

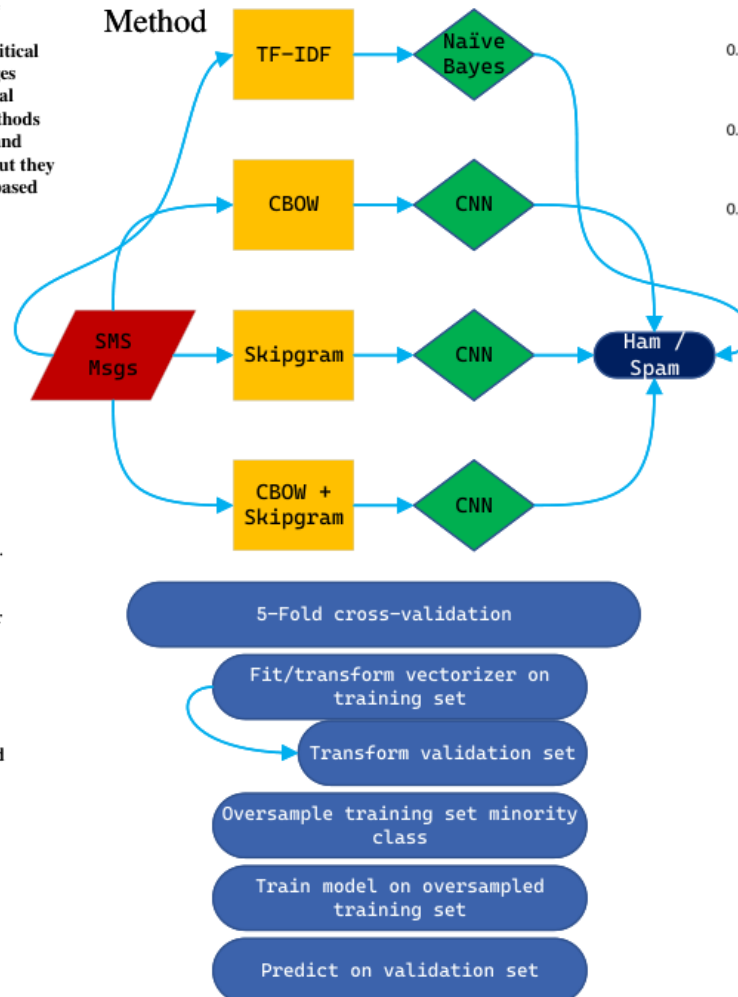
Classifying Text Messages as Spam Using Convolutional Neural Networks

Elliot Farmer Garcia
Clark Nelson

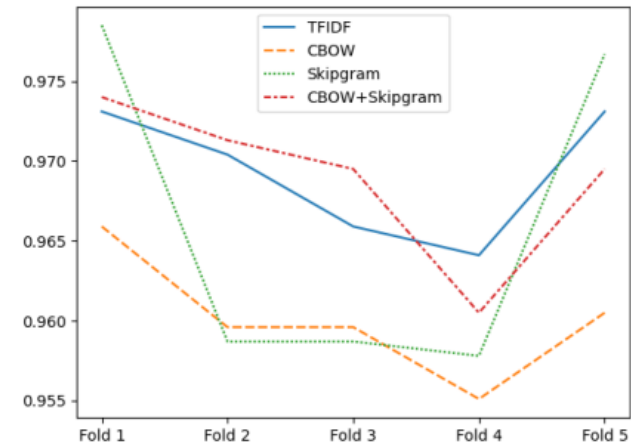
Department of Computer Science
University of Houston

Abstract—Natural language processing (NLP), a branch of machine learning dedicated to programmatically interpreting voice and text data, has developed into a very rich and mature discipline, with several viable techniques being applied to various problems in the NLP space. This study compares the efficacy of two leading NLP feature extraction methods, continuous bag-of-words (CBOW) and skipgram, by feeding their vectorized results into a Convolutional Neural Network (CNN). These techniques are applied to classifying SMS text messages as either “spam” or “ham” (i.e., not spam), an increasingly critical domain — by June 2022, over 66 billion SMS spam messages were received and caused an estimated \$9 billion in financial losses. The combination of these two feature extraction methods performs better (.9641 accuracy) than individually (.9585 and .9632 accuracies for CBOW and skipgram, respectively), but they are unable to statistically outperform a more naive count-based approach (.9693 accuracy).

Method



Results



Background

Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition," in *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, Nov. 1998, doi: 10.1109/5.726791.

T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," arXiv:1301.3781.

G.V. Cormack, J.M.G. Hidalgo, and E.P. Sáenz, "Feature engineering for mobile (SMS) spam filtering," in *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '07)*. Association for Computing Machinery, New York, NY, USA, pp. 871-872, 2007, doi: 10.1145/1277741.1277951

Conclusion

- All word vectorization methods resulted in highly classifiable messages
- The most effective and fastest method was the combination of TF-IDF and Naïve Bayes
- Combining CBOW and Skipgram marginally improved results

Recommendations

- FastText
- BERT
- RoBERTa

Acknowledgements

- Amaury Lendasse
- Peggy Lindner
- Nouhad Rizk
- Ed Ratner