A Simple Approach to Learning Unsupervised Multilingual Embeddings

Pratik Jawanpuria, Mayank Meghwanshi, Bamdev Mishra
Microsoft, India
{pratik.jawanpuria,mamegh,bamdevm}@microsoft.com

Abstract

Recent progress on unsupervised cross-lingual embeddings in the bilingual setting has given the impetus to learning a shared embedding space for several languages. A popular framework to solve the latter problem is to solve the following two sub-problems jointly: 1) learning unsupervised word alignment between several language pairs, and 2) learning how to map the monolingual embeddings of every language to shared multilingual space. In contrast, we propose a simple approach by decoupling the above two sub-problems and solving them separately, one after another, using existing techniques. We show that this proposed approach obtains surprisingly good performance in tasks such as bilingual lexicon induction, cross-lingual word similarity, multilingual document classification, and multilingual dependency parsing. When distant languages are involved, the proposed approach shows robust behavior and outperforms existing unsupervised multilingual word embedding approaches.

1 Introduction

Learning cross-lingual word representations has been the focus of many recent works (Mikolov et al., 2013; Faruqui and Dyer, 2014; Artetxe et al., 2016). It aims at learning a shared embedding space for words across two (bilingual word embedding) or more languages (multilingual word embedding or MWE), by mapping similar words (or concepts) across different languages close to each other in a shared embedding space. Such a representation is useful in various applications such as cross-lingual text classification (Klementiev et al., 2012), building bilingual lexicons (Mikolov et al., 2013), cross-lingual information retrieval (Vulić and Moens, 2015), and machine translation (Gu et al., 2018), to name a few.

Mikolov et al. (2013) showed that the geometric arrangement of word embeddings could be (approximately) preserved by linearly transforming the word embeddings from one language space to another. Subsequently, several works have explored learning bilingual word embeddings in both supervised (Xing et al., 2015; Artetxe et al., 2016, 2018a; Smith et al., 2017; Jawanpuria et al., 2019) and unsupervised (Zhang et al., 2017a,b; Conneau et al., 2018; Artetxe et al., 2018b; Alvarez-Melis and Jaakkola, 2018; Hoshen and Wolf, 2018; Grave et al., 2019; Jawanpuria et al., 2020a) settings.

Representing word embeddings of many languages in a common shared space is desirable to allow knowledge transfer between different languages. Chen and Cardie (2018) are among the first to propose unsupervised learning of MWEs. They extend the GAN-based iterative refinement procedure for learning bilingual word embeddings (Conneau et al., 2018) to the multilingual setting. However, adversarial training has known concerns of optimization stability with distant language pairs (Søgaard et al., 2018). Alaux et al. (2019) propose a joint optimization framework for learning bilingual lexicons and mappings between several pairs of languages. They obtain the bilingual lexicons using the Gromov-Wasserstein approach (Alvarez-Melis and Jaakkola, 2018) and mapping operators between languages using the RCSLS algorithm (Joulin et al., 2018). Heyman et al. (2019) propose to learn the shared multilingual space by incrementally adding languages to it, one in each iteration. Their approach is based on a reformulation of the bilingual self-learning algorithm proposed by Artetxe et al. (2018b).

This work proposes a two-stage framework for learning a shared MWE space in the unsupervised setting. The two stages aim at solving the following sub-problems: a) generating bilingual lexicons between a few pairs of languages, and *subsequently*

Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, pages 2995–3001,
November 16–20, 2020.

b) learning the mapping operators between languages in a shared multilingual space. The subproblems are separately solved using existing techniques. In contrast, existing unsupervised multilingual approaches (Chen and Cardie, 2018; Heyman et al., 2019; Alaux et al., 2019) solve the above subproblems jointly. Though it appears like a simple baseline approach, the proposed framework provides the robustness and versatility often desired while learning an effective multilingual space for distant languages, which is a challenging setting for unsupervised methods (Søgaard et al., 2018; Glavaš et al., 2019; Vulić et al., 2019).

We evaluate our approach on the bilingual lexicon induction (BLI) task, cross-lingual word similarity task, and two downstream multilingual tasks: document classification and dependency parsing. We summarize our findings below.

