CROSS-LINGUAL ALIGNMENT VS JOINT TRAINING: A COMPARATIVE STUDY AND A SIMPLE UNIFIED FRAMEWORK

Zirui Wang* , Jiateng Xie* , Ruochen Xu, Yiming Yang, Graham Neubig, Jaime Carbonell

ICLR 2020

Previous work

- Two main approaches to learn multi-/cross-lingual embeddings:

  \+ Alignment: learns a mapping between two languages.

  \+ Joint training: learns unified multilingual embeddings via monolingual and cross-lingual objectives.

Previous work: Alignment method

- Given seed parallel embeddings: $\{x_i, y_i\}_{i=1}^{D}$

  - Optimize the objective function w.r.t W:

$$\min_{W} \sum_{i=1}^{D} \ell(\mathbf{W}\mathbf{x}_i, \mathbf{y}_i) \text{ s.t. } WW^\top = I$$

  - Solution: $W = UV^\top$

    where

$$X_D = [x_1, x_2, \cdots, x_D]^\top$$

$$Y_D = [y_1, y_2, \cdots, y_D]^\top$$

$$Y_D^\top X_D = U \sum V^\top$$

Previous work: Joint training

- Use multilingual training data.
- Use shared vocabulary.
- Jointly optimize monolingual and cross-lingual objectives to learn multilingual embeddings.

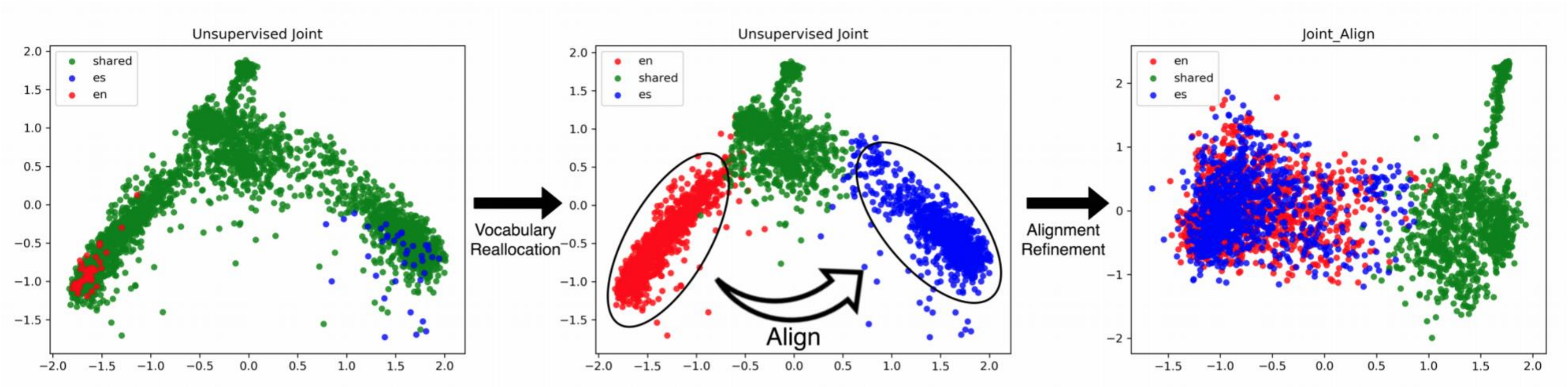
$$\mathcal{L}_J = \mathcal{L}_1 + \mathcal{L}_2 + \mathcal{R}(L_1, L_2)$$

Problem

- Alignment:
  + Relies on two disjoint sets of embeddings.
- Joint training:
  + Overshares among languages.
  + Not leverage bilingual dictionaries.

Proposed framework

- 3-step framework:



Proposed framework

- Vocabulary Reallocation: remove a word w from shared list if:

$$\frac{1-\gamma}{\gamma} \leq r \leq \frac{\gamma}{1-\gamma}$$

where $$r = \frac{T_{L_2}}{T_{L_1}} \cdot \frac{C_{L_1}(w)}{C_{L_2}(w)}$$

Results

	en-es	es-en	en-fr	fr-en	en-de	de-en	en-it	it-en	en-ru	ru-en	en-zh	zh-en	avg
Alignment Methods													
(1) MUSE (Conneau et al., 2018a)	81.7	83.3	82.3	82.1	74.0	72.0	77.7	78.2	44.0	59.1	32.5	31.4	66.5
(2) VECMAP (Artetxe et al., 2018a)	82.3	84.7	82.3	83.6	75.1	74.3	-	-	49.2	65.6	0.0	0.0	-
(3) DeMa-BWE (Zhou et al., 2019)	82.8	84.9	83.1	83.5	77.2	74.4	-	-	49.2	65.7	42.5	37.9	-
(4) Procrustes (Smith et al., 2017)	81.4	82.9	81.1	82.4	73.5	72.4	77.5	77.9	51.7	63.7	42.7	36.7	68.7
(5) GeoMM (Jawanpuria et al., 2019)	81.4	85.5	82.1	84.1	74.7	76.7	77.9	80.9	51.3	67.6	49.1	45.3	71.4
(6) RCSLS (Joulin et al., 2018b)	84.1	86.3	83.3	84.1	79.1	76.3	78.5	79.8	57.9	67.2	45.9	46.4	72.4
(7) RCSLS + IN (Zhang et al., 2019)	83.9	-	**83.9**	-	78.1	-	79.1	-	57.9	-	48.6	-	-
Joint Traing Methods													
(8) Unsupervised Joint	33.4	36.6	42.2	47.4	39.5	41.4	36.8	38.8	4.0	3.5	17.9	10.2	29.3
(9) Supervised Joint (Duong et al., 2016)	79.7	79.8	78.1	76.7	67.5	68.9	74.4	74.1	41.8	51.8	46.7	43.3	65.2
(10) Joint - Replace	48.2	47.7	49.4	52.1	46.5	46.9	43.8	45.8	20.3	36.6	32.7	34.1	42.0
Joint Align Framework													
(11) Joint_Align (w/o AR)	55.9	62.8	61.8	67.0	49.1	54.6	50.2	51.4	8.7	8.2	19.4	18.2	42.3
(12) Joint_Align + MUSE	81.4	84.2	82.8	83.6	74.2	72.2	77.5	81.5	45.0	58.3	36.1	35.3	67.7
(13) Joint_Align + RCSLS (w/o VR)	34.2	37.0	41.2	46.8	34.0	35.6	35.3	35.1	7.7	5.2	20.2	15.7	29.0
(14) Joint_Align + GeoMM	82.6	85.7	82.5	84.2	75.5	77.2	78.2	81.4	52.4	67.7	50.4	46.5	72.0
(15) Joint_Align + RCSLS	**84.7**	**87.9**	83.5	**85.6**	**79.6**	**78.0**	**80.6**	**84.0**	**59.8**	**67.8**	**54.3**	**48.7**	**74.5**