (см **Языковые модели**).

Модель перевода

Может быть,

$$P(F|E) = \frac{count(F, E)}{count(E)}$$

Но это не будет работать (см **Языковые модели**).

Идея:

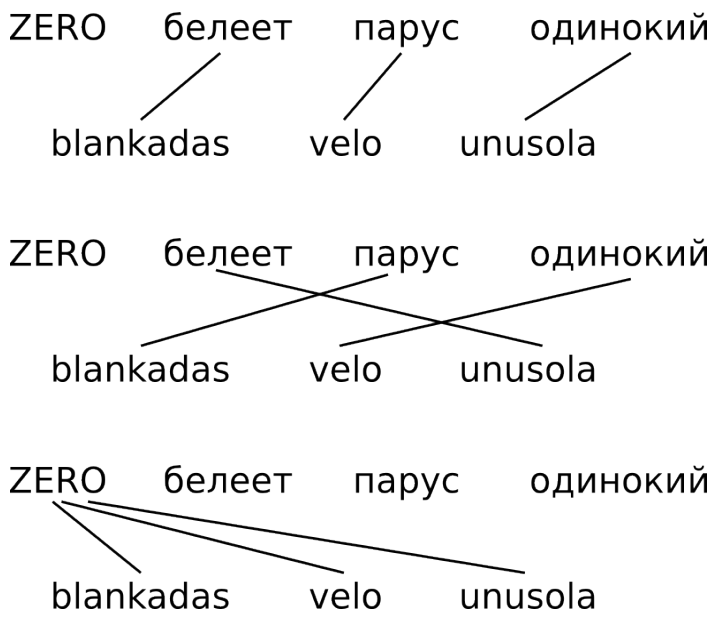
разделить F и E на фрагменты (токены)

вычислять $P(F|E)$ на основе вероятностей сопоставления фрагментов

IBM Model 1

IBM Model 1

Выравнивание



Каждому токену f_i соответствует только однин токен e_j или *ZERO*-токен (мы будем считать, что это e_0).

IBM Model 1

Используем выравнивания α для вычисления $P(F|E)$:

$$P(F|E) = \sum_{\alpha} P(F, \alpha|E)$$

IBM Model 1

Используем выравнивания α для вычисления $P(F|E)$:

$$P(F|E) = \sum_{\alpha} P(F, \alpha|E)$$

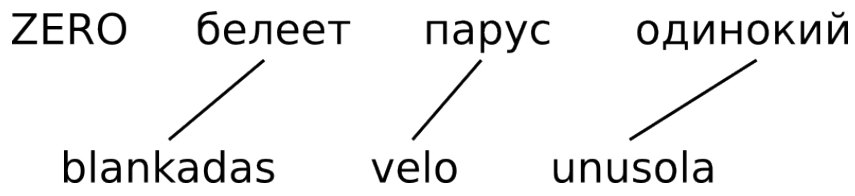
В IBM Model 1 $P(F, \alpha|E)$ имеет вид:

$$P(F, \alpha|E) = \frac{\epsilon}{(|E|+1)^{|F|}} \prod_{j=1}^{|F|} t(f_j|e_{\alpha(j)})$$

- все выравнивания α равновероятны;
- ϵ - нормирующий коэффициент.

IBM Model 1

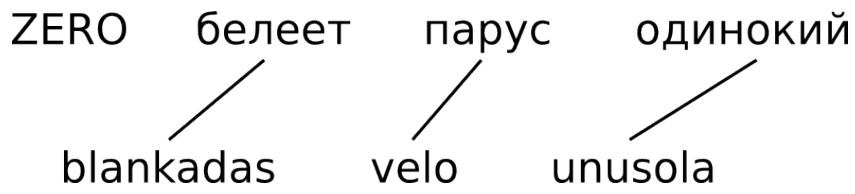
$P(\cdot|ZERO) \longrightarrow$ ..., blankadas: 0.05, ..., velo: 0.02, ..., unusola: 0.01
$P(\cdot|белеет) \longrightarrow$ ..., blankadas: 0.1, ..., velo: 0.01, ..., unusola: 0.005
$P(\cdot|парус) \longrightarrow$ velo: 0.1, ..., blankadas: 0.003, ..., unusola: 0.002
$P(\cdot|одинокий) \longrightarrow$ unusola: 0.07, ..., velo: 0.04, ..., blankadas: 0.03, ...



$$P(F, \alpha|E) =$$

IBM Model 1

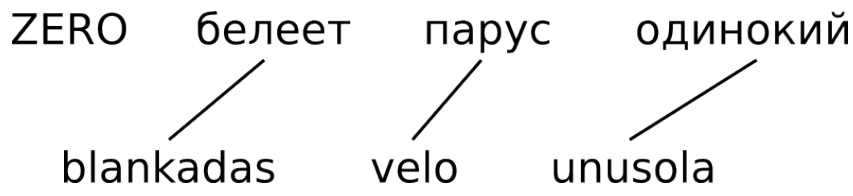
$P(\cdot|ZERO) \longrightarrow$ ..., blankadas: 0.05, ..., velo: 0.02, ..., unusola: 0.01
$P(\cdot|белеет) \longrightarrow$ ..., blankadas: 0.1, ..., velo: 0.01, ..., unusola: 0.005
$P(\cdot|парус) \longrightarrow$ velo: 0.1, ..., blankadas: 0.003, ..., unusola: 0.002
$P(\cdot|одинокий) \longrightarrow$ unusola: 0.07, ..., velo: 0.04, ..., blankadas: 0.03, ...



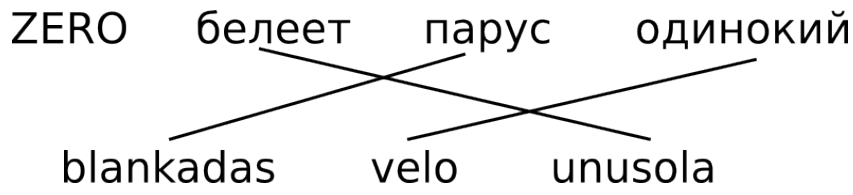
$$P(F, \alpha|E) = \frac{\epsilon}{4^3} \times 0.1 \times 0.1 \times 0.07 = 0.000010937\epsilon$$

IBM Model 1

$P(\cdot|\textit{ZERO}) \longrightarrow$ ..., blankadas: 0.05, ..., velo: 0.02, ..., unusola: 0.01
$P(\cdot|\textit{белеет}) \longrightarrow$ ..., blankadas: 0.1, ..., velo: 0.01, ..., unusola: 0.005
$P(\cdot|\textit{парус}) \longrightarrow$ velo: 0.1, ..., blankadas: 0.003, ..., unusola: 0.002
$P(\cdot|\textit{одинокий}) \longrightarrow$ unusola: 0.07, ..., velo: 0.04, ..., blankadas: 0.03, ...



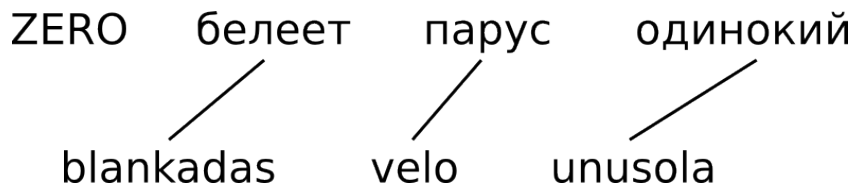
$$P(F, \alpha|E) = \frac{\epsilon}{4^3} \times 0.1 \times 0.1 \times 0.07 = 0.000010937\epsilon$$



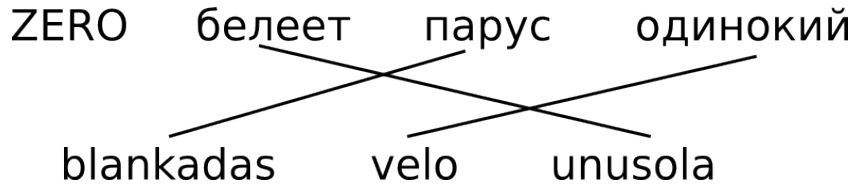
$$P(F, \alpha|E) =$$

IBM Model 1

$P(\cdot|\textit{ZERO}) \longrightarrow$ ..., blankadas: 0.05, ..., velo: 0.02, ..., unusola: 0.01
$P(\cdot|\textit{белеет}) \longrightarrow$ ..., blankadas: 0.1, ..., velo: 0.01, ..., unusola: 0.005
$P(\cdot|\textit{парус}) \longrightarrow$ velo: 0.1, ..., blankadas: 0.003, ..., unusola: 0.002
$P(\cdot|\textit{одинокий}) \longrightarrow$ unusola: 0.07, ..., velo: 0.04, ..., blankadas: 0.03, ...



$$P(F, \alpha|E) = \frac{\epsilon}{4^3} \times 0.1 \times 0.1 \times 0.07 = 0.000010937\epsilon$$



$$P(F, \alpha|E) = \frac{\epsilon}{4^3} \times 0.005 \times 0.003 \times 0.04 = 0.000000009\epsilon$$

IBM Model 1

Пусть есть корпус пар предложений

$$\mathfrak{C} = \{(F_1, E_1), (F_2, E_2), ..., (F_m, E_m)\}$$

Нужно максимизировать правдоподобие:

$$P(\mathfrak{C}) = \prod_{i=1}^{m} P(F_i|E_i)$$

$$P(F|E) = \sum_{\alpha} P(F, \alpha|E)$$

$$P(F, \alpha|E) = \frac{\epsilon}{(|E|+1)^{|F|}} \prod_{j=1}^{|F|} t(f_j|e_{\alpha(j)})$$

IBM Model 1

EM-алгоритм

- нам нужно оценить параметры $t(f|e)$ на основе параллельного корпуса,
- ... но мы не знаем правильные выравнивания!