- For a group consisting of similar languages, all multilingual approaches, including ours, benefit from transfer learning across languages and achieve similar BLI performance.
- In challenging scenarios involving distant languages, existing unsupervised approaches fail to learn an effective multilingual space. The proposed approach, however, is robust and outperforms other multilingual methods in such settings.
- The proposed approach performs better than existing methods on the cross-lingual word similarity, the document classification, and the dependency parsing tasks.

2 Unsupervised Multilingual Multi-stage Framework

We propose the following framework for unsupervised learning of MWEs:

- generate unsupervised word alignment between a few pairs of languages, and then
- use the above knowledge to learn the shared multilingual space.

We solve the above two stages sequentially using known techniques. Our methodology contrasts with the existing unsupervised MWE methods (Alaux et al., 2019; Chen and Cardie, 2018; Heyman et al., 2019), which learn the unsupervised word alignments and the cross-lingual word embedding mappings jointly. Despite its apparent simplicity, we empirically observe that the proposed approach illustrates remarkable generalization ability and robustness. We summarize the proposed approach in Algorithm 1 and discuss the details below.

Algorithm 1 Proposed Algorithmic Framework

Input: Monolingual embeddings $\mathbf{X}_i$ for each language L_i and an undirected, connected graph $G(V, E)$ with $V = \{L_1, \ldots, L_n\}$.

/***Stage 1:** Generate bilingual lexicons $\mathbf{Y}_{ij}$*/
for each unordered pair $(L_i, L_j) \in E$ **do**
  $\mathbf{Y}_{ij} \leftarrow \texttt{UnsupWordAlign}(\mathbf{X}_i, \mathbf{X}_j)$
end for

/***Stage 2:** Learn MWE in a shared latent space*/
Run `GeoMM` on $G(V, E)$ with monolingual embeddings $\mathbf{X}_i$ for all languages L_i and bilingual lexicons $\mathbf{Y}_{ij}$ for language pairs $(L_i, L_j) \in E$

The output of `GeoMM`:
a) metric $\mathbf{B}$ (a positive definite matrix), and
b) orthogonal matrices $\mathbf{U}_i$ $\forall i = 1, \ldots, n$.

/*Represent word embedding x of language L_i in the common multilingual space*/
$x \rightarrow \mathbf{B}^{\frac{1}{2}} \mathbf{U}_i^\top x$.

2.1 Stage 1: Generating Bilingual Lexicons

We first generate bilingual lexicons for a few language pairs using existing unsupervised bilingual word alignment algorithms (Artetxe et al., 2018b; Alvarez-Melis and Jaakkola, 2018). The lexicons are learned in the bilingual setting, independent of each other. Our framework allows using different unsupervised bilingual word alignment algorithms for different language pairs as our second stage is agnostic to this process. More generally, one may obtain bilingual lexicons for language pairs using various algorithms/resources: unsupervised, weakly-supervised with bootstrapping (Artetxe et al., 2017), human supervision, etc. Such flexibility in getting bilingual lexicons is often desirable in real-world applications (Søgaard et al., 2018; Glavaš et al., 2019; Vulić et al., 2019). To the best of our knowledge, existing unsupervised MWE approaches do not discuss[1] applicability to such hybrid settings.

We experiment with two unsupervised bilingual word alignment algorithms (Artetxe et al., 2018b;

[1]Heyman et al. (2019), for example, state that their approach is impractical in the supervised setting as it requires pairwise dictionaries for all pair of languages.

Alvarez-Melis and Jaakkola, 2018) to generate bilingual lexicons, described in Section 2.3. It should be emphasized that the lexicons are learned only for a few language pairs. For instance, in our experiments, $n-1$ bilingual lexicons are generated for n languages.

2.2 Stage 2: Multilingual Word Embeddings

We now learn the MWEs using the bilingual lexicons obtained from the first stage. To achieve our objective, we propose to employ the Geometry-aware Multilingual Mapping (GeoMM) algorithm (Jawanpuria et al., 2019).

