Pre-trained Data Augmentation for Text Classification

Brazilian Conference on Intelligent Systems
BRACIS 2020: Intelligent Systems pp 551-565 | Cite as

Conference paper
First Online: 13 October 2020

Part of the Lecture Notes in Computer Science book series (LNCS, volume 12319)

Abstract

Data augmentation is a widely adopted method for improving model performance in image classification tasks. Although it still not as ubiquitous in Natural Language Processing (NLP) community, some methods have already been proposed to increase the amount of training data using simple text transformations or text generation through language models. However, recent text classification tasks need to deal with domains characterized by a small amount of text and informal writing, e.g., Online Social Networks content, reducing the capabilities of current methods. Facing these challenges by taking advantage of the pre-trained language models, low computational resource consumption, and model compression, we proposed the *PRE-trained Data AugmenTOR* (PREDATOR) method. Our data augmentation method is composed of two modules: the Generator, which synthesizes new samples grounded on a lightweight model, and the Filter, that selects only the high-quality ones. The experiments comparing Bidirectional Encoder Representations from Transformer (BERT), Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) and Multinomial Naive Bayes (NB) in three datasets exposed the effective improvement of accuracy. It was obtained 28.5% of accuracy improvement with LSTM on the best

scenario and an average improvement of 8% across all scenarios. PREDATOR was able to augment real-world social media datasets and other domains, overcoming the recent text augmentation techniques.

Keywords

Data augmentation Text classification Online social networks
The authors would like to thank the financial support of the National Council for Scientific and Technological Development (CNPq) of Brazil - Grant of Project 420562/2018-4 - and Fundação Araucária.

References

1. 1.
   Anaby-Tavor, A., et al.: Not enough data deep learning to the rescue. arXiv preprint arXiv:1911.03118 (2019)
2. 2.
   Bengio, Y., Ducharme, R., Vincent, P., Jauvin, C.: A neural probabilistic language model. J. Mach. Learn. Res. **3**(2), 1137–1155 (2003)zbMATHGoogle Scholar
3. 3.
   Cubuk, E.D., Zoph, B., Mane, D., Vasudevan, V., Le, Q.V.: Autoaugment: learning augmentation policies from data (2019).https://arxiv.org/pdf/1805.09501.pdf
4. 4.
   Demšar, J.: Statistical comparisons of classifiers over multiple data sets. J. Mach. Learn. Res. **7**, 1–30 (2006)MathSciNetzbMATHGoogle Scholar
5. 5.
   Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 1 (Long and Short Papers), pp. 4171–4186. Association for Computational Linguistics, Minneapolis, Minnesota,

June 2019. https://doi.org/10.18653/v1/N19-1423, https://www.aclweb.org/anthology/N19-1423

6. 6.

Edunov, S., Ott, M., Auli, M., Grangier, D.: Understanding back-translation at scale. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 489–500. Association for Computational Linguistics, Brussels, Belgium, Oct-Nov 2018. https://doi.org/10.18653/v1/D18-1045, https://www.aclweb.org/anthology/D18-1045

7. 7.

Fan, A., Lewis, M., Dauphin, Y.: Hierarchical neural story generation. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 889–898. Association for Computational Linguistics, Melbourne, Australia, July 2018. https://doi.org/10.18653/v1/P18-1082, https://www.aclweb.org/anthology/P18-1082

8. 8.

Gokaslan, A., Cohen, V.: Openwebtext corpus (2019). http://Skylion007.github.io/OpenWebTextCorpus

9. 9.

Hinton, G., Vinyals, O., Dean, J.: Distilling the knowledge in a neural network. In: NIPS Deep Learning and Representation Learning Workshop (2015). http://arxiv.org/abs/1503.02531

10. 10.

Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997). https://doi.org/10.1162/neco.1997.9.8.1735CrossRefGoogle Scholar

11. 11.

Holtzman, A., Buys, J., Du, L., Forbes, M., Choi, Y.: The curious case of neural text degeneration. arXiv preprint arXiv:1904.09751 (2019)

12. 12.

Howard, J., Ruder, S.: Universal language model fine-tuning for text classification. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 328–339. Association for Computational Linguistics, Melbourne, Australia, July 2018. https://doi.org/10.18653/v1/P18-1031, https://www.aclweb.org/anthology/P18-1031

13. 13.

Hu, Z., Tan, B., Salakhutdinov, R.R., Mitchell, T.M., Xing, E.P.: Learning data manipulation for augmentation and weighting. In: Wallach, H., Larochelle, H., Beygelzimer, A., dÁlché-Buc, F., Fox, E., Garnett, R. (eds.) Advances in Neural Information Processing Systems, vol. 32, pp. 15764–15775. Curran Associates, Inc. (2019). http://papers.nips.cc/paper/9706-learning-data-manipulation-for-augmentation-and-weighting.pdf

14. 14.