IBM Model 1

EM-алгоритм

- нам нужно оценить параметры $t(f|e)$ на основе параллельного корпуса,
- ... но мы не знаем правильные выравнивания!

Проблема курицы и яйца:

- если бы мы знали правильные выравнивания,
  то мы могли бы оценить параметры $t(f|e)$.
- если бы мы знали параметры $t(f|e)$,
  то мы могли бы оценить вероятности выравниваний.

IBM Model 1

EM-алгоритм

- нам нужно оценить параметры $t(f|e)$ на основе параллельного корпуса,
- ... но мы не знаем правильные выравнивания!

Проблема курицы и яйца:

- если бы мы знали правильные выравнивания,
  то мы могли бы оценить параметры $t(f|e)$.
- если бы мы знали параметры $t(f|e)$,
  то мы могли бы оценить вероятности выравниваний.

Решение - EM-алгоритм.

EM-алгоритм на примере

EM-алгоритм на примере

Параллельный корпус $\mathfrak{C}$:

- (*ZERO дом*) – (*a house*)
- (*ZERO книга*) – (*a book*)

EM-алгоритм на примере

Параллельный корпус $\mathfrak{C}$:

- $(ZERO\ дом) - (a\ house)$
- $(ZERO\ книга) - (a\ book)$

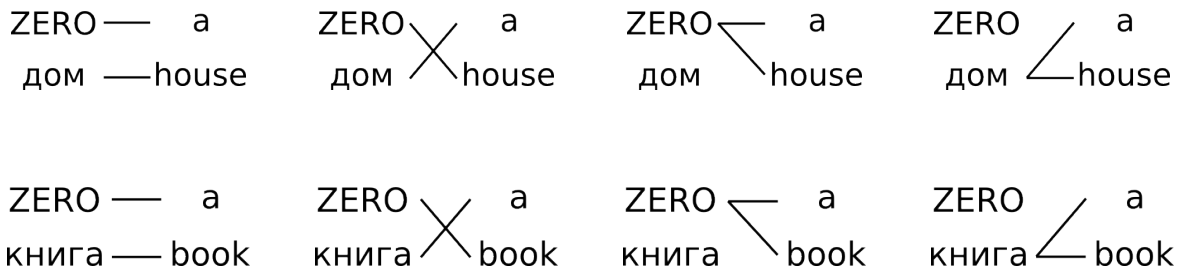
Начальная инициализация $t(f|e)$

$$
\begin{array}{cc}
t(a|ZERO) & 1/3 \\
t(house|ZERO) & 1/3 \\
t(book|ZERO) & 1/3
\end{array}
$$

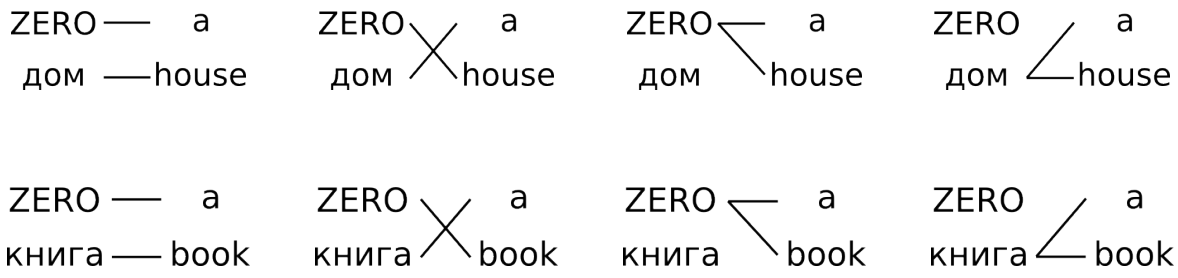
$$
\begin{array}{cc}
t(a|дом) & 1/3 \\
t(house|дом) & 1/3 \\
t(book|дом) & 1/3
\end{array}
$$

$$
\begin{array}{cc}
t(a|книга) & 1/3 \\
t(house|книга) & 1/3 \\
t(book|книга) & 1/3
\end{array}
$$

EM-алгоритм на примере

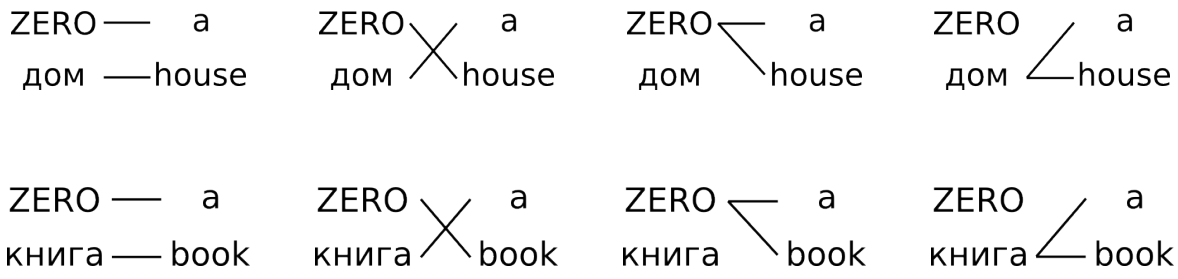


EM-алгоритм на примере



$P(F, \alpha|E)$ $\frac{1}{9}z$ $\frac{1}{9}z$ $\frac{1}{9}z$ $\frac{1}{9}z$

EM-алгоритм на примере



$P(F, \alpha \mid E)$	$\frac{1}{9}z$	$\frac{1}{9}z$	$\frac{1}{9}z$	$\frac{1}{9}z$
$P(\alpha \mid F, E)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

EM-алгоритм на примере

Пересчитываем параметры t:

$$
\begin{array}{cc}
c(a|ZERO) & 1/4 + 1/4 + 1/4 + 1/4 = 1 \\
c(house|ZERO) & 1/4 + 1/4 = 1/2 \\
c(book|ZERO) & 1/4 + 1/4 = 1/2
\end{array}
$$

$$
\begin{array}{cc}
c(a|дом) & 1/4 + 1/4 = 1/2 \\
c(house|дом) & 1/4 + 1/4 = 1/2 \\
c(book|дом) & 0
\end{array}
$$

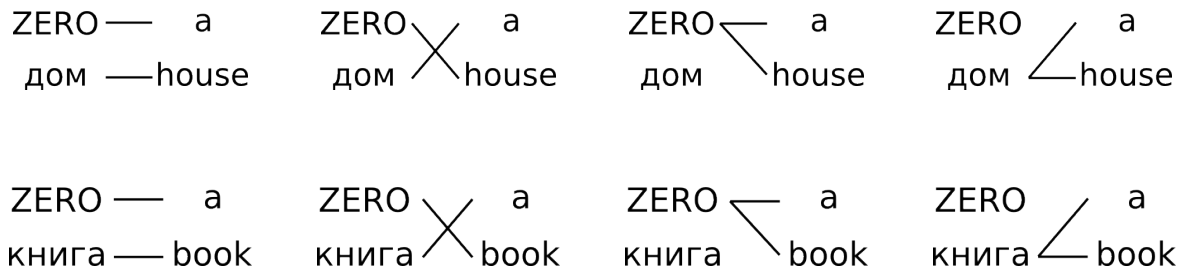
$$
\begin{array}{cc}
c(a|книга) & 1/4 + 1/4 = 1/2 \\
c(house|книга) & 0 \\
c(book|книга) & 1/4 + 1/4 = 1/2
\end{array}
$$

EM-алгоритм на примере

Пересчитываем параметры t:

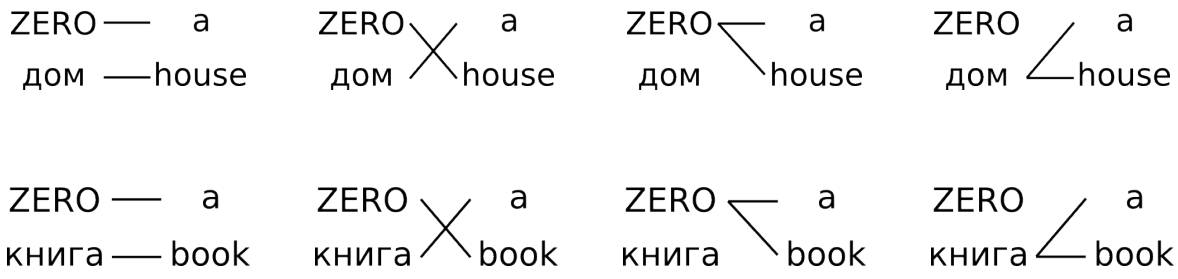
$t(a\|ZERO)$	$1/3$	$\mathbf{1/2}$
$t(house\|ZERO)$	$1/3$	$\mathbf{1/4}$
$t(book\|ZERO)$	$1/3$	$\mathbf{1/4}$
$t(a\|дом)$	$1/3$	$\mathbf{1/2}$
$t(house\|дом)$	$1/3$	$\mathbf{1/2}$
$t(book\|дом)$	$1/3$	$\mathbf{0}$
$t(a\|книга)$	$1/3$	$\mathbf{1/2}$
$t(house\|книга)$	$1/3$	$\mathbf{0}$
$t(book\|книга)$	$1/3$	$\mathbf{1/2}$