The setting of GeoMM may be formalized as an undirected, connected graph, whose nodes represent languages and edges between nodes imply the availability of bilingual dictionaries (for the corresponding language pairs). GeoMM represents multiple languages in a common latent space by learning language-specific rotations for each language ($d \times d$ orthogonal matrix $\mathbf{U}_i$ for each language L_i) and a Mahalanobis metric common across languages (a $d \times d$ symmetric positive-definite matrix $\mathbf{B}$), where d is the dimensionality of the monolingual word embeddings. The rotation matrices align the language embeddings to a common latent space, while the (shared) metric $\mathbf{B}$ governs how distances are measured in this latent space. Both the language-specific parameters ($\mathbf{U}_i\ \forall L_i$) and the shared parameter ($\mathbf{B}$) are learned via a joint optimization problem (Jawanpuria et al., 2019, Equation 3). The function that maps a word embedding x from language L_i's space to the shared latent space is given by: $x \rightarrow \mathbf{B}^{\frac{1}{2}}\mathbf{U}_i^\top x$.

2.3 Implementation Details

We develop two variants of the proposed approach, which differ in the unsupervised bilingual word alignment algorithm employed in the first stage. Both the variants use the GeoMM algorithm in the second stage.

SL-GeoMM: In this method, we employ the self-learning algorithm of Artetxe et al. (2018b) for generating bilingual lexicons (`UnsupWordAlign` subroutine in Algorithm 1). We simplify the self-learning algorithm for our purpose by using its unsupervised initialization followed by stochastic dictionary induction (without any pre/post -processing steps).

GW-GeoMM: We also experiment with the Gromov-Wasserstein (GW) word alignment algorithm (Alvarez-Melis and Jaakkola, 2018) as the `UnsupWordAlign` subroutine in Algorithm 1. The GW algorithm learns a doubly stochastic matrix. To further obtain a bilingual lexicon, we additionally run a CSLS (cross-domain similarity local scaling) based refinement procedure (Conneau et al., 2018).

3 Experiments

The proposed methods **SL-GeoMM** and **GW-GeoMM** are compared against existing unsupervised multilingual word embeddings approaches **UMWE** (Chen and Cardie, 2018) and **UMH** (Alaux et al., 2019) on various BLI and downstream tasks. As a bilingual baseline, we also include state-of-the-art unsupervised bilingual word embeddings approach **BilingUnsup** (Artetxe et al., 2018b) in our BLI experiments. In addition to gauging the effectiveness of the proposed two-staged framework, the experiments also study the multilingual approaches' robustness, especially when distant languages are involved. The evaluated tasks are detailed below.

Bilingual lexicon induction (BLI): We evaluate on the MUSE (Conneau et al., 2018) and the VecMap (Dinu and Baroni, 2015; Artetxe et al., 2018a) datasets. Following (Chen and Cardie, 2018; Alaux et al., 2019), we report Precision@1 in the BLI experiments and employ the CSLS based inference (Conneau et al., 2018).

Cross-lingual word similarity (CLWS): The CLWS task is evaluated using the SemEval 2017 dataset (Camacho-Collados et al., 2017).

Multilingual dependency parsing (MLDP): In this task (Ammar et al., 2016), we evaluate the quality of learned multilingual embeddings on ML-Parsing dataset sampled from the Universal Dependencies 1.1 corpus (Agić et al., 2015). The dataset has twelve languages: Bulgarian, Czech, Danish, German, Greek, English, Spanish, Finnish, French, Hungarian, Italian, and Swedish.

Multilingual document classification (MLDC): This task (Ammar et al., 2016) is evaluated on the ReutersMLDC dataset, which has documents in seven languages: Danish, German, English, Spanish, French, Italian, and Swedish.

More details of the experimental settings and additional results are discussed in the technical report (Jawanpuria et al., 2020b).

	de-xx	en-xx	es-xx	fr-xx	it-xx	pt-xx	xx-de	xx-en	xx-es	xx-fr	xx-it	xx-pt	avg.
SL-GeoMM	**70.5**	80.0	81.7	79.7	**80.9**	**80.9**	**69.9**	80.6	82.3	83.1	79.6	78.2	**79.0**
GW-GeoMM	69.3	80.2	81.2	78.9	80.3	79.9	69.0	**81.7**	81.7	82.0	78.7	76.7	78.3
UMWE	70.4	**80.6**	**82.0**	**79.8**	80.6	80.6	69.5	77.4	**83.5**	**84.1**	**80.4**	**79.0**	**79.0**
UMH	69.2	79.9	81.8	79.4	80.6	80.6	69.0	80.7	82.3	82.8	79.0	77.6	78.6
BilingUnsup	60.9	76.9	75.6	72.7	75.2	75.3	61.6	76.1	77.2	75.9	73.6	72.2	72.8

