Evaluating Sentence Representations

3rd Practical, Unsupervised Language Learning

18 April 2018

1   Introduction

In the 2nd practical, you implemented and trained three different models to learn the word embeddings: The skip-gram [3], the Bayesian skip-gram [1], and Embed-Align [5]. You have evaluated the performance of these three models on the lexical substitution task. In this practical, your task is to compare these models using **SentEval** [2]. SentEval, facebook evaluation toolkit for sentence embeddings, is a library for evaluating the quality of sentence embeddings by applying them on a broad and diverse set of downstream tasks called "transfer" tasks. The reason they are called transfer tasks is that the sentence embeddings are not explicitly optimized on them.

For this practical, you do not need to use your own implementations of the models. And you need to make sure to use models that are properly trained on a large enough corpus. We will provide you with the pre-trained models of Embed-Align and BSG. For the skip-gram model, you can use an implementation from any other source, but you should acknowledge the source and also you should train it on the English side of the Europarl (from practical 2). That's to make the models more comparable.

2   Sentence Embeddings

So far, you have been working with word embeddings. For this practical, first, you need to obtain sentence embeddings, vector representations of sentences, from the word embedding. E.g. in case of the skip-gram model, the simplest thing you can do is to average the word vectors of all the words in a given sentence. Next, you will use SentEval to investigate the qualities and properties of the sentence embeddings.

3   Resources

Here are some useful resources to help you get started.

- The SentEval paper:

- https://arxiv.org/pdf/1803.05449.pdf

- The link to the SentEval repository on github:

  - https://github.com/facebookresearch/SentEval

- A tutorial on how to install and use SentEval to evaluate GloVe [4], when GloVe word vectors are averaged for the words in a sentence to compute the sentence representation:

  - https://uva-slpl.github.io/ull/resources/practicals/practical3/senteval_example.ipynb

- A notebook to test a pre-trained EmbedAlign on SentEval

  - https://github.com/uva-slpl/embedalign/blob/master/notebooks/senteval_embedalign.ipynb

4 Write a report

Write a 4 page report using the TACL style files, briefly explaining:

- the methods

- the evaluation results

- what you have learned about the properties of the three word representation models

More advice on how to write a report can be found here: https://uva-slpl.github.io/ull/resources/slides/ULL-final-report.pdf

You can obtain the TACL style files from here:

- https://www.transacl.org/ojs/index.php/tacl/about/submissions#authorGuidelines

Submit your report in a pdf format on blackboard with the title **ULL-Practical3-FullName1_FullName2**.

Submission deadline for the report is **23:59 on Thursday, 31 May**. Submit only one report per group. Upload your codes on a github repository and put a link to the repository in your report. Also, add the instructions on how to run your codes for each part of the practical to the github ReadMe.

References

[1] Arthur Bražinskas, Serhii Havrylov, and Ivan Titov. Embedding words as distributions with a bayesian skip-gram model. *arXiv preprint arXiv:1711.11027*, 2017.

[2] Alexis Conneau and Douwe Kiela. Senteval: An evaluation toolkit for universal sentence representations. *arXiv preprint arXiv:1803.05449*, 2018.

[3] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[4] Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.

[5] Miguel Rios, Wilker Aziz, and Khalil Sima'an. Deep generative model for joint alignment and word representation. *arXiv preprint arXiv:1802.05883*, 2018.