EM-алгоритм на примере



$P(F, \alpha|E)$ $\frac{1}{4}z$ $\frac{1}{8}z$ $\frac{1}{8}z$ $\frac{1}{4}z$

EM-алгоритм на примере



$P(F, \alpha \mid E)$	$\frac{1}{4}z$	$\frac{1}{8}z$	$\frac{1}{8}z$	$\frac{1}{4}z$
$P(\alpha \mid F, E)$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$

EM-алгоритм на примере

Пересчитываем параметры t:

$$
\begin{array}{cc}
c(a|ZERO) & 1/3 + 1/6 + 1/3 + 1/6 = 1 \\
c(house|ZERO) & 1/6 + 1/6 = 1/3 \\
c(book|ZERO) & 1/6 + 1/6 = 1/3
\end{array}
$$

$$
\begin{array}{cc}
c(a|дом) & 1/3 + 1/6 = 1/2 \\
c(house|дом) & 1/3 + 1/3 = 2/3 \\
c(book|дом) & 0
\end{array}
$$

$$
\begin{array}{cc}
c(a|книга) & 1/3 + 1/6 = 1/2 \\
c(house|книга) & 0 \\
c(book|книга) & 1/3 + 1/3 = 2/3
\end{array}
$$

EM-алгоритм на примере

Пересчитываем параметры t:

$t(a\|ZERO)$	1/3	1/2	**3/5**
$t(house\|ZERO)$	1/3	1/4	**1/5**
$t(book\|ZERO)$	1/3	1/4	**1/5**
$t(a\|дом)$	1/3	1/2	**3/7**
$t(house\|дом)$	1/3	1/2	**4/7**
$t(book\|дом)$	1/3	0	**0**
$t(a\|книга)$	1/3	1/2	**3/7**
$t(house\|книга)$	1/3	0	**0**
$t(book\|книга)$	1/3	1/2	**4/7**

EM-алгоритм

Перебор по всем выравниваниям α - неэффективно.
$P(F|E)$ можно вычислить по формуле:

$$P(F|E) = \sum_{\alpha} P(F, \alpha|E) = \frac{\epsilon}{(|E|+1)^{|F|}} \prod_{j=1}^{|F|} \sum_{i=0}^{|E|} t(f_j|e_i)$$

Тогда

$$P(\alpha|F, E) = \frac{\frac{\epsilon}{(|E|+1)^{|F|}} \prod_{j=1}^{|F|} t(f_j|e_{\alpha(j)})}{\frac{\epsilon}{(|E|+1)^{|F|}} \prod_{j=1}^{|F|} \sum_{i=0}^{|E|} t(f_j|e_i)} = \prod_{j=1}^{|F|} \frac{t(f_j|e_{\alpha(j)})}{\sum_{i=0}^{|E|} t(f_j|e_i)}$$

EM-алгоритм

Формулы пересчета параметров:

$$c(f|e;F,E) = \frac{t(f|e)}{\sum_{i=0}^{|E|} t(f|e_i)} \sum_{j=1}^{|F|} \delta(f,f_j) \sum_{i=0}^{|E|} \delta(e,e_i)$$

$$t(f|e;\mathfrak{C}) = t(f|e;\{(F,E)\}) = \frac{\sum_{(F,E)} c(f|e;F,E)}{\sum_{\tilde{f}} \sum_{(F,E)} c(\tilde{f}|e;F,E)}$$

δ - символ Кронекера

EM-алгоритм итерационный:

- до сходимости или
- некоторое фиксированное количество итераций

IBM Models, фразовый перевод

- Model 1: пословный перевод
- Model 2: + глобальное переупорядочивание слов
- Model 3: + вставка / удаление слов
- Model 4: + локальное переупорядочивание слов
- Model 5: + разрешение конфликта вставки различных слов на одно и то же место

Фразовый перевод:

- предложения можно делить на фразы, а не просто на токены
- фразы могут быть произвольными (n-граммы) - не обязательно "корректными с точки зрения лингвистики"

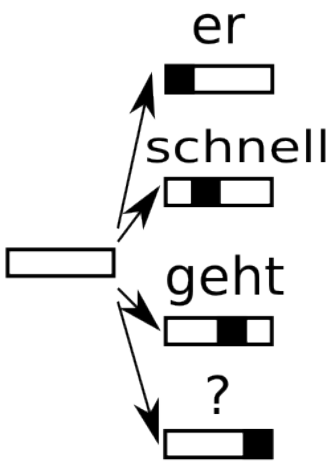
P. F. Brown, V. J. Della Pietra, S. A. Della Pietra, R. L. Mercer
The mathematics of statistical machine translation: parameter estimation

Декодирование

Декодирование

Полный перебор

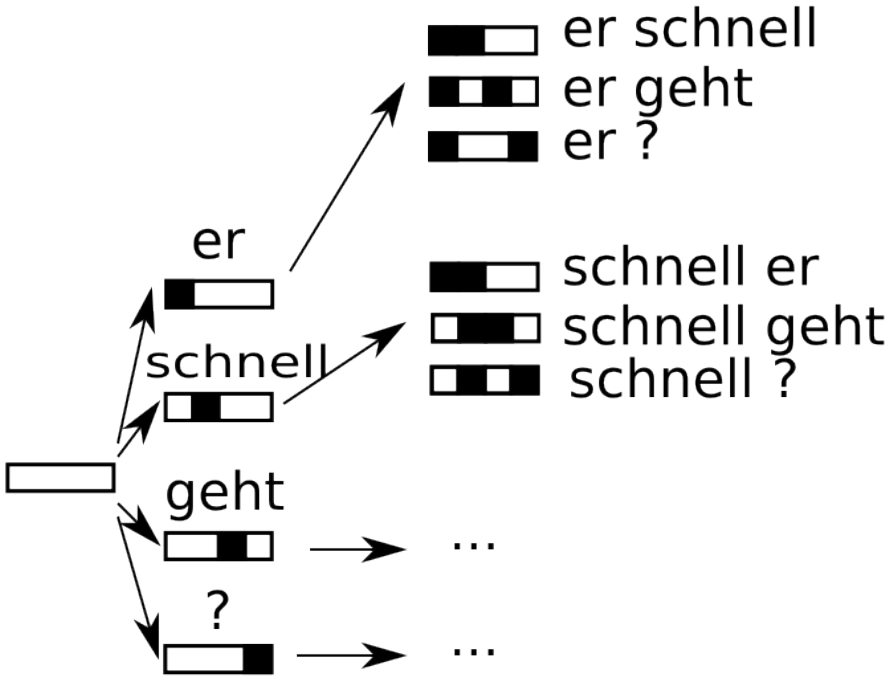
Он быстро идет?



Декодирование

Полный перебор

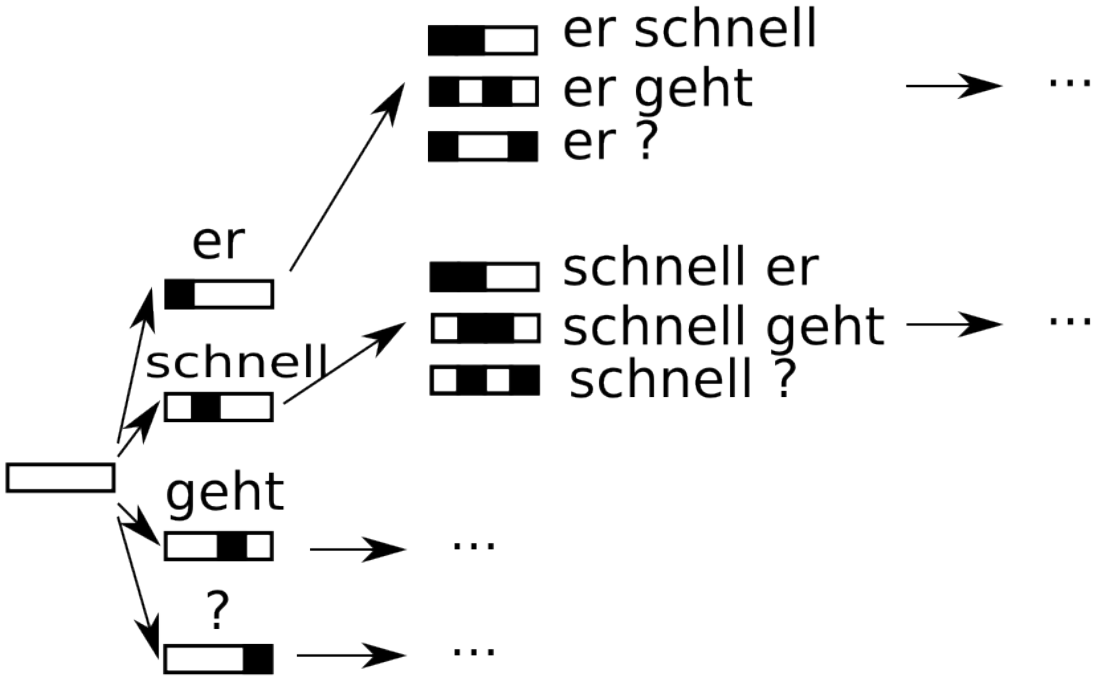
Он быстро идет?



