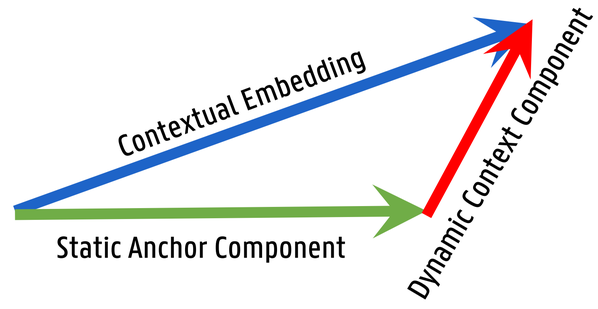
language and translation technology team

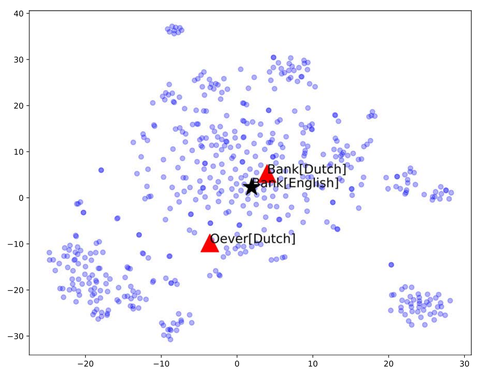
Investigating the Quality of Static Anchor Embeddings from Transformers for Under-Resourced Languages

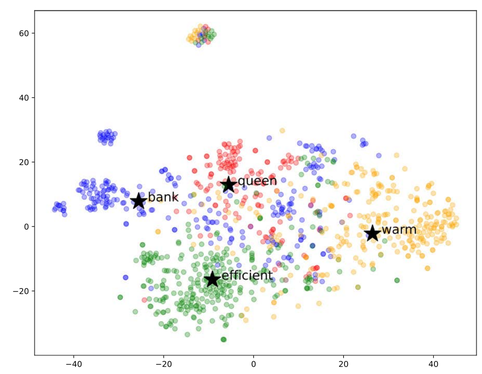
Pranaydeep Singh, Orphée De Clercq, Els Lefever

Can BERT Embeddings be broken down into Static Anchor Components and then aligned for Cross-Lingual Inference?

Step 1: Verify the Static Anchor hypothesis for BERT
Step 2: Evaluate the strengths of the Anchors in monolingual setting
Step 3: Align Anchors from multiple languages, evaluate cross-lingual setting







1

- Obtain Anchors by averaging for all contexts
- BERT models for EN, HI, ZH, RU, NL
- Contexts from 1 million Wikipedia sents

2

- Monolingual evaluation -> Sentiment lexicon for all 5 langs
- Test lexical strength instead of semantic understanding
- 2000 words for training, 400 for testing
- Outcome: FastText > Anchors

Language	FastText	Static Anchors
EN	**0.8425**	0.7575
HI	**0.8125**	0.5625
NL	**0.7300**	0.5750
RU	**0.7575**	0.7175
ZH*	0.5200	**0.5780**

3

- Cross-lingual Setting 1: BLI
  - Eng-Target and Target-Eng setting from MUSE dicts
  - 2 Setups: Full Dict and 1K Pairs
  - Aligned with Vecmap Supervised setting

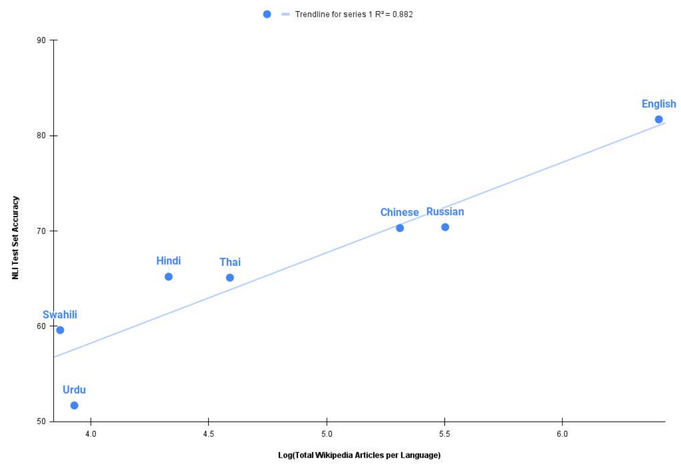
- Cross-lingual Setting 2: XNLI
  - Zero-Shot: Train on EN, Test on Target
  - Data only for HI, RU and ZH
  - No fine-tuning of LM unlike other SOTA approaches

	EN-HI	HI-EN	EN-NL	NL-EN	EN-RU	RU-EN	EN-ZH	ZH-EN
FASTTEXT EMBEDDINGS WITH VECMAP								
Full Train Set	0.5679	0.7098	0.8604	0.8467	0.6465	0.8137	0.8325	0.549
1k Supervision	0.4864	0.5268	0.8234	0.7660	0.5525	0.7561	-	-
ALIGNED ANCHORS WITH VECMAP								
Full Train Set	0.4955	0.5994	0.6382	0.7350	0.6210	0.8043	0.8010	0.4510
1k Supervision	0.3620	0.2997	0.2300	0.3860	0.3276	0.5940	-	-

Model	**HI**	**RU**	**ZH**
XNLI Transfer Learning Baseline	0.563	0.578	0.588
mBERT (Devlin et al., 2019)	0.600	0.638	-
XLM (MLM) (Lample and Conneau, 2019)	0.657	0.731	0.719
MonoTrans (Artetxe et al., 2020)	**0.660**	0.704	0.703
RAMEN (Tran, 2020)	0.656	**0.736**	**0.737**
CL ELMo (Schuster et al., 2019)	0.548	-	-
CL-anchor-BERT	0.583	0.644	0.662

Poor Lexical Perf due to limited vocab vs FastText
Alignment Issues due to Isomorphism assumption
Inadequate Anchors due to low context for some words
Language Performance directly correlated to Wikipedia Data

Better Transfer Languages than English for UR languages
Anchors from Joint Models for Isomorphism
Better Evaluations (MUSE:poor quality)(XNLI: Dense Task)



GHENT UNIVERSITY

References

Schuster, T., Ram, O., Barzilay, R., and Globerson, A. (2019). Cross-lingual alignment of contextual word embeddings, with applications to zero-shot dependency parsing.

Artetxe, M., Labaka, G., and Agirre, E. (2017). Learning bilingual word embeddings with (almost) no bilingual data.

Artetxe, M., Labaka, G., and Agirre, E. (2018). A robust self-learning method for fully unsupervised cross-lingual mappings of word embeddings.