Limits of Deep Learning

Simon Jenni
University of Bern
Email: simujenni@students.unibe.ch

Simon Brulhart
University of Fribourg
Email: simon.brulhart@unifr.ch

***Abstract*—The abstract goes here.**

I. Introduction

Recent state of the art natural language models represent words by embedding them as vectors in a continuous vector space [1] [2]. These embeddings are learned as weights of a recurrent neural network language model. It has been demonstrated that the distributed representation of words as vectors in a common vector-space captures not just syntactic similarities (e.g. cat is close to $cats$) but semantic similarities as well. Concretely, the model's ability to perform word-arithmetic has been demonstrated [3]. To exemplify this, let f be the word-embedding function. Ideally, we would then have an embedding where distances in vector-space give a measure of semantic-similarity between the words. For two word pairs such as $(king, man)$ and $(queen, woman)$ which have the same relation (or similarity) we would obtain $f(king) - f(man) = f(queen) - f(woman)$ or equivalently $f(king) - f(man) + f(woman) = f(queen)$.

The goal of this work is to further evaluate the performance of the popular word2vec tool on these tasks. word2vec is an implementation of the Continuous Bag of Words (CBOW) and Skip-Gram models introduced in [1]. Both of these models are based on shallow two-layer neural networks with architectures depicted in Figure 1 and Figure 2. Both architectures take 1-hot-encoded words as inputs which get then mapped into into an N-dimensional vector via the weight matrix W. The hidden unit then accumulates the sum of these vectors for all the input words.

A. Subsection Heading Here

Subsection text here.

1) Subsubsection Heading Here: Subsubsection text here.

II. Conclusion

The conclusion goes here.

Acknowledgment

The authors would like to thank...

References

[1] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[2] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation." in *EMNLP*, vol. 14, 2014, pp. 1532–1543.

[3] T. Mikolov, W.-t. Yih, and G. Zweig, "Linguistic regularities in continuous space word representations." in *HLT-NAACL*, 2013, pp. 746–751.

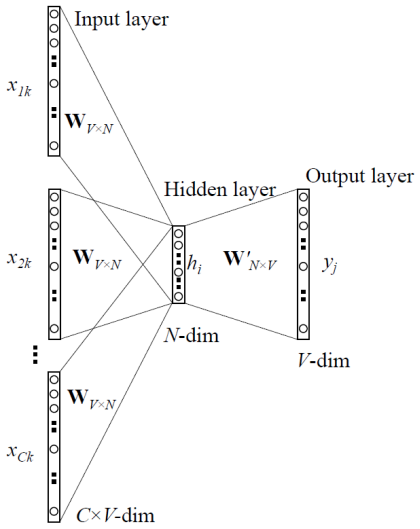

Fig. 1: Architecture of the CBOW model.

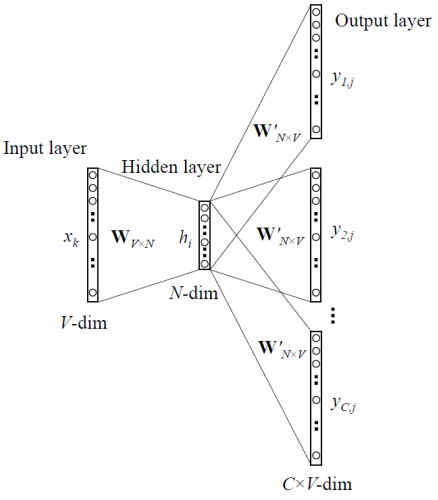

Fig. 2: Architecture of the Skip-Gram model.