Декодирование

Полный перебор

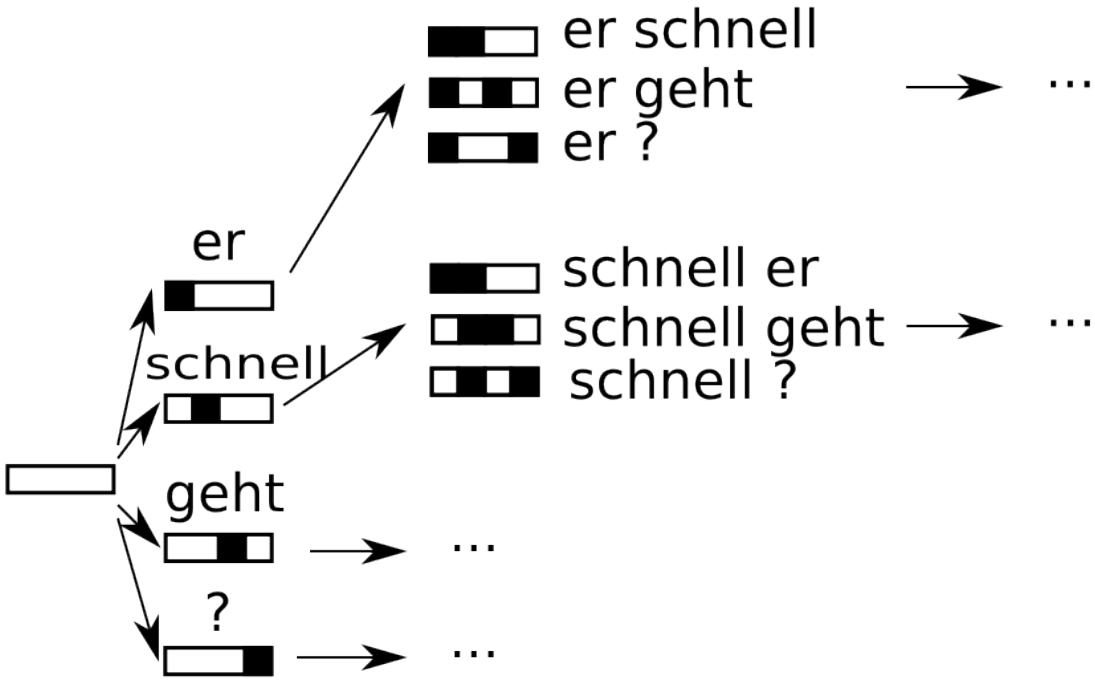
Он быстро идет?



Декодирование

Полный перебор

Он быстро идет?



Полный перебор неэффективен!

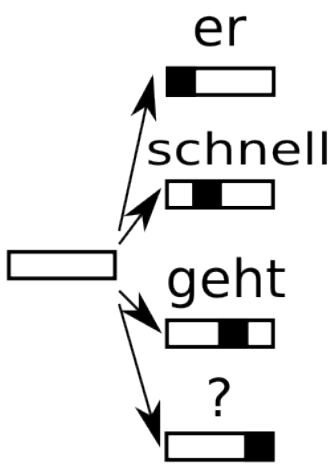
Декодирование

Перебор с отсечением

Декодирование

Перебор с отсечением

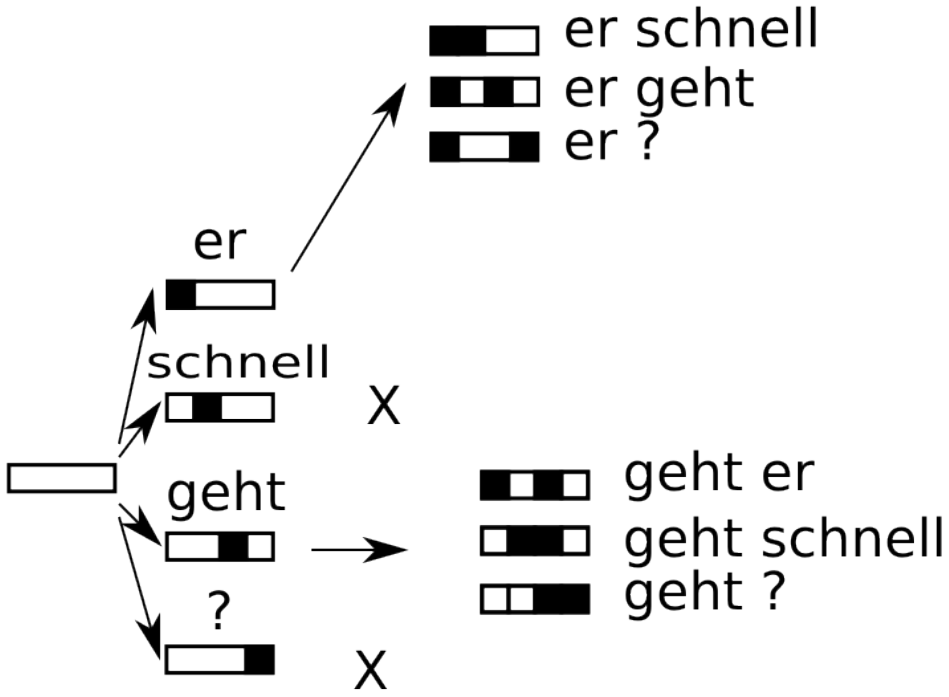
Он быстро идет?



Декодирование

Перебор с отсечением

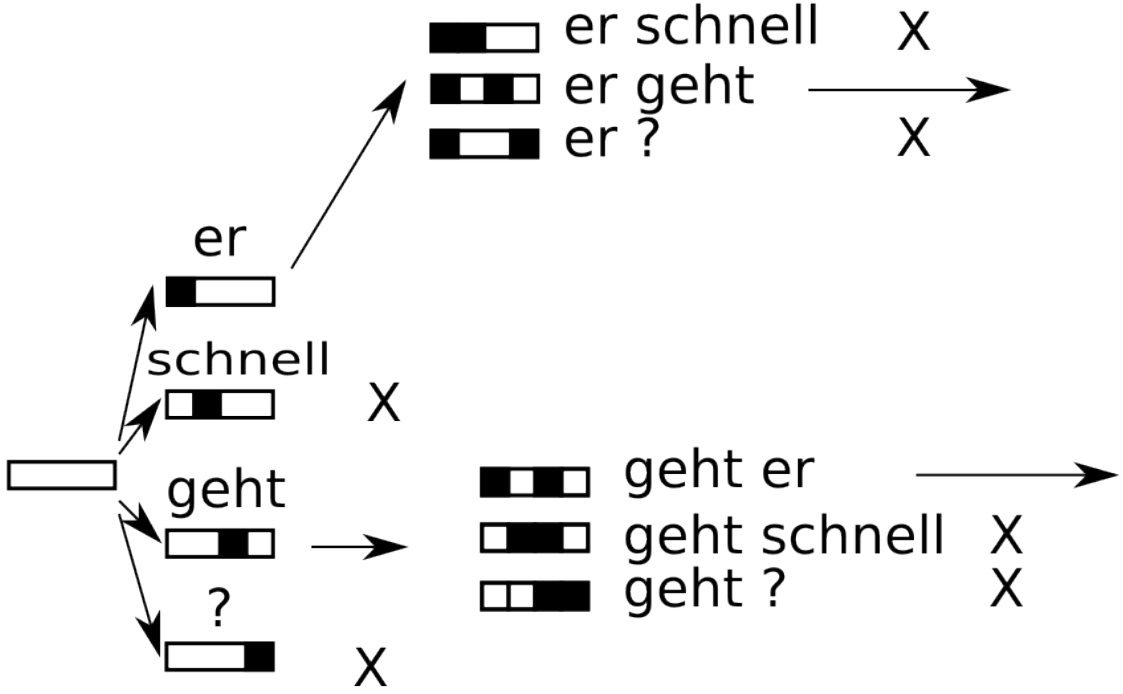
Он быстро идет?



