Corrigendum to “Improving Negative Sampling for Word Representation using Self-embedded Features” by Long Chen et al, in the Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, Pages 99-107

In the article, Improving Negative Sampling for Word Representation using Self-embedded Features by Long Chen et al, published in the Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, Pages 99-107, authors would like to add the following clarification in section 4.

As mentioned in Section 2, negative sampling by approximating maximum-inner-product function has been investigated in other research fields, e.g., [33] for the recommender system domain. Moreover, by understanding the similarity of the sampling motivation in the field of word embedding to that in the recommender system, intuitively, we could adapt the sampler developed for recommender system to solve our problem. Hence, for self-contained purpose, we would use similar notations (but with totally different meanings) in [33] to present in detail how to perform ranking-invariant normalization, rank mixture and sampling from rank mixture. In fact, the feasible sampling methods are not limited to the one introduced below, other samplers having a similar intuition (e.g., [46]) may also have an effect on improving the original popularity-based sampler. To our best knowledge, this is the first research work that introduces a dynamic negative sampling method for the field of word embedding.

(See DOI: <https://doi.org/10.1145/3159652.3159695>)