

Unsupervised Machine Translation

Носков Федор
Дискуссионный клуб NLP

14 сентября 2020 г.

Общий алгоритм

Algorithm 1 Unsupervised MT

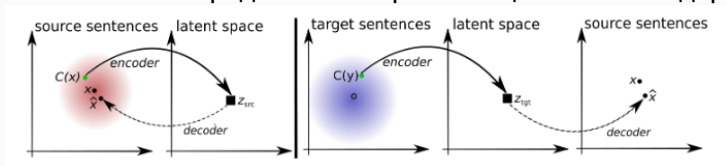
Ensure: P_s, P_t — language models, $P_{s \rightarrow t}^{(0)}, P_{t \rightarrow s}^{(0)}$ — initial translation models

- 1: **for** $k = 1$ to N **do**
- 2: **Back-translation:** Generate source and target sentences using the current models, $P_{t \rightarrow s}^{(k-1)}$ and $P_{s \rightarrow t}^{(k-1)}$, factoring in language models, P_s and P_t
- 3: Train new translation models $P_{t \rightarrow s}^{(k)}$ and $P_{s \rightarrow t}^{(k)}$ using the generated sentences
- 4: **end for**

Details

Language models

- $L^{LM} = \mathbb{E}_{x \sim \mathcal{S}}[-\log P_{s \rightarrow s}(x|C(x))] + \mathbb{E}_{y \sim \mathcal{T}}[-\log P_{t \rightarrow t}(y|C(y))] \rightarrow \min$
- $C(x)$ — нечто вроде аугментации, некоторые слова с некоторой вероятностью пропускаются, с некоторой переставляются, но на не очень большое расстояние
- Фактически перед нами исправляющий автоэнкодер:



Details

Translation model

- $u^*(y) = \operatorname{argmax} P_{t \rightarrow s}(u|y)$
- $v^*(x) = \operatorname{argmax} P_{s \rightarrow t}(v|x)$
- Loss

$$L^{back} = \mathbb{E}_{y \sim \mathcal{T}}[-\log P_{s \rightarrow t}(y|u^*(y))] + \mathbb{E}_{x \sim \mathcal{S}}[-\log P_{t \rightarrow s}(x|v^*(x))]$$

Details

Архитектура

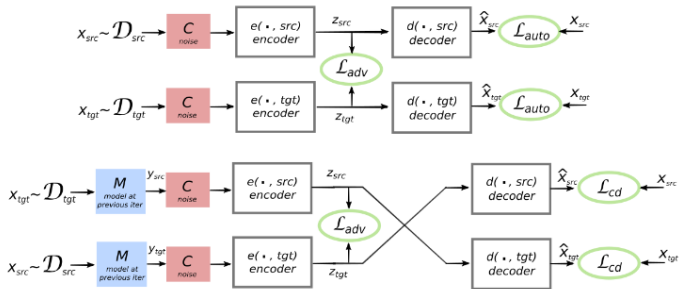


Рис. 1: Схема работы сети (на лосс смотреть не надо)

Details

Общий принцип работы

- Два языка — target и source
- Энкодер — переводит предложение в вектор некоего (скрытого) пространства
- Декодер — переводит вектор из скрытого пространства в предложения. И энкодеру, и декодеру перед скормливанием предложения следует сообщить, на каком языке они работают

Details

Общий принцип работы

- Модель $P_{s \rightarrow s}$ составляется из $enc(\cdot, src)$ и декодера $dec(\cdot, src)$ (т.е. энкодеру и декодеру сообщается, что они работают с языком source), ее обучают исправлять предложения, испорченные с помощью функции C (см. предыдущие слайды)
- Модель $P_{s \rightarrow t}$ составляется из $enc(\cdot, src)$ и $dec(\cdot, trgt)$, и она учится переводить
- Модели $P_{t \rightarrow t}$ и модели $P_{t \rightarrow s}$ работают аналогично
- Обучение всех моделей происходит одновременно путем минимизации ошибки $L^{LM} + L^{back}$ (см. предыдущие слайды)

Performance

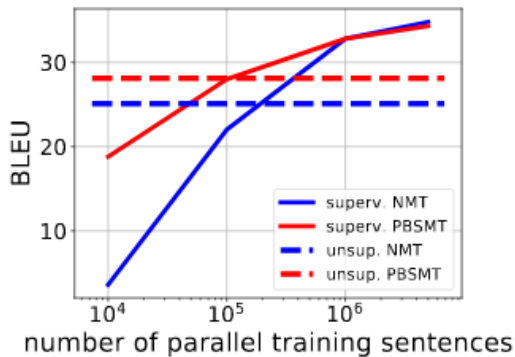


Рис. 2: Comparison between unsupervised and supervised approaches