Декодирование

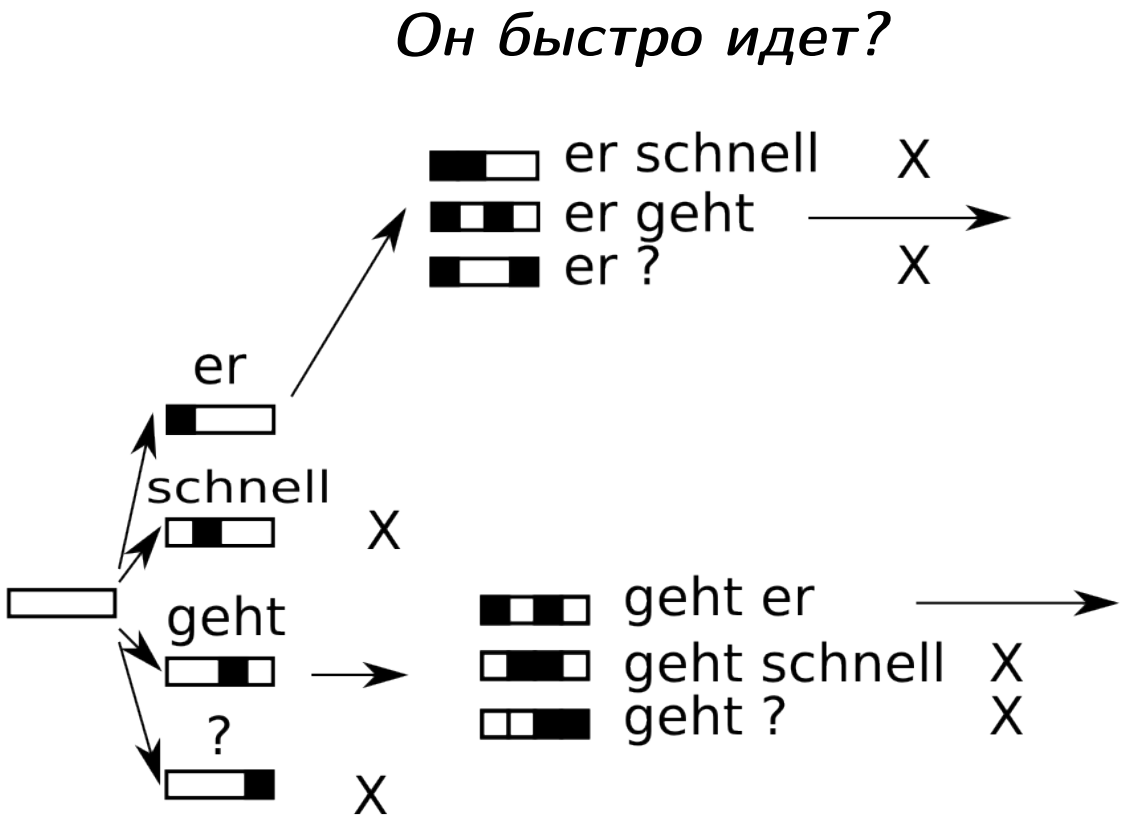
Перебор с отсечением

Он быстро идет?



Декодирование

Перебор с отсечением



Оценка частичных переводов:

$$P(F|E)P(E)$$

Декодирование

Перебор с отсечением

Оценка перевода:

$$score(F) = P(F|E)P(E)$$

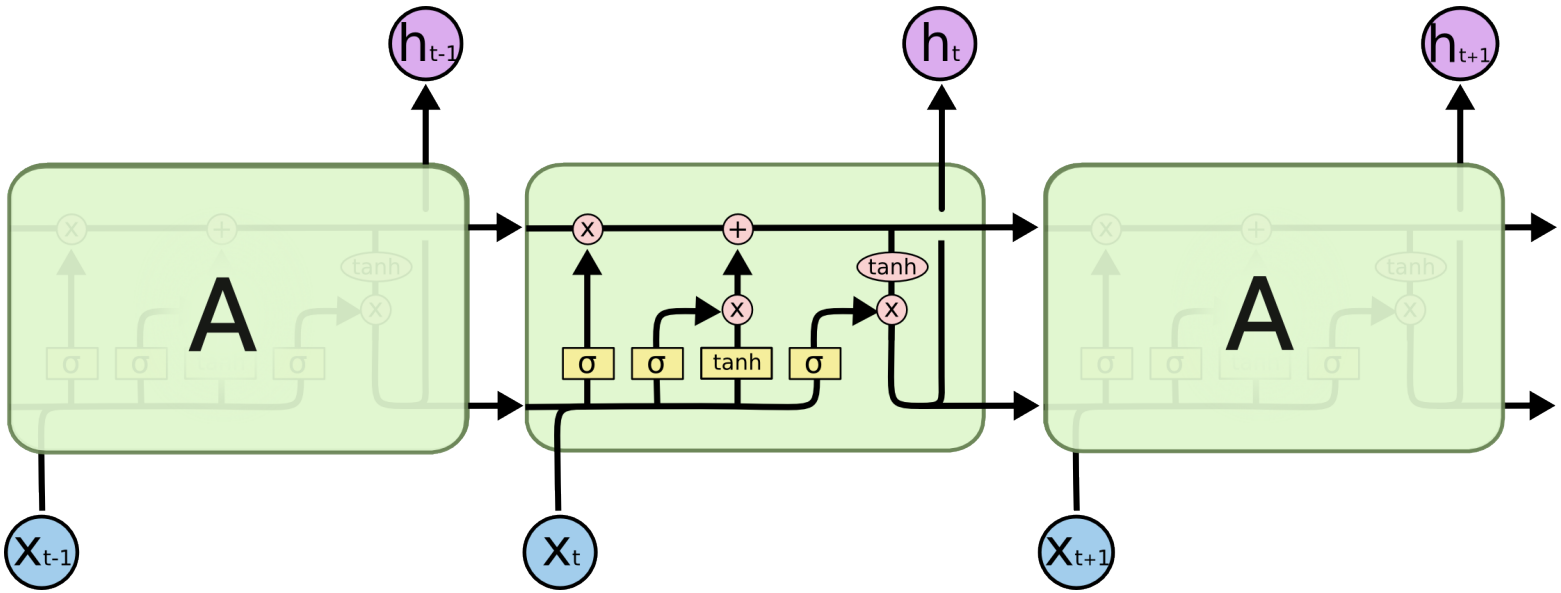
Он быстро идет?

er geht schnell?

geht er schnell?

Нейросетевой машинный перевод

LSTM



Understanding LSTM Networks

http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Сверточные нейронные сети - CNN

Фильтр

2	-1	-1
-1	0	0
-1	0	1

Результат

2	1	3	1	2
3	4	2	0	1
2	0	1	2	3
1	3	2	4	0

Сверточные нейронные сети - CNN

#1

Фильтр

2	-1	-1
-1	0	0
-1	0	1

Результат

-4		

2 2	1 -1	3 -1	1	2
3 -1	4 0	2 0	0	1
2 -1	0 0	1 1	2	3
1	3	2	4	0

Сверточные нейронные сети - CNN

#2

Фильтр

2	-1	-1
-1	0	0
-1	0	1

Результат

-4	-4	

2	1 2	3 -1	1 -1	2
3	4 -1	2 0	0 0	1
2	0 -1	1 0	2 1	3
1	3	2	4	0

Сверточные нейронные сети - CNN

#3

Фильтр

2	-1	-1
-1	0	0
-1	0	1

Результат

-4	-4	3

2	1	3 $_{2}$	1 $_{-1}$	2 $_{-1}$
3	4	2 $_{-1}$	0 $_{0}$	1 $_{0}$
2	0	1 $_{-1}$	2 $_{0}$	3 $_{1}$
1	3	2	4	0

