Word Translation Without Parallel Data

Кнышов Александр, группа 151.

Мотивация

- Для машинного перевода нужно большое количество параллельных данных

- Это довольно дорого, для некоторых пар языков таких данных очень мало

- Хочется научиться переводить любую языковую пару достаточно хорошо и не очень затратно

Что было раньше(Procrustes problem)

- Пусть у нас есть параллельный словарь из n = 5000 пар слов (x, y) размерности d.

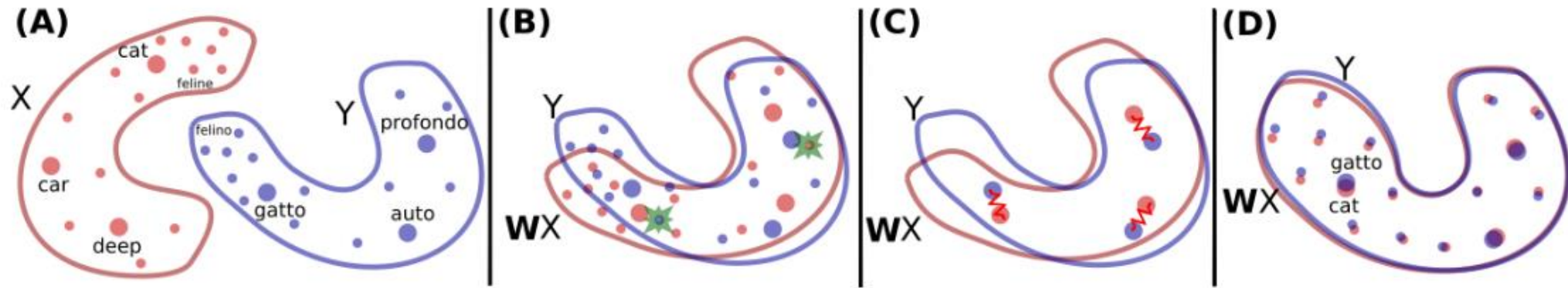$$W^\star = \operatorname*{argmin}_{W \in M_d(\mathbb{R})} \|WX - Y\|_{\mathrm{F}}$$

- Добавим ортогональность матрицы W.

$$W^\star = \operatorname*{argmin}_{W \in M_d(\mathbb{R})} \|WX - Y\|_{\mathrm{F}} \quad \text{s.t } WW^T = I$$
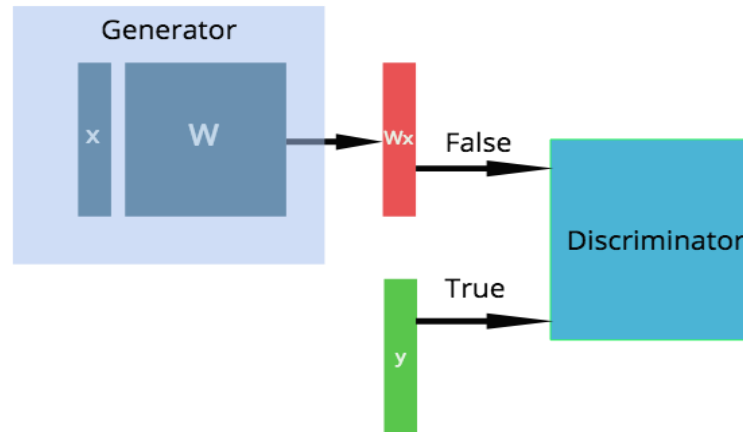
- Решением будет являться

$$W^\star = \operatorname*{argmin}_{W \in O_d(\mathbb{R})} \|WX - Y\|_{\mathrm{F}} = UV^T, \text{with } U\Sigma V^T = \mathrm{SVD}(YX^T).$$

Модель

1. Adversarial learning
2. Refinement
3. Translation

Domain-adversarial setting

Дискриминатор

$$\mathcal{L}_D(\theta_D|W) = -\frac{1}{n}\sum_{i=1}^{n} \log P_{\theta_D}\left(\text{source}=1\big|Wx_i\right) - \frac{1}{m}\sum_{i=1}^{m} \log P_{\theta_D}\left(\text{source}=0\big|y_i\right).$$

Генератор

$$\mathcal{L}_W(W|\theta_D) = -\frac{1}{n}\sum_{i=1}^{n} \log P_{\theta_D}\left(\text{source}=0\big|Wx_i\right) - \frac{1}{m}\sum_{i=1}^{m} \log P_{\theta_D}\left(\text{source}=1\big|y_i\right).$$