Table 1: Average Precision@1 for BLI on six European languages from the MUSE dataset. The results are obtained for every combination of source-target language pair.

	cs-xx	da-xx	de-xx	en-xx	es-xx	fr-xx	it-xx	nl-xx	pl-xx	pt-xx	ru-xx	
SL-GeoMM	**65.1**	61.3	**64.1**	**70.2**	**69.3**	**68.1**	**68.7**	**67.4**	**66.0**	**68.7**	**63.3**	
GW-GeoMM	64.6	**61.7**	**64.1**	70.0	**69.3**	68.0	**68.7**	67.1	65.6	68.3	62.4	
UMWE	57.6	54.1	56.8	63.1	62.9	61.5	61.9	0.0	58.6	61.6	56.3	
UMH	63.7	60.8	62.8	68.8	68.9	67.5	68.0	66.1	64.2	67.8	61.9	
BilingUnsup	61.8	58.7	58.4	64.9	65.0	63.3	64.5	63.7	62.0	64.4	59.3	
	xx-cs	xx-da	xx-de	xx-en	xx-es	xx-fr	xx-it	xx-nl	xx-pl	xx-pt	xx-ru	avg.
SL-GeoMM	**53.6**	61.9	**69.5**	**75.0**	**76.3**	**75.7**	**72.4**	**70.1**	**55.2**	**74.0**	**48.4**	**66.6**
GW-GeoMM	53.1	**62.1**	69.3	74.6	**76.3**	75.6	72.3	70.0	55.1	73.8	47.6	66.3
UMWE	49.5	57.6	60.3	63.1	68.4	67.9	65.2	0.0	51.0	66.3	45.0	54.0
UMH	52.9	60.4	68.3	74.1	75.6	74.6	71.4	68.6	54.8	72.6	47.4	65.5
BilingUnsup	51.0	56.6	64.5	69.6	71.7	70.1	67.7	66.1	53.6	68.8	46.3	62.4

Table 2: Average Precision@1 for BLI on eleven European languages from the MUSE dataset. The results are obtained for every combination of source-target language pair.

3.1 Results on Standard BLI Setting

Table 1 reports the BLI results on a group of six relatively close European languages (Alaux et al., 2019): German, English, Spanish, French, Italian, and Portuguese. We observe that the proposed two-stage methods, GW-GeoMM and SL-GeoMM, obtain scores on par with state-of-the-art methods, UMWE and UMH. Thus, multilingual approaches can learn an effective multilingual space for close-by languages. We also observe that all the multilingual approaches outperform BilingUnsup, highlighting the benefits of transfer learning.

3.2 Results on Robust BLI Setting

We evaluate the robustness of the methods to distant languages by including five other European languages (Czech, Danish, Dutch, Polish, Russian) (Alaux et al., 2019) to the previous setup. Table 2 reports the summarized results. The proposed methods, GW-GeoMM and SL-GeoMM, perform better than UMH and UMWE for every language. We also observe that UMWE fails at mapping Dutch language embeddings in the multilingual space even though Dutch is close to English. However, in a separate bilingual experiment, UMWE learns an effective English-Dutch cross-lingual space (obtaining an average en-nl and nl-en score of 75.2). This contrasting behavior of the GAN-based UMWE algorithm between the bilingual and multilingual settings is possibly due to its optimization instability (Søgaard et al., 2018).

We also evaluate the methods in a highly diverse language group: Arabic, German, English, French, Hindi, and Russian. Table 3 reports the BLI performance on each language pair. We observe that the proposed SL-GeoMM learns a highly effective multilingual space and obtains the best overall result, illustrating its robustness in this challenging setting. On the other hand, other multilingual approaches fail to learn a reasonably good multilingual space. For instance, GW-GeoMM, UMWE, and UMH fail to obtain a good BLI score (< 1 Precision@1) in 10, 16, and 18 language pairs, respectively. Below, we analyze their results.

