Fancy Title

Benno Kruit
10576223
benno.kruit@student.uva.nl

Sara Veldhoen
10545298
sara.veldhoen@student.uva.nl

Abstract

This is an abstract

1 Introduction

2 Related Work

Some research has focused on the induction of multilingual word embeddings, using both different techniques to obtain word representations, and different approaches to the cross-lingual aspects. The evaluation methods applied also vary a lot.

2.1 Linear Mapping

According to Mikolov et al. (Mikolov et al., 2013), the vector space of word representations in different languages are geometrically similar, because words in languages are grounded in real world concepts. It is therefore possible to find a linear mapping between these vector spaces.

The approach is to first train word embeddings on large monolingual data for both languages separately, using the `word2vec` implementation. In the reported experiments, the so-called CBOW architecture is used, that predicts a word given its context in both directions. Notably, the authors also propose a way to include some phrases: multi-word expressions. This may prove useful for translation, as one multiple words can together express a concept that has a single word in another language.

Using a relatively small set of gold standard word translations, in this case obtained from Google Translate, a transformation matrix W is searched. The training objective is to minimize the distance between words that are translations of one another.

The evaluation is performed on a test set of gold-standard word translations, again from Google Translate. The word representation in the source language is transformed using W, and a ranked list of the nearest words in the target language is the output. The precision at ranks 1 and 5 is reported.

2.2 Multitask Learning

Distributed representations for a pair of languages are induced jointly by Klementiev and Titov (Klementiev et al., 2012). Words in both languages are represented in a single vector space.

The induction is treated as a multitask learning problem where each task corresponds to a single word. The training influences other tasks depending on the task-relatedness. The latter is derived from co-occurrence statistics in bilingual parallel data: the number of alignment links between that word and its (supposed) translations.

The word representations are induced in a neural language model architecture. The n preceding words form the context, their representations are concatenated to form a context vector. The probability of the next word occuring is predicted from this vector. The training procedure aims to find the word representations that minimize the data (log) likelihood: $L(\theta) = \sum_{t=1}^{T} \log \hat{P}_\theta(w_t|w_{t-n+1:t-1})$.

The method is evaluated on a real-world task: crosslingual document classification. Topic annotations are available for documents in one of the languages, and the system predicts the topics of documents in the other language. The jointly induced word representation outperform two other approaches to the problem: glossing (where every word in the document is translated separately, based on word alignments) and Machine Translation.

2.3 Joint Learning from Sentence Embeddings

Unlike the previous approaches, Hermann and Blunsum (Hermann and Blunsom, 2013) start from sentence alignments, which share the same semantics. The assumption is that some function

can describe the composition of word embeddings into a sentence embedding. For the sake of argument, the authors use a simple bag-of-words additive interpretation of composition. The word embeddings are induced jointly for both languages from these sentence-embeddings, by minimizing the distance between both sums of word embeddings. In order to make sure the weights won't be reduced to zero, similarity between unaligned sentence embeddings is penalized.

The same evaluation as in (Klementiev et al., 2012) is applied, i.e. the document classification task. Furthermore, the authors present a graphical qualitative analysis. In (Hermann, 2014), one of the authors expands this approach by evaluating on a larger number of language pairs.

3 Experiments

3.1 Evaluation

We evaluate our mutlilingual word embeddings on the same real-world task as (Klementiev et al., 2012) and (Hermann and Blunsom, 2013): crosslingual document classification. The task is based on Reuters corpora, which has topic-annotated documents in English, French, German, Italian and Spanish. These languages are also in the Europarl data. Only documents that are assigned a single topic are used.

Each document is represented by the average of the representations of its tokens (in (Klementiev et al., 2012)), or sentences (in (Hermann and Blunsom, 2013)). An averaged version of the perceptron algorithm is trained for document classification in one language, and tested on data in another.

We thus compute classification accuracy to compare the existing approaches described above to each other and to the new models introduced in **??**.

4 Results and discussion

5 Conclusion

References

[Hermann and Blunsom2013] Karl Moritz Hermann and Phil Blunsom. 2013. Multilingual distributed representations without word alignment.

[Hermann2014] Karl Moritz Hermann. 2014. *Distributed Representations for Compositional Semantics*. Ph.D. thesis.

[Klementiev et al.2012] Alexandre Klementiev, Ivan Titov, and Binod Bhattarai. 2012. Inducing crosslingual distributed representations of words.

[Mikolov et al.2013] Tomas Mikolov, Quoc V Le, and Ilya Sutskever. 2013. Exploiting similarities among languages for machine translation.