Improving Distributional Similarity with Lessons Learned from Word Embeddings

Omer Levy Yoav Goldberg Ido Dagan
Computer Science Department
Bar-Ilan University
Ramat-Gan, Israel
{omerlevy,yogo,dagan}@cs.biu.ac.il

July 22, 2015

Errata

This errata clarifies the issue of how out-of-vocabulary (OOV) words were treated during evaluation, overriding any information provided in the original publication.

The accompanying software package, `hyperwords`, represents OOV words as zero-vectors. This evaluation scheme was used for $win = 10$. However, the experiments for $win = 2$ and $win = 5$ were run using an older version of our software. The deprecated version does not represent OOV words at all, and omits them from the evaluation, essentially reducing the number of test instances. In principle, this evaluation strategy is misleading, since it gives an advantage to representations with smaller vocabularies by filtering out the harder test instances – rare words.

Fortunately, this bug had very little effect in practice; the main shift in performance occurs when evaluating on Luong et al.'s rare-word dataset. More importantly, the claims made in the paper still hold, since every algorithm was given access to every vocabulary during training. In addition, the window-size parameter (win) was controlled, and results from different window sizes were not directly compared.

Acknowledgements

We would like to thank Guillaume Pitel for finding this error and notifying us.