Сверточные нейронные сети - CNN

#4

Фильтр

2	-1	-1
-1	0	0
-1	0	1

Результат

-4	-4	3
-1		

2	1	3	1	2
3 2	4 -1	2 -1	0	1
2 -1	0 0	1 0	2	3
1 -1	3 0	2 1	4	0

Сверточные нейронные сети - CNN

#5

Фильтр

2	-1	-1
-1	0	0
-1	0	1

Результат

-4	-4	3
-1	7	

2	1	3	1	2
3	4 2	2 -1	0 -1	1
2	0 -1	1 0	2 0	3
1	3 -1	2 0	4 1	0

Сверточные нейронные сети - CNN

#6

Фильтр

2	-1	-1
-1	0	0
-1	0	1

Результат

-4	-4	3
-1	7	0

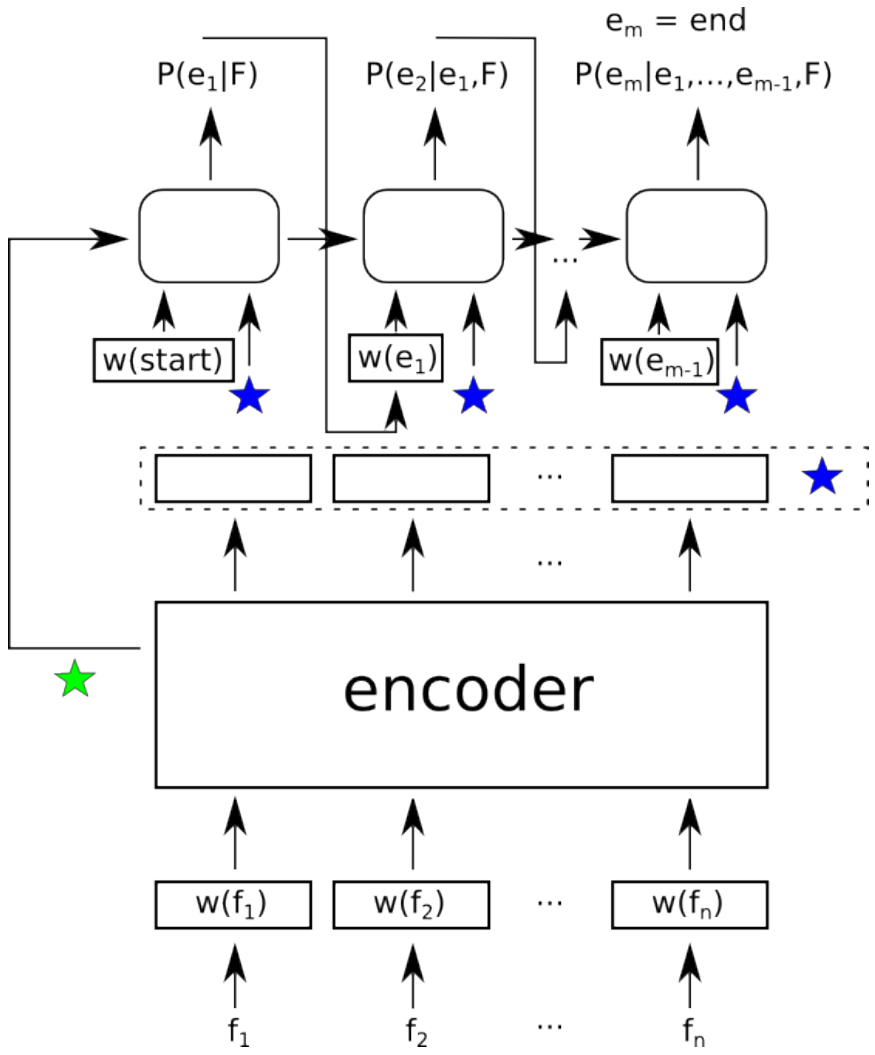
2	1	3	1	2
3	4	2_{2}	0_{-1}	1_{-1}
2	0	1_{-1}	2_{0}	3_{0}
1	3	2_{-1}	4_{0}	0_{1}

Нейросетевой машинный перевод

- обучение по корпусу параллельных текстов:
  $\mathfrak{C} = \{(F_1, E_1), (F_2, E_2), ..., (F_m, E_m)\}$
- ограничение размера словаря ($< 40k$), все остальные слова заменяются на специальный токен *UNK*

