From English to Foreign Languages: Transfer Pretrained Language Model

Noskov Fedor
group 792, MIPT

September 28, 2020

Agenda

- Problem
- Setup
- Results

Problem

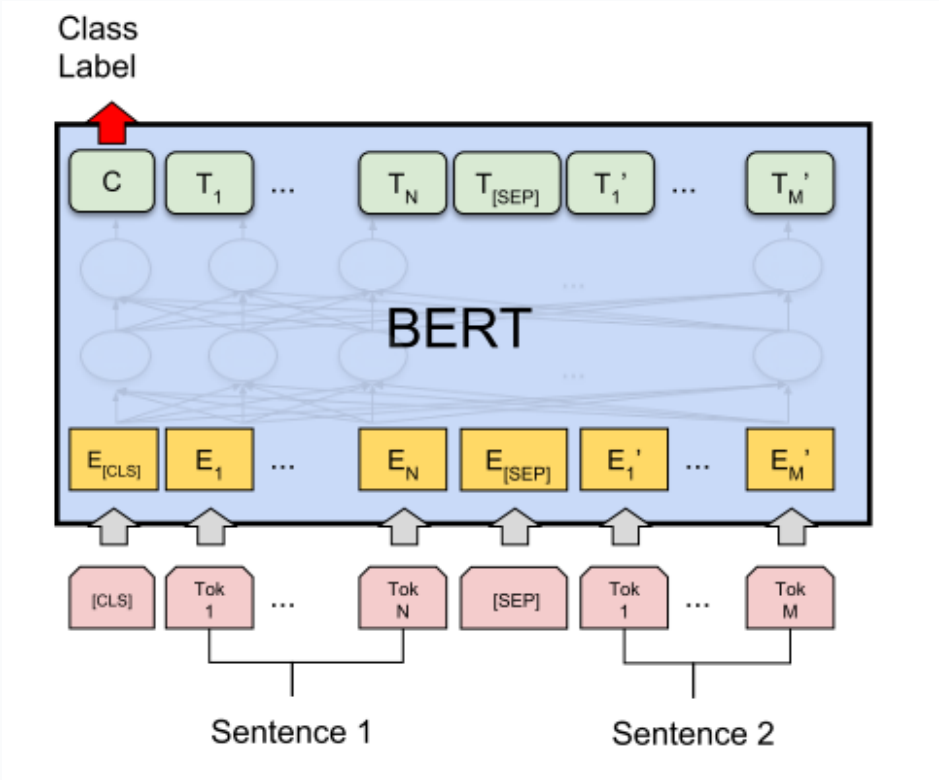

Figure 1: BERT
(src: https://medium.com/the-artificial-impostor/zero-shot-cross-lingual-transfer-with-multilingual-bert-9fe111e02bb)

Setup

Main formula:

$$E'_i = \sum_{j=1}^{n} \alpha_{ij} E_j$$

- E'_i — эмбеддинг для i-того слова в новом языке
- E_j — эмбеддинг BERT'а для английского слова j
- α_{ij} — вероятность перевести i-тое словое j-тым

