Unitary Entity Representations for Knowledge Base Completion

Anonymous EMNLP submission

Abstract

Word alignment is an important task. Encounter noisyness. We propose a novel method to mitigate the noisyness. We varify effectiveness of our approach.

1 Introduction

Multi-lingual word alignment is important to transferring knowledge between the languages by the carrier of word, which is an application of domain adaption. [COMMENT: [1] What is main problem in this paper? Domain adaption Word align or else? Unclear. [2] I can't understand what "by the carrier of word, which is an application of domain adaption" means.] [TO DO: clarify main massage] Based on the fundamental concept in cognitive language that the physiological basis of human beings plays an important role in formation of concepts and language (Add reference here), even the disparate origin of the languages has led to barrier of comprehension, same cognitive mechanism determine the feasibility to transfer the knowledge without supervision. [COMMENT: [1] I can't understant grammar of this sentence. [2] Do not use unpractical words such as "cognitive language" and "physiological basis of human beings". [3] I can't understand what this sentense means.] [TO DO: fix grammar] For example, with the cognitive concept that the capital is a center of a entity in country category, " Tomboy " in French should have a high similarity with " London " in English. [COMMENT: [1]Not use "cognitive concept". [2]I can't understand what this example explains. [3] What is the main problem? what is the role of this example?] [TO DO: clarify them.]

There are several attempts on applying word distribution model in multi-lingual alignment model, which we call it multi-lingual distribution model. [COMMENT: [1] What is "word distribution model"? [2] What is the Research Question in this paper? "word distribution model" is related to it? Why you explain several attempts about it?] [TO DO: You MUST explain our motivations BEFORE describe technical facters.]

Word embedding has obtained huge success in wide range NLP downstream tasks like sentiment classification and reading comprehension e. T C (Add reference here) On the other side, multilingual distribution model aims at solving multilingual tasks ranging from as machine translation to cross-lingual search engine, which try to get several monolingual embeddings and their correspondences. Multi-lingual distribution model has been proved to be useful in cross-lingual transfer learning (Mikolov et al, 2013).

Mikolov et al. (2013a) found that linear relationship exists on two words embedding spaces by an empirical research, and show learning the linear mapping with a small starting dictionary can overmatch the neural network. As a further work, orthogonal transformations have been shown effective because its ability to keep distance invariance after transformation (Xing et al, 2015).

Nevertheless, in realistic conditions, there still exist some language pairs which don not have too much parallel data for learning. To figure out the problem, semi-supervised and unsupervised methods are developed recently. They have obtained obvious result even better than supervised methods. We focus on full-unsupervised method of this research.

Word embeddings suffer from drawback of uneven quality cause by different frequency of words. In unsupervised settings, because the word embeddings are the only resource we can use, this problem tends to be a significant factor influencing the accuracy of alignment. To settle this problem, (Conneau et, 2017) only use mutual nearest words pair through training to generate orthogonal mapping. On the other side, (David et, 2018) explicitly sets weight to decay rate by frequency as hyper parameters to each language to abate the influence of low frequency words.

However, all of these methods only consider the deficiency in monolingual but pay no attention to the noise in bilingual alignment. There are two main difficulties in bilingual alignment, the first originated from different semantic of identical entity by cultural difference. For example, different countries may have distinctive attitudes to an entity like color and person. Secondly, high frequency words not always have exact correspondence with / between the other language such as auxiliary word and symbol which have less information for alignment.

In this research, to settle this problem, we use Sinkhorn algorithm (Sinkhorn, 1964) to accustom weights of all words which can reduce transformation loss between two embedding sets. Our method achieves remarkable result and can observably improve the robustness of model to different experiment settings.

The contributions to this paper are as follows :

- We give a systematic introduction to Multilingual distribution models.
- We present a novel model by modifying the weight to loss of each word automatically and achieve better result than previous State-of-art works.

2 Related Work

awesome related work!

We need introduce previous work related to unsupervised word alignment in order.

2.1 Supervised Word Alignment

In supervised sett, since (Miklov, 2013) show the availability of linear mapping in embedding alignment, many other studies has / had tried to refine the transformation function to improve alignment accuracy. (Xing, 2015) put forward the inconsistency problem caused by improper distance metric of inner product, and solve it by replacing inner product by normalization and orthogonal transform. Beside mapping with a transformation matrix directly, (Fairqui, 2014; Artetxe, 2016; Lifu, 2018) tried mapping different languages into a common space. Some other mapping methods also have been proposed, for example, (Lazaridou, 2015) use max-margin to avoid hubness problem and non-linear projection method has also been explored by (Chandar A P, 2014).

To face up to situation with small and even no seed dictionary, lots of prior works gave several solutions. Start from the learning mechanism, self-learning (Artetxe, 2016; Artetxe, 2017) has shown its power to align the words with little supervised data. On the other side, motivated by finding seed pairs automatically by heuristic method, (Chandar A P, 2014) tried to use document-level parallel corpus to help training the auto-encoder between the two language sentences. On the contrary of such explicit method, (Hauer, 2017; Smith, 2017) simply use similar spelling and identical character to generate new seed dictionary pairs. (Vulic, 2016) analyzed the importance of seed dictionary in word alignment, and demonstrate that some external resource of word-level alignment is unnecessary, and some inexpensive automatically induced lexicons even do better. Further more, (Hauer, 2017) use google translation system to generate the seed word pairs and also acquire good result.

2.2 Unsupervised Word Alignment

Although the demand for seed dictionary has descended a lot, various alignment tools and corpus are still necessary. Hence, unsupervised methods are explored recently, all in all, most of them can be categorized into three types. Firstly, about GAN based method, (Miceli, 2016) use GAN to train generator to mapping source embedding into target space and discriminator to classify the mapped embeddings and real target embeddings. Base on this method, (Zhang, 2017a) use Wasserstein GAN, (Conneau, 2017) replace linear mapping layer to replace the origin model and (Zhang, 2017b) add orthogonal parameterization to the model. On the other side, (Zhang, 2017a) adopted a totally different method which regards word alignment as an optimal transport problem. On basis of this work, (Grave, 2018) use batch iteration to solve the OT problem by stinkhorn algorithm, however, it suffers from optimization problem due to the uncertainty of distance between two spaces. Moreover, (Melis, 2018) adopted Gromov-Wasserstein method to get rid of the problem mentioned above. Finally, there are also two methods use statistical properties to align embedding (Cao, 2016; Artetxe, 2018).

2.3 Importance weights in Previous work

To solve the uneven quality problems of word embeddings, most supervised methods only use the high frequency words to learn the transformation mapping. On the unsupervised settings, (Conneau, 2017) propose mutual nearest neighbourhood method which only uses / used reliable word pairs learned by model to acquire orthogonal matrix and achieve great generalization on less frequency words alignment. (Melis, 2018) decrease the influence of low frequency words by sett hyper parameter to decay weight by words frequency manually.

3 Proposal

awesome proposal!

4 Experiments

awesome experiments.

We need to show effectiveness of our proposal compared with previous work. what kinds of experiments are needed?

1 word embedddings are noisy

2 noisy embeddings are harmful for word alignment

3 removing noisy embedding is helpful for word alignment

4 error analysis

5 Conclusion

awesome conclusion! (Aho and Ullman, 1972)

References

Alfred V. Aho and Jeffrey D. Ullman. 1972. *The Theory of Parsing, Translation and Compiling*, volume 1. Prentice-Hall, Englewood Cliffs, NJ.