For low-resource languages, these large datasets don’t exist. Bilingual lexical induction, the task of translating individual words or phrases without relying on parallel text, is especially useful in these settings.

Many foreign-language speakers use the internet—and upload images—on websites in their native language. Is it possible to learn word translations from these images?

We show that incorporating image similarity into a state-of-the-art word translation system improves translation accuracy.

Introduction

For high-resource languages, machine translation is made possible due to large parallel text corpora. For low-resource languages, these large datasets don’t exist. Bilingual lexical induction, the task of translating individual words or phrases without relying on parallel text, is especially useful in these settings.

Many foreign-language speakers use the internet—and upload images—on websites in their native language. Is it possible to learn word translations from these images?

We show that incorporating image similarity into a state-of-the-art word translation system improves translation accuracy.

When are images useful?

Images are most helpful for translating concrete words. They are less useful for abstract words.

By building a model which predicts word concreteness, we can selectively choose when to rely on images for word translation.

We trained a 2-layer perceptron to predict word concreteness using images corresponding to the 40,000 English words annotated by (Brybaet al. 2014) as training data.

input to the model was the feature-wise mean and standard deviation of the images’ Alexnet embeddings.

Top-1 accuracy across a selection of high-resource and low-resource languages. Word concreteness predictions seem to be less useful for low-resource languages.

Extending Text-based Systems

• We extended a state-of-the-art text-based system for word translation by (Wijaya et al. 2017) that uses Bayesian Personalized Ranking (BPR).

• BPR’s translation rankings are reranked using image similarity (IMG), concreteness scores (Cnc), as well as the original text-based features (TXT) as input.

• This is done by training a 2-layer preceptor to, given a word and candidate translation, classify whether the translation is correct.

Our predicted concreteness scores allow us to analyze the gradual degradation in word translation performance as words become more abstract.

Discussion

• We introduce a challenge set for word-translation using images.

• We show that weakly annotated images provide substantial signal for word translation, and our dataset is the first to include images for low-resource languages.

• Future work should show the applicability of image data more broadly in multi-model and resource-scarce NLP.

• Our dataset furthers research on the ability of images to represent parts-of-speech beyond nouns; most existing vision datasets focus only on concrete objects or adjectives in a limited subject scope.

• Our model for predicting concreteness allows for more nuanced analysis of word translation quality.