- The Gromov-Wasserstein alignment algorithm (Alvarez-Melis and Jaakkola, 2018), used in the first stage of GW-GeoMM, fails to align English and Hindi words. However, this misalignment does not adversely affect GW-GeoMM on language pairs not involving Hindi as GW-GeoMM performs similar to SL-GeoMM on those language pairs.
- UMH employs the Gromov-Wasserstein (GW) alignment formulation in its joint learning framework. As observed with GW-GeoMM, UMH also does not learn suitable Hindi embeddings in the MWE space. However, UMH also fails to learn

	ar-de	ar-en	ar-fr	ar-hi	ar-ru	de-en	de-fr	de-hi	de-ru	en-fr	en-hi	en-ru	fr-hi	fr-ru	hi-ru	
SL-GeoMM	46.2	49.5	56.5	39.4	34.1	74.6	75.2	**38.4**	**45.2**	**82.5**	**39.0**	**49.7**	**42.7**	**47.1**	**29.7**	
GW-GeoMM	**46.5**	**50.5**	58.1	0.0	33.6	74.0	**75.5**	0.0	44.4	**82.5**	0.0	47.7	0.0	46.2	0.0	
UMWE	0.0	45.0	**58.4**	**41.4**	0.0	0.0	0.0	0.0	40.2	81.9	36.4	0.0	42.6	0.0	0.0	
UMH	0.1	0.1	0.3	0.0	0.1	**74.7**	72.5	0.0	44.7	82.0	0.0	46.5	0.0	44.5	0.0	
BilingUnsup	**46.5**	46.4	55.0	36.9	**35.2**	70.8	61.9	31.8	43.6	79.8	31.3	44.1	36.1	44.9	24.9	
	de-ar	en-ar	fr-ar	hi-ar	ru-ar	en-de	fr-de	hi-de	ru-de	fr-en	hi-en	ru-en	hi-fr	ru-fr	ru-hi	avg.
SL-GeoMM	31.1	35.5	37.4	**29.7**	33.9	**75.1**	**70.7**	**45.5**	**61.7**	82.9	**47.6**	**65.6**	**51.9**	**66.6**	**39.9**	**50.8**
GW-GeoMM	**31.5**	35.6	37.5	0.0	32.8	74.6	70.5	0.0	61.3	83.1	0.0	62.9	0.0	65.8	0.0	37.2
UMWE	0.1	**37.6**	**39.8**	23.7	0.1	0.0	0.0	0.0	55.7	79.5	34.1	0.0	48.4	0.0	0.0	22.2
UMH	0.2	0.1	0.2	0.0	0.1	74.3	69.9	0.0	60.8	**83.2**	0.0	62.5	0.0	65.1	0.0	26.1
BilingUnsup	30.8	29.4	37.7	28.7	**35.0**	72.0	61.3	42.0	59.6	78.7	37.6	59.2	45.4	62.6	32.3	46.7

Table 3: Average Precision@1 for BLI on a diverse group of six languages (MUSE dataset). The results are obtained for every combination of source-target pair.

	CLWS	MLDC	MLDP
SL-GeoMM	0.724	**90.3**	**71.0**
GW-GeoMM	**0.725**	89.7	69.9
UMWE	0.706	88.3	**71.0**
UMH	0.718	90.0	70.6

Table 4: Average Spearman correlation, average accuracy, and average unlabeled attachment score (UAS) on the CLWS, MLDC, and MLDP tasks, respectively.

suitable Arabic embeddings in the MWE space even though the GW algorithm learns an effective bilingual alignment of English and Arabic words. Misalignment of one language's embeddings in the MWE space adversely affects other languages in the joint learning approaches like UMH.

• The GAN-based approach, UMWE, learns two groups of aligned languages in the shared multilingual space. The first group consists of Arabic, English, French, and Hindi languages. However, these languages are misaligned with the other group consisting of German and Russian. Such grouping cannot be attributed to language similarity (e.g., English and German are closer than English and Arabic) and maybe an outcome of optimization stability (Søgaard et al., 2018).

3.3 Cross-lingual Word Similarity Results

Table 4, first column, reports the SemEval 2017 cross-lingual word similarity (CLWS) task's results on four languages: English, German, Spanish, and Italian. For each method, we consider the MWEs of the four languages learned in the second BLI experiment (corresponding to Table 2) for the CLWS evaluation. We observe that the proposed approaches, SL-GeoMM and GW-GeoMM, obtain the best results.