Igawa, R.A., et al.: Account classification in online social networks with lbca and wavelets. Inform. Sci. **332**, 72–83 (2016)CrossRefGoogle Scholar

15. 15.

Kim, Y.: Convolutional neural networks for sentence classification. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1746–1751. Association for Computational Linguistics, Doha, Qatar, October 2014. https://doi.org/10.3115/v1/D14-1181, https://www.aclweb.org/anthology/D14-1181

16. 16.

Kobayashi, S.: Contextual augmentation: data augmentation by words with paradigmatic relations. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 2 (Short Papers), pp. 452–457. Association for Computational Linguistics, New Orleans, Louisiana, June 2018. https://doi.org/10.18653/v1/N18-2072, https://www.aclweb.org/anthology/N18-2072

17. 17.

Kolomiyets, O., Bethard, S., Moens, M.F.: Model-portability experiments for textual temporal analysis. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers, Vol. 2, pp. 271–276. HLT '11, Association for Computational Linguistics, USA (2011)Google Scholar

18. 18.

Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)Google Scholar

19. 19.

Kumar, V., Choudhary, A., Cho, E.: Data augmentation using pre-trained transformer models. arXiv preprint arXiv:2003.02245 (2020)

20. 20.
Petroni, F., et al.: Language models as knowledge bases. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pp. 2463–2473. Association for Computational Linguistics, Hong Kong, China, November 2019. https://doi.org/10.18653/v1/D19-1250, https://www.aclweb.org/anthology/D19-1250

21. 21.
Sanh, V., Debut, L., Chaumond, J., Wolf, T.: Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108 (2019)

22. 22.
Schwartz, R., Dodge, J., Smith, N.A., Etzioni, O.: Green ai. ArXiv abs/1907.10597 (2019)Google Scholar

23. 23.
Sennrich, R., Haddow, B., Birch, A.: Improving neural machine translation models with monolingual data. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 86–96. Association for Computational Linguistics, Berlin, Germany, Auguest 2016. https://doi.org/10.18653/v1/P16-1009, https://www.aclweb.org/anthology/P16-1009

24. 24.
Shleifer, S.: Low resource text classification with ulmfit and backtranslation. arXiv preprint arXiv:1903.09244 (2019)

25. 25.
Socher, R., et al.: Recursive deep models for semantic compositionality over a sentiment treebank. In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pp. 1631–1642 (2013)Google Scholar

26. 26.
Strubell, E., Ganesh, A., McCallum, A.: Energy and policy considerations for deep learning in NLP. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 3645–3650. Association for Computational Linguistics, Florence,

Italy, July 2019. https://doi.org/10.18653/v1/P19-1355, https://www.aclweb.org/anthology/P19-1355

27. 27.

Wang, S., Manning, C.D.: Baselines and bigrams: simple, good sentiment and topic classification. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers, vol. 2, pp. 90–94. Association for Computational Linguistics (2012)Google Scholar

28. 28.

Wei, J., Zou, K.: EDA: easy data augmentation techniques for boosting performance on text classification tasks. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP-IJCNLP), pp. 6382–6388. Association for Computational Linguistics, Hong Kong, China, November 2019. https://doi.org/10.18653/v1/D19-1670, https://www.aclweb.org/anthology/D19-1670

29. 29.

Wolf, T., et al.: Huggingface's transformers: state-of-the-art natural language processing. ArXiv abs/1910.03771 (2019)Google Scholar

30. 30.

Wong, S.C., Gatt, A., Stamatescu, V., McDonnell, M.D.: Understanding data augmentation for classification: when to warp. In: 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA), pp. 1–6. IEEE (2016)Google Scholar

31. 31.

Yogatama, D., Dyer, C., Ling, W., Blunsom, P.: Generative and discriminative text classification with recurrent neural networks. arXiv preprint arXiv:1703.01898 (2017)

32. 32.

Yu, A.W., Dohan, D., Luong, T., Zhao, R., Chen, K., Le, Q.: Qanet: combining local convolution with global self-attention for reading comprehension (2018). https://openreview.net/pdf?id=B14TlG-RW

33. 33.

Zhang, X., Zhao, J., LeCun, Y.: Character-level convolutional networks for text classification. In: Proceedings of the 28th International Conference on Neural Information Processing Systems, Vol. 1, pp. 649–657. NIPS'15, MIT Press, Cambridge, MA, USA (2015)Google Scholar