Ортогональность

$$W \leftarrow (1+\beta)W - \beta(WW^T)W$$

Refinement

- По полученной W хотим построить высококачественный словарь
- Выбираем слова, в которых уверена модель
- Решаем Procrustes problem

$$W^\star = \operatorname*{argmin}_{W \in O_d(\mathbb{R})} \|WX - Y\|_F = UV^T, \text{with } U\Sigma V^T = \text{SVD}(YX^T).$$

Cross-domain similarity local scaling (CSLS)

- K-NN несимметрична
- Hubness problem

$\mathcal{N}_T(x)$ - K ближайших соседей для x из target языка T

$$r_{\mathrm{T}}(W x_s) = \frac{1}{K} \sum_{y_t \in \mathcal{N}_{\mathrm{T}}(W x_s)} \cos(W x_s, y_t),$$

$$\mathrm{CSLS}(W x_s, y_t) = 2 \cos(W x_s, y_t) - r_{\mathrm{T}}(W x_s) - r_{\mathrm{S}}(y_t).$$

Выбор лучшей модели

- Возьмем 10k самых популярных слов из source
- Найдем для них лучшие переводы с точки зрения CSLS
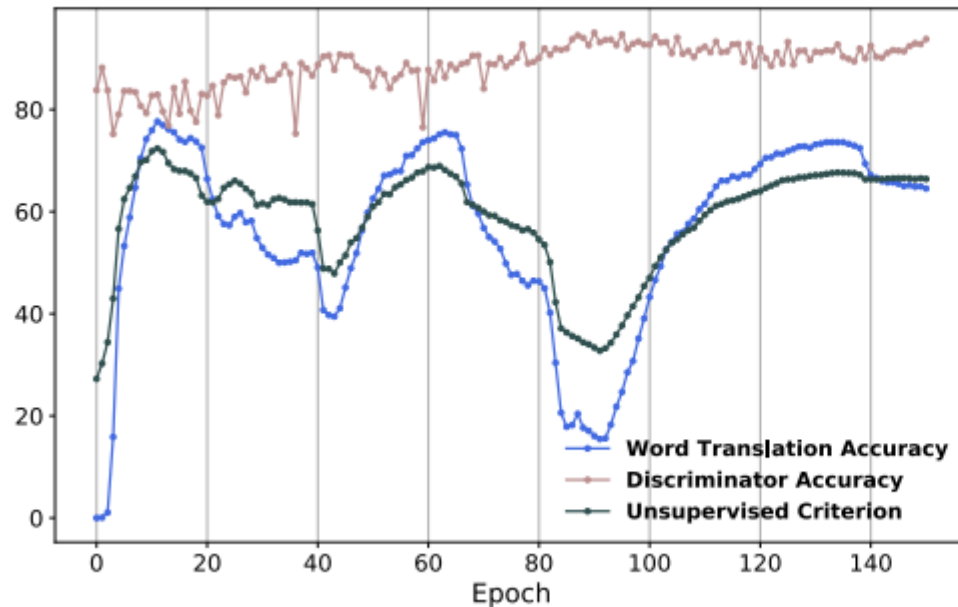- Посчитаем среднюю косинусную меру всех пар слов

Figure 2: Unsupervised model selection. Correlation between our unsupervised validation criterion (black line) and actual word translation accuracy (blue line). In this particular experiment, the selected model is at epoch 10. Observe how our criterion is well correlated with translation accuracy.

Результаты

	en-es	es-en	en-fr	fr-en	en-de	de-en	en-ru	ru-en	en-zh	zh-en	en-eo	eo-en
Methods with cross-lingual supervision and fastText embeddings												
Procrustes - NN	77.4	77.3	74.9	76.1	68.4	67.7	47.0	58.2	40.6	30.2	22.1	20.4
Procrustes - ISF	81.1	82.6	81.1	81.3	71.1	71.5	49.5	63.8	35.7	**37.5**	29.0	27.9
Procrustes - CSLS	81.4	82.9	81.1	**82.4**	73.5	**72.4**	**51.7**	**63.7**	**42.7**	36.7	**29.3**	25.3
Methods without cross-lingual supervision and fastText embeddings												
Adv - NN	69.8	71.3	70.4	61.9	63.1	59.6	29.1	41.5	18.5	22.3	13.5	12.1
Adv - CSLS	75.7	79.7	77.8	71.2	70.1	66.4	37.2	48.1	23.4	28.3	18.6	16.6
Adv - Refine - NN	79.1	78.1	78.1	78.2	71.3	69.6	37.3	54.3	30.9	21.9	20.7	20.6
Adv - Refine - CSLS	**81.7**	**83.3**	**82.3**	82.1	**74.0**	72.2	44.0	59.1	32.5	31.4	28.2	**25.6**