3.4 Results on Downstream Applications

For each multilingual method, we first learn a shared multilingual space (as in BLI setup), followed by application-specific evaluation. Table 4, second and third columns, reports the multilingual document classification (MLDC) and multilingual document parsing (MLDP) tasks' performance, respectively. We observe that both the proposed two-stage approaches perform well on the downstream tasks with SL-GeoMM obtaining the best results.

4 Conclusion

We study a two-stage framework for learning unsupervised multilingual word embeddings. The two stages correspond to unsupervised generation of bilingual lexicons for a few language pairs and subsequently learning a shared latent multilingual space. We propose to solve each of them with existing techniques (Artetxe et al., 2018b; Alvarez-Melis and Jaakkola, 2018; Jawanpuria et al., 2019). Though the proposed framework seems simple compared to the joint optimization methods (Chen and Cardie, 2018; Alaux et al., 2019; Heyman et al., 2019), our main contribution has been to show that it is a strong performer. Empirical results on several different benchmarks on bilingual lexicon induction, cross-lingual word similarity, multilingual document classification, and multilingual document parsing tasks show remarkably good performance and robustness of the proposed framework. The proposed framework has the flexibility to be easily employed in hybrid setups where supervision is available for a few language pairs but is unavailable for others. Overall, our results encourage the development of simple multi-stage models for learning multilingual word embeddings.

References

Željko Agić, Maria Jesus Aranzabe, Aitziber Atutxa, Cristina Bosco, Jinho Choi, Marie-Catherine de Marneffe, Timothy Dozat, Richrd Farkas, Jennifer Foster, Filip Ginter, Iakes Goenaga, Koldo Gojenola, Yoav Goldberg, Jan Hajič, Anders Trrup Johannsen, Jenna Kanerva, Juha Kuokkala, Veronika Laippala, Alessandro Lenci, Krister Lindén, Nikola Ljubešić, Teresa Lynn, Christopher Manning, Hctor Alonso Martínez, Ryan McDonald, Anna Missilä, Simonetta Montemagni, Joakim Nivre, Hanna Nurmi, Petya Osenova, Slav Petrov, Jussi Piitulainen, Barbara Plank, Prokopis Prokopidis, Sampo Pyysalo, Wolfgang Seeker, Mojgan Seraji, Natalia Silveira, Maria Simi, Kiril Simov, Aaron Smith, Reut Tsarfaty, Veronika Vincze, and Daniel Zeman. 2015. Universal Dependencies 1.1.

Jean Alaux, Edouard Grave, Marco Cuturi, and Armand Joulin. 2019. Unsupervised hyper-alignment for multilingual word embeddings. In *Proceedings of the International Conference on Learning Representations*. URL: `https://github.com/facebookresearch/fastText/tree/master/alignment`.

David Alvarez-Melis and Tommi S. Jaakkola. 2018. Gromov-wasserstein alignment of word embedding spaces. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. URL: https://github.com/dmelis/otalign.

Waleed Ammar, George Mulcaire, Yulia Tsvetkov, Guillaume Lample, Chris Dyer, and Noah A. Smith. 2016. Massively multilingual word embeddings. Technical report, arXiv preprint arXiv:1602.01925. URL: https://github.com/wammar/multilingual-embeddings-eval-portal.

Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2016. Learning principled bilingual mappings of word embeddings while preserving monolingual invariance. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 2289–2294.

Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2017. Learning bilingual word embeddings with (almost) no bilingual data. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 451–462.

Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2018a. Generalizing and improving bilingual word embedding mappings with a multi-step framework of linear transformations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 5012–5019.

Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2018b. A robust self-learning method for fully unsupervised cross-lingual mappings of word embeddings. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 789–798. URL: https://github.com/artetxem/vecmap.

Jose Camacho-Collados, Mohammad Taher Pilehvar, Nigel Collier, and Roberto Navigli. 2017. Semeval-2017 task 2: Multilingual and cross-lingual semantic word similarity. In *Proceedings of the 11th International Workshop on Semantic Evaluation*.