Setup

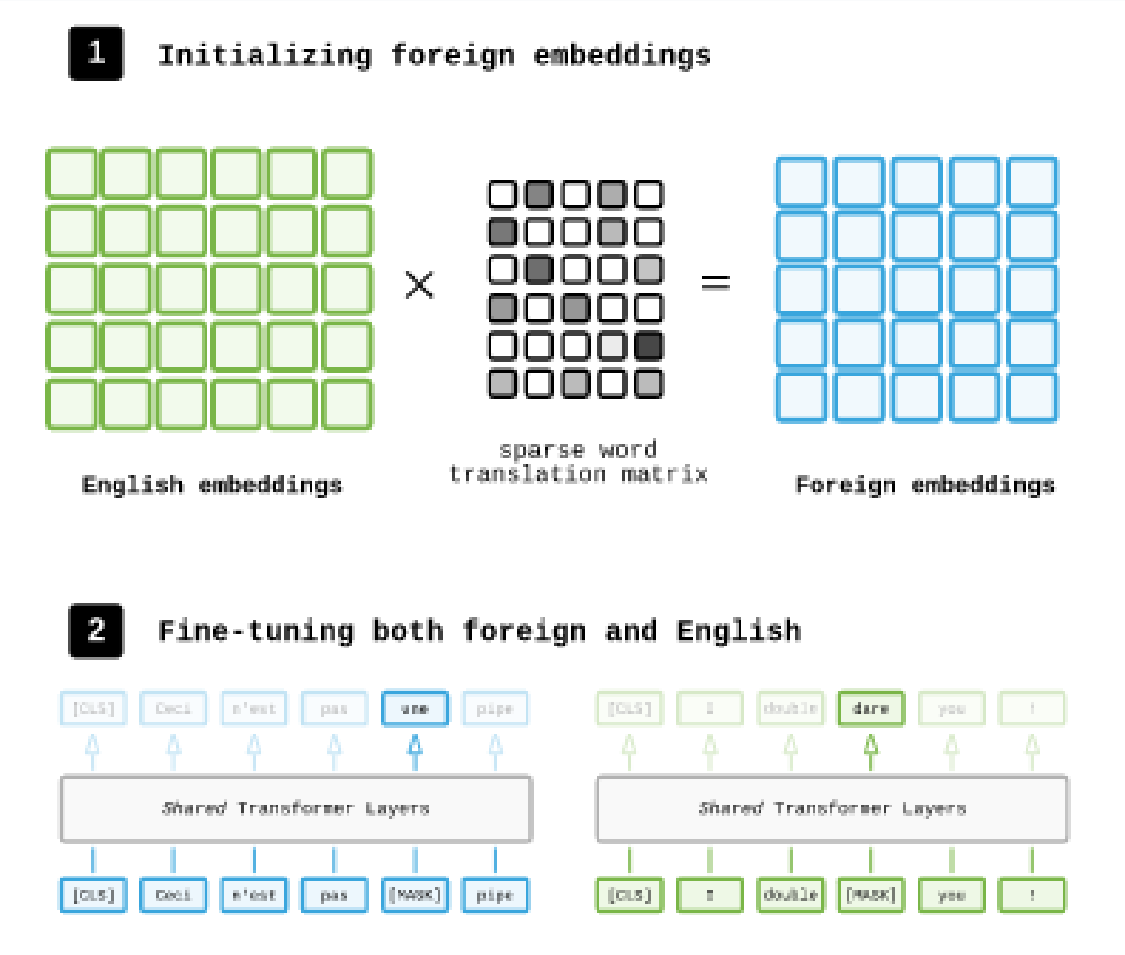

Figure 2: Setup
(src: https://arxiv.org/pdf/2002.07306.pdf)

Results

		fr	vi	zh	ru	ar	hi	avg
(Conneau et al., 2018)	⊖	67.7	66.4	65.8	65.4	64.8	64.1	65.7
(Artetxe and Schwenk, 2019)	⊞	71.9	72.0	71.4	71.5	71.4	65.5	70.6
(Lample and Conneau, 2019) (MLM)	⊖	76.5	72.1	71.9	73.1	68.5	65.7	71.3
(Lample and Conneau, 2019) (MLM+TLM)	⊞	78.7	76.1	76.5	75.3	73.1	**69.6**	74.9
mBERT (Wu and Dredze, 2019)	⊖	73.8	69.5	69.3	69.0	64.9	60.0	67.8
RAMEN$_{\text{BASE}}$								
+ BERT	⊖	75.2	71.8	70.7	71.1	69.3	62.8	70.1
	⊞	75.1	72.5	71.9	70.8	69.7	63.5	70.6
+ RoBERTa	⊖	79.9	75.9	73.7	73.6	71.9	65.6	73.4
	⊞	80.3	75.6	76.2	**75.8**	73.1	68.1	74.9
RAMEN$_{\text{LARGE}}$								
+ BERT	⊖	78.1	74.8	74.5	73.7	70.8	64.5	72.7
	⊞	78.0	75.1	71.3	74.0	71.8	66.1	72.7
+ RoBERTa	⊖	**81.3**	**76.2**	76.3	75.6	**73.5**	64.5	74.6
	⊞	81.0	**76.2**	**76.8**	75.0	72.9	68.2	75.0

Figure 3: Zero-shot classification results on XNLI
(src: https://arxiv.org/pdf/2002.07306.pdf)

Summary

- Problem is solved
- Model is built
- Results achieved SOTA

Instead of conclusion

Figure 4: src: https://www.codemotion.com/magazine/dev-hub/machine-learning-dev/bert-how-google-changed-nlp-and-how-to-benefit-from-this/