Table 1: Word translation retrieval P@1 for our released vocabularies in various language pairs. We consider 1,500 source test queries, and 200k target words for each language pair. We use fastText embeddings trained on Wikipedia. NN: nearest neighbors. ISF: inverted softmax. ('en' is English, 'fr' is French, 'de' is German, 'ru' is Russian, 'zh' is classical Chinese and 'eo' is Esperanto)

Результаты

	English to Italian			Italian to English		
	P@1	P@5	P@10	P@1	P@5	P@10
Methods with cross-lingual supervision (WaCky)						
Mikolov et al. (2013b) [†]	33.8	48.3	53.9	24.9	41.0	47.4
Dinu et al. (2015)[†]	38.5	56.4	63.9	24.6	45.4	54.1
CCA[†]	36.1	52.7	58.1	31.0	49.9	57.0
Artetxe et al. (2017)	39.7	54.7	60.5	33.8	52.4	59.1
Smith et al. (2017)[†]	43.1	60.7	66.4	38.0	58.5	63.6
Procrustes - CSLS	44.9	61.8	66.6	38.5	57.2	63.0
Methods without cross-lingual supervision (WaCky)						
Adv - Refine - CSLS	45.1	60.7	65.1	38.3	57.8	62.8
Methods with cross-lingual supervision (Wiki)						
Procrustes - CSLS	63.7	78.6	81.1	56.3	76.2	80.6
Methods without cross-lingual supervision (Wiki)						
Adv - Refine - CSLS	**66.2**	**80.4**	**83.4**	**58.7**	**76.5**	**80.9**

Table 2: English-Italian word translation average precisions (@1, @5, @10) from 1.5k source word queries using 200k target words. Results marked with the symbol [†] are from Smith et al. (2017). Wiki means the embeddings were trained on Wikipedia using fastText. Note that the method used by Artetxe et al. (2017) does not use the same supervision as other supervised methods, as they only use numbers in their initial parallel dictionary.

Результаты

	English to Italian			Italian to English		
	P@1	P@5	P@10	P@1	P@5	P@10
Methods with cross-lingual supervision						
Mikolov et al. (2013b) [†]	10.5	18.7	22.8	12.0	22.1	26.7
Dinu et al. (2015) [†]	45.3	72.4	80.7	48.9	71.3	78.3
Smith et al. (2017) [†]	54.6	72.7	78.2	42.9	62.2	69.2
Procrustes - NN	42.6	54.7	59.0	53.5	65.5	69.5
Procrustes - CSLS	**66.1**	77.1	80.7	**69.5**	**79.6**	**83.5**
Methods without cross-lingual supervision						
Adv - CSLS	42.5	57.6	63.6	47.0	62.1	67.8
Adv - Refine - CSLS	65.9	**79.7**	**83.1**	69.0	**79.7**	83.1

Table 3: **English-Italian sentence translation retrieval**. We report the average P@k from 2,000 source queries using 200,000 target sentences. We use the same embeddings as in Smith et al. (2017). Their results are marked with the symbol [†].

Результаты

	en-eo	eo-en
Dictionary - NN	6.1	11.9
Dictionary - CSLS	11.1	14.3

Table 5: BLEU score on English-Esperanto.
Although being a naive approach, word-by-word translation is enough to get a rough idea of the input sentence. The quality of the generated dictionary has a significant impact on the BLEU score.

Source	mi kelkfoje parolas kun mia najbaro tra la barilo .
Hypothesis	sorry sometimes speaks with my neighbor across the barrier .
Reference	i sometimes talk to my neighbor across the fence .
Source	la viro malanta ili ludas la pianon .
Hypothesis	the man behind they plays the piano .
Reference	the man behind them is playing the piano .
Source	bonvole protektu min kontra tiuj malbonaj viroj .
Hypothesis	gratefully protects hi against those worst men .
Reference	please defend me from such bad men .

Список литературы

- https://arxiv.org/pdf/1710.04087.pdf
- https://mi-sigma.com/2018/06/19/conneau-et-al-2018-word-translation-without-parallel-data-summary/
- https://wiki.math.uwaterloo.ca/statwiki/index.php?title=Word_translation_without_parallel_data