Xilun Chen and Claire Cardie. 2018. Unsupervise multilingual word embeddings. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. URL: https://github.com/ccsasuke/umwe.

Alexis Conneau, Guillaume Lample, Marc'Aurelio Ranzato, Ludovic Denoyer, and Hervé Jégou. 2018. Word translation without parallel data. In *Proceedings of the International Conference on Learning Representations*. URL: https://github.com/facebookresearch/MUSE.

Georgiana Dinu and Marco Baroni. 2015. Improving zero-shot learning by mitigating the hubness problem. In *Workshop track of International Conference on Learning Representations*.

Manaal Faruqui and Chris Dyer. 2014. Improving vector space word representations using multilingual correlation. In *Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics*, pages 462–471.

Goran Glavaš, Robert Litschko, Sebastian Ruder, and Ivan Vulić. 2019. How to (properly) evaluate cross-lingual word embeddings: On strong baselines, comparative analyses, and some misconceptions. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 710–712.

Edouard Grave, Armand Joulin, and Quentin Berthet. 2019. Unsupervised alignment of embeddings with Wasserstein Procrustes. In *International Conference on Artificial Intelligence and Statistics*.

Jiatao Gu, Hany Hassan, Jacob Devlin, and Victor OK Li. 2018. Universal neural machine translation for extremely low resource languages. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

Geert Heyman, Bregt Verreet, Ivan Vulić, and Marie-Francine Moens. 2019. Learning unsupervised multilingual word embeddings with incremental multilingual hubs. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, pages 1890–1902.

Yedid Hoshen and Lior Wolf. 2018. Non-adversarial unsupervised word translation. Technical report, arXiv preprint arXiv:1801.06126v3.

Pratik Jawanpuria, Arjun Balgovind, Anoop Kunchukuttan, and Bamdev Mishra. 2019. Learning multilingual word embeddings in latent metric space: A geometric approach. *Transactions of the Association for Computational Linguistics*, 7:107–120.

Pratik Jawanpuria, Mayank Meghwanshi, and Bamdev Mishra. 2020a. Geometry-aware domain adaptation for unsupervised alignment of word embeddings. In *Annual Meeting of the Association for Computational Linguistics (ACL)*.

Pratik Jawanpuria, Mayank Meghwanshi, and Bamdev Mishra. 2020b. A simple approach to learning unsupervised multilingual embeddings. Technical report, arXiv preprint arXiv:2004.05991.

Armand Joulin, Piotr Bojanowski, Tomas Mikolov, Edouard Grave, and Hervé Jégou. 2018. Loss in translation: Learning bilingual word mapping with a retrieval criterion. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*.

Alexandre Klementiev, Ivan Titov, and Binod Bhattarai. 2012. Inducing crosslingual distributed representations of words. In *Proceedings of the International Conference on Computational Linguistics: Technical Papers*, pages 1459–1474.

Tomas Mikolov, Quoc V Le, and Ilya Sutskever. 2013. Exploiting similarities among languages for machine translation. Technical report, arXiv preprint arXiv:1309.4168.

Samuel L. Smith, David H. P. Turban, Steven Hamblin, and Nils Y. Hammerla. 2017. Aligning the fastText vectors of 78 languages. URL: https://github.com/Babylonpartners/fastText_multilingual.

Anders Søgaard, Sebastian Ruder, and Ivan Vulić. 2018. On the limitations of unsupervised bilingual dictionary induction. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pages 778–788.

Ivan Vulić, Goran Glavaš, Roi Reichart, and Anna Korhonen. 2019. Do we really need fully unsupervised cross-lingual embeddings? In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*.

Ivan Vulić and Marie-Francine Moens. 2015. Monolingual and cross-lingual information retrieval models based on (bilingual) word embeddings. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 363–372.

Chao Xing, Dong Wang, Chao Liu, and Yiye Lin. 2015. Normalized word embedding and orthogonal transform for bilingual word translation. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1006–1011.

Meng Zhang, Yang Liu, Huanbo Luan, and Maosong Sun. 2017a. Adversarial training for unsupervised bilingual lexicon induction. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 1959–1970.

Meng Zhang, Yang Liu, Huanbo Luan, and Maosong Sun. 2017b. Earth mover's distance minimization for unsupervised bilingual lexicon induction. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1934–1945.