Нейросетевой машинный перевод

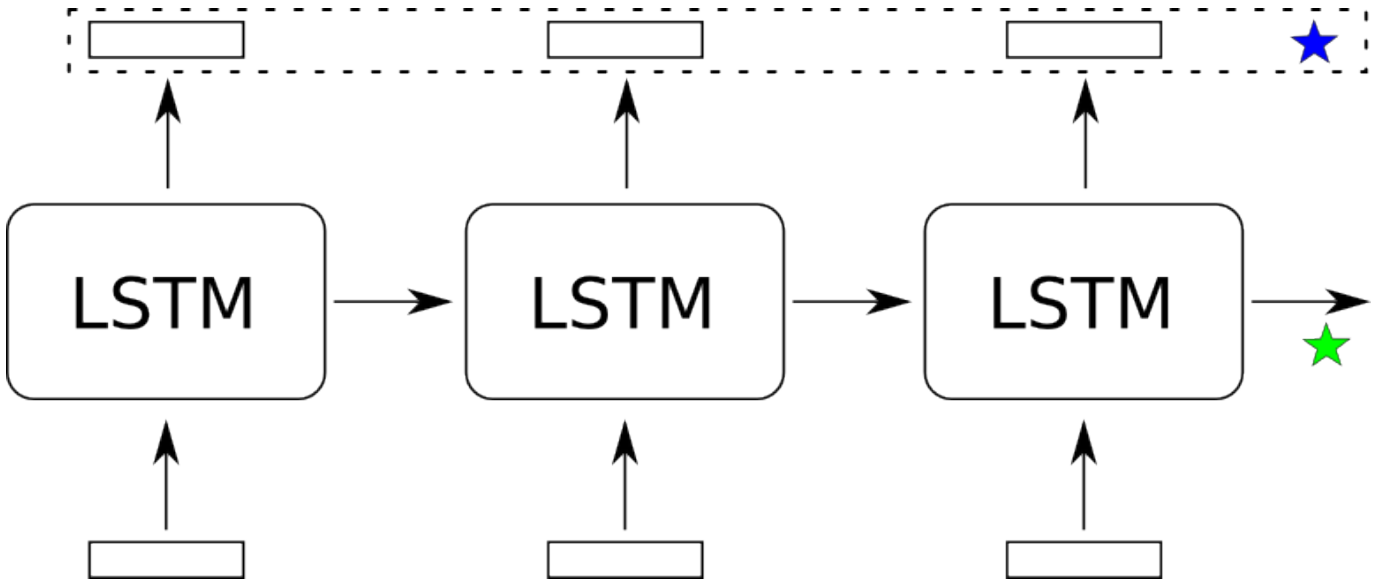
Архитектура кодировщик-декодировщик



I. Sutskever, O. Vinyals, Q. V.Le
Sequence to sequence learning with neural networks

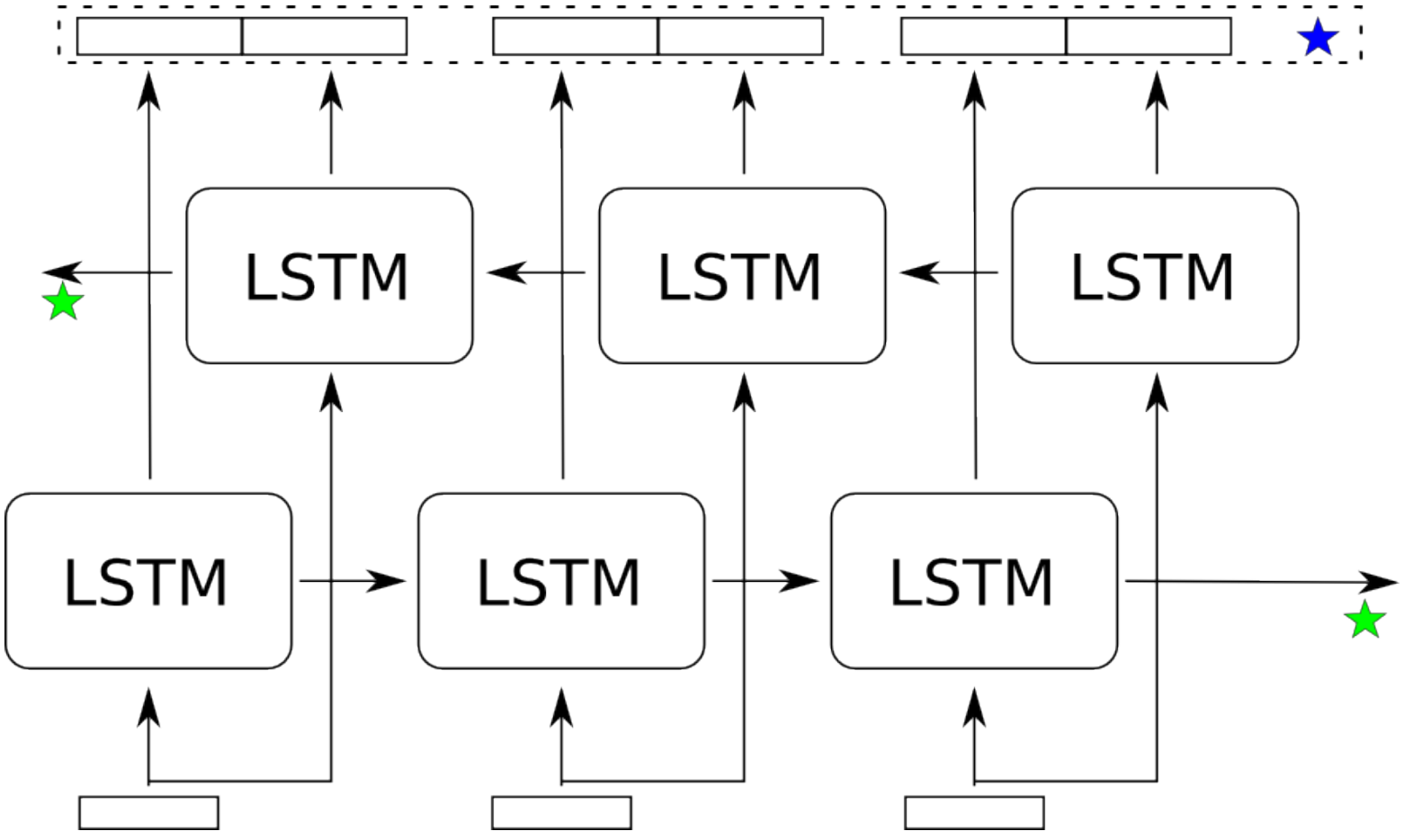
Нейросетевой машинный перевод

LSTM-кодировщик



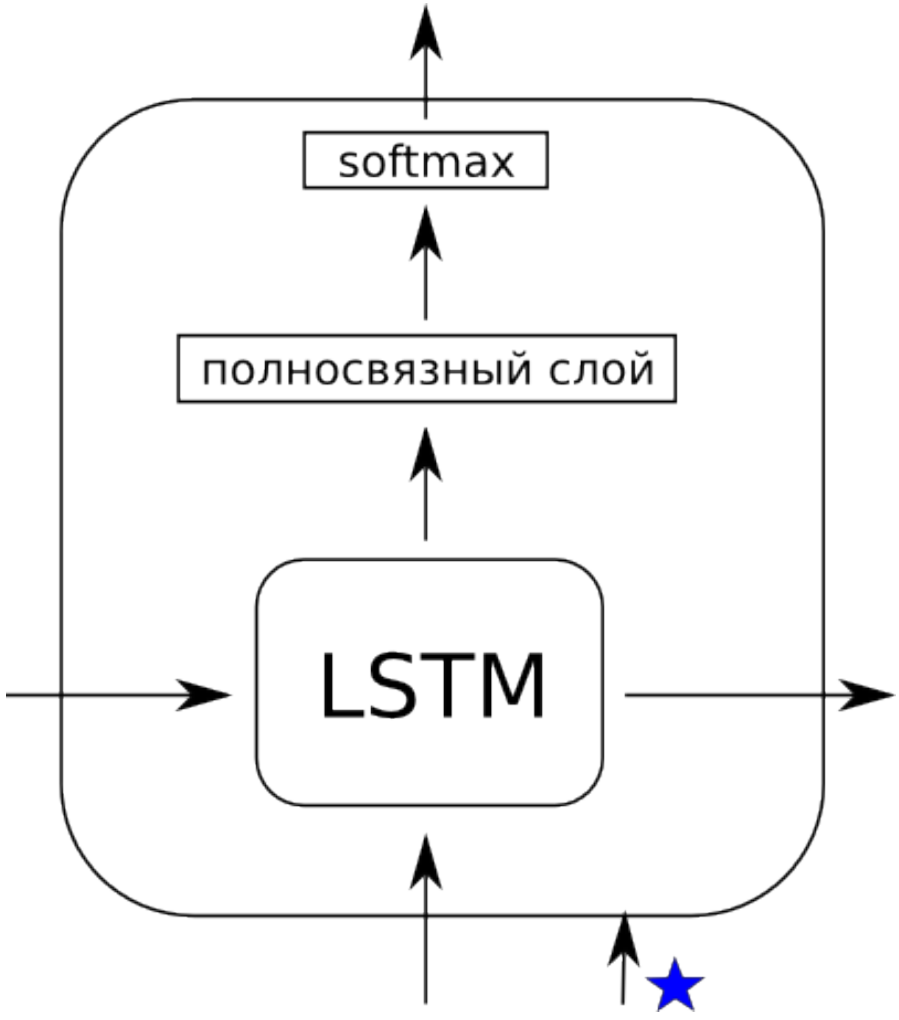
Нейросетевой машинный перевод

BiLSTM-кодировщик



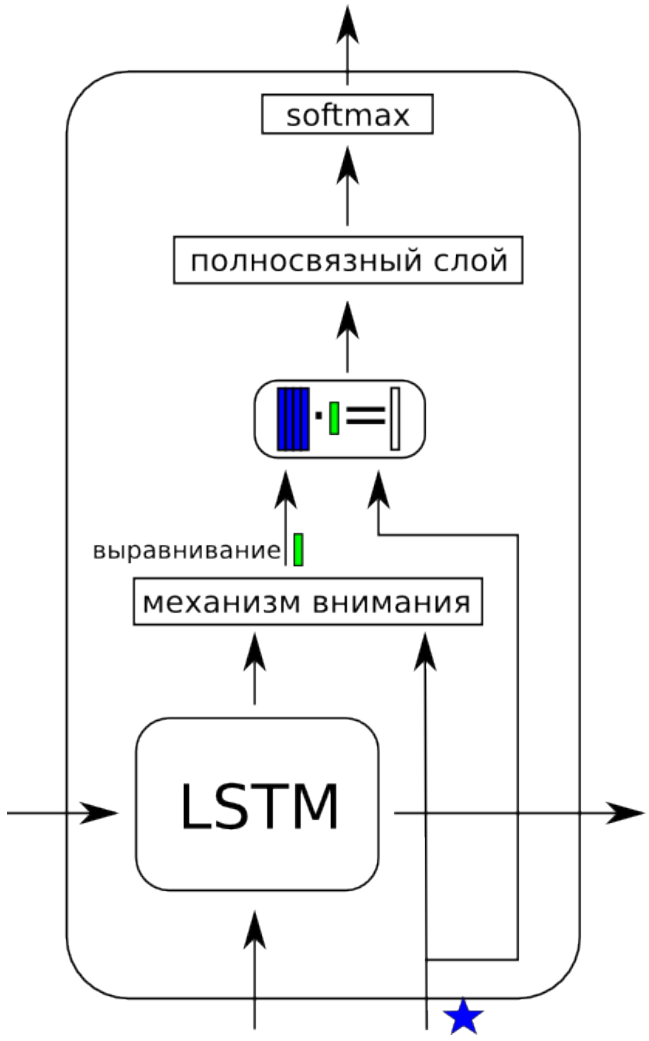
Нейросетевой машинный перевод

Простой декодировщик



Нейросетевой машинный перевод

Декодировщик с вниманием



D. Bahdanau, K. Cho, Y. Bengio
Neural machine translation by jointly learning to align and translate

Нейросетевой машинный перевод

Механизм внимания

$h_1, h_2, ..., h_n$ - вектора для токенов, полученные из кодировщика
s_j - выходной вектор LSTM-ячейки

$$\alpha(h_1, h_2, ..., h_n; s_j) = softmax(score(h_1, s_j), score(h_2, s_j), ..., score(h_n, s_j))$$

Механизмы внимания:

M. T. Luong, H. Pham, C. D. Manning
Effective approaches to attention-based neural machine translation

Нейросетевой машинный перевод

Механизм внимания

$h_1, h_2, ..., h_n$ - вектора для токенов, полученные из кодировщика
s_j - выходной вектор LSTM-ячейки

$$\alpha(h_1, h_2, ..., h_n; s_j) = softmax(score(h_1, s_j), score(h_2, s_j), ..., score(h_n, s_j))$$

Механизмы внимания:

- Bahdanau:

$$score(h_i, s_j) = v_\alpha^T tanh(W_\alpha^B s_j + U_\alpha^B h_i)$$

M. T. Luong, H. Pham, C. D. Manning
Effective approaches to attention-based neural machine translation

Нейросетевой машинный перевод

Механизм внимания

$h_1, h_2, ..., h_n$ - вектора для токенов, полученные из кодировщика
s_j - выходной вектор LSTM-ячейки

$$\alpha(h_1, h_2, ..., h_n; s_j) = softmax(score(h_1, s_j), score(h_2, s_j), ..., score(h_n, s_j))$$

Механизмы внимания:

- Bahdanau:

$$score(h_i, s_j) = v_\alpha^T tanh(W_\alpha^B s_j + U_\alpha^B h_i)$$

- Luong:

$$score(h_i, s_j) = h_i^T W_\alpha^L s_j$$

M. T. Luong, H. Pham, C. D. Manning
Effective approaches to attention-based neural machine translation

Нейросетевой машинный перевод

Механизм внимания

$h_1, h_2, ..., h_n$ - вектора для токенов, полученные из кодировщика
s_j - выходной вектор LSTM-ячейки

$$\alpha(h_1, h_2, ..., h_n; s_j) = softmax(score(h_1, s_j), score(h_2, s_j), ..., score(h_n, s_j))$$

Механизмы внимания:

- Bahdanau:

$$score(h_i, s_j) = v_\alpha^T tanh(W_\alpha^B s_j + U_\alpha^B h_i)$$

- Luong:

$$score(h_i, s_j) = h_i^T W_\alpha^L s_j$$

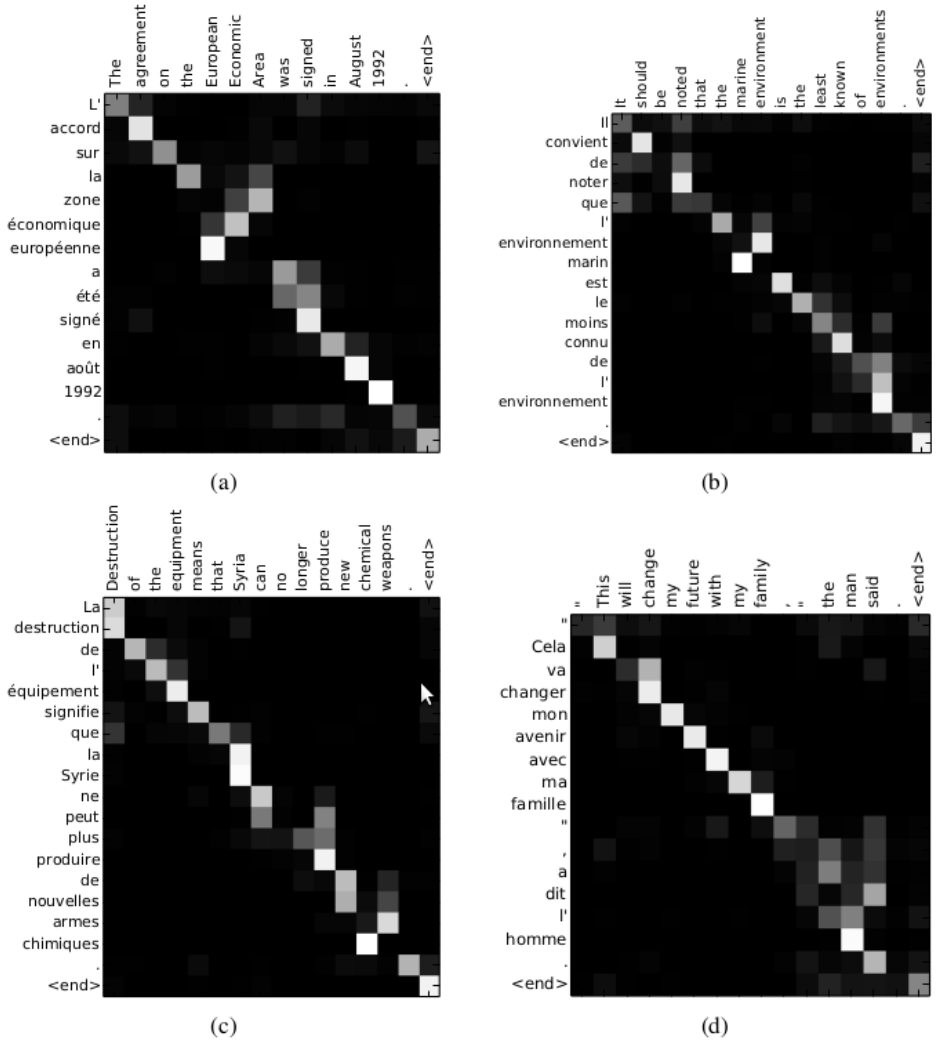
- concat:

$$score(h_i, s_j) = W_\alpha^c [h_i; s_j]$$

M. T. Luong, H. Pham, C. D. Manning
Effective approaches to attention-based neural machine translation

Нейросетевой машинный перевод

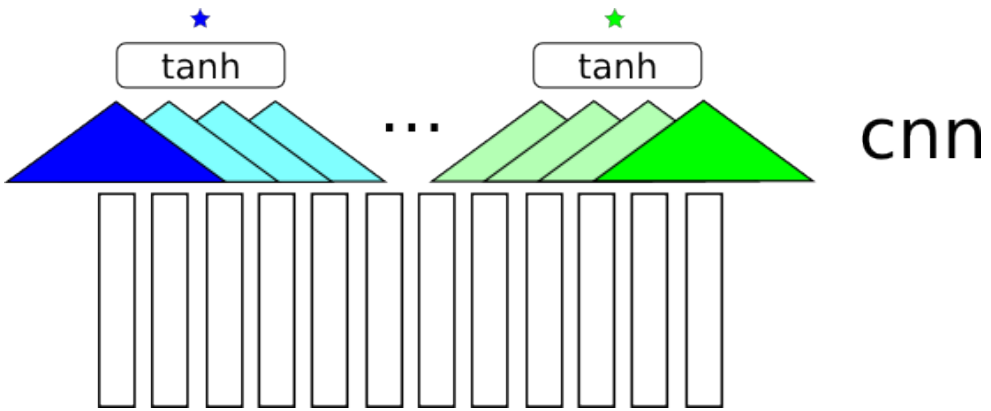
Выравнивания

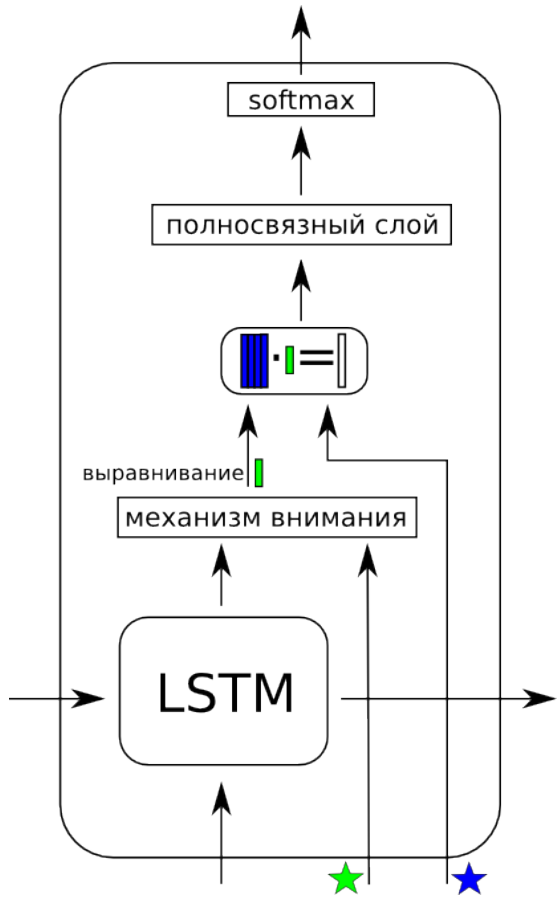


D. Bahdanau, K. Cho, Y. Bengio
Neural machine translation by jointly learning to align and translate

Нейросетевой машинный перевод

CNN-кодировщик





J. Gehring, M. Auli, D. Grangier, Y. N. Dauphin
A Convolutional Encoder Model for Neural Machine Translation

Гибридный машинный перевод

Ранжирование предложений, полученных системой статистического машинного перевода:

- модель языка - на основе реккурентной нейронной сети (вместо n-грамм)
- модель перевода - на основе архитектуры кодировщик-декодировщик

Оценка качества

- BLEU
- WER

Оценка качества

BLEU

$$BLEU = BP \times \sqrt[N]{\prod_{i=1}^{N} P'_i}$$

- обычно $N = 4$
- P'_n - модифицированная точность (в числителе вклад каждой n-граммы не может превышать ее максимальной кратности в правильных предложениях).
- BP - штраф за краткость:

$$BP = \begin{cases} 1 & c > r \\ e^{1-r/c} & c \leq r \end{cases}$$

c - длина предложения системы

r - длина экспертного предложения с наибольшим количеством совпадающих n-грамм

Оценка качества - BLEU

Пример

Система:
the the quick fox jumps after the the fox

Эксперты:
the brown fox jumps
the quick fox jumps over the dog

Оценка качества - BLEU

Пример

Система:
the the quick fox jumps after the the fox

Эксперты:
the brown fox jumps
the quick fox jumps over the dog

$$P'_1 = \frac{the + the + quick + fox + jumps}{9} = \frac{5}{9}$$

Оценка качества - BLEU

Пример

Система:
the the quick fox jumps after the the fox

Эксперты:
the brown fox jumps
the quick fox jumps over the dog

$$P'_1 = \frac{the + the + quick + fox + jumps}{9} = \frac{5}{9}$$

$$P'_2 = \frac{3}{8} \qquad P'_3 = \frac{2}{7} \qquad P'_4 = \frac{1}{6}$$

Оценка качества - BLEU

Пример

Система:
the the quick fox jumps after the the fox

Эксперты:
the brown fox jumps
the quick fox jumps over the dog

$$P'_1 = \frac{the + the + quick + fox + jumps}{9} = \frac{5}{9}$$

$$P'_2 = \frac{3}{8} \qquad P'_3 = \frac{2}{7} \qquad P'_4 = \frac{1}{6}$$

$$BLEU = 1 \times \sqrt[4]{\frac{5}{9} \times \frac{3}{8} \times \frac{2}{7} \times \frac{1}{6}} = 0.315598454$$

Оценка качества

WER

WER (word error rate) - нормализованное количество операций редактирования, необходимых для преобразования выданного системой предложения в правильное (составленное экспертом).

$$WER = \frac{S + D + I}{N}$$

где

- S - количество замен
- D - количество удалений
- I - количество вставок
- N - количество токенов в правильном предложении

Ошибки машинного перевода

Ошибки машинного перевода

Google Переводчик

- It's raining cats and dogs.

- He was wearing his heart on his sleeve after the meeting with his boss.

Ошибки машинного перевода

Google Переводчик

- It's raining cats and dogs.
- Льет как из ведра.

- He was wearing his heart on his sleeve after the meeting with his boss.
- После встречи со своим боссом он был одет в свое сердце на рукаве.

Ошибки машинного перевода

Google Переводчик

- It's raining cats and dogs.
- Льет как из ведра.

- He was wearing his heart on his sleeve after the meeting with his boss.
- После встречи со своим боссом он был одет в свое сердце на рукаве.
- Он не мог скрыть своих чувств после встречи с начальником.

Ошибки машинного перевода

Google Переводчик

- There's a book on a table. Take it.

Ошибки машинного перевода

Google Переводчик

- There's a book on a table. Take it.
- На столе есть книга.

Ошибки машинного перевода

Google Переводчик

- There's a book on a table. Take it.
- На столе есть книга. Возьми это.

Ошибки машинного перевода

Google Переводчик

- There's a book on a table. Take it.
- На столе есть книга. Возьми это.

- На столе лежит книга. Возьми её.
- There's a book on the table. Take her.

Следующая лекция

Разрешение кореферентности

лектор: Сысоев Андрей Анатольевич