can manifest themselves as out-of-context words such as 'well' and 'continent' in the given example or as a grammatical mistake, where 'enjoy' becomes 'enjoys'.
Table 8 provides a few examples of text corruption and subsequent corrections for qualitative comparison. The samples were picked randomly from the medium error variant of the Amazon dataset. [4] We can see that the error correction model is quite accurate at simple corrections like 'wive herb fingers' to 'with her fingers' in example 1, and 'i booth hits' to 'i bought this' in example 2. It is also able to accurately predict complex correction requiring more context, e.g., 'malnutrition' is correctly predicted to be 'nutritional' given the positive sentiment of the review. In a more ambiguous situation where the exact word is harder to predict, the model still makes a sensible prediction. For instance, 'bomb' is corrected to 'combo' in example 3, and 'isle' to the verb 'ship' in example 2, which, albeit not correct, is a sensible prediction in the context, similar in meaning to the ground truth word 'send'.. However, it often fails to correct obvious errors, especially when the out-of-context word is a fairly common one. For example, 'hot of style' in example 1 and 'gorgeous and screen' in example 3 are left uncorrected. This model is entirely independent of the error detection model. A two-component design feeding the prediction of the detection model into the correction model may improve the results.

7. Conclusion

This paper proposes a method to induce typographical errors based on realistic error modeling, which is used to induce two novel typographic error datasets from different domains, each generated at three different error levels. We show that BiLSTM neural networks are fairly effective in detecting these errors and that Transformer networks show potential in correcting them. Our data is freely available online at `http://typo.nlproc.org` for the community to make further progress on this challenging task.

8. References

Aramaki, E. (2010). Typo corpus. `http://luululu.com/tweet/`.

Baba, Y. and Suzuki, H. (2012). How are spelling errors generated and corrected?: a study of corrected and uncorrected spelling errors using keystroke logs. In *Proceedings of ACL 2012*, pages 373–377.

Banko, M. and Brill, E. (2001). Scaling to very very large corpora for natural language disambiguation. In *Proceedings of ACL 2001*, pages 26–33.

Bigert, J., Ericson, L., and Solis, A. (2003). Autoeval and missplel: Two generic tools for automatic evaluation. *Proceedings of NoDaLiDa, Reykjavik, Iceland*, 3.

Carlson, A. and Fette, I. (2007). Memory-based context-sensitive spelling correction at web scale. In *Sixth international conference on Machine Learning and Applications 2007 (ICMLA 2007)*, pages 166–171. IEEE.

Carlson, A. J., Rosen, J., and Roth, D. (2001). Scaling up context-sensitive text correction. *Urbana*, 51:61801.

[4] A constraint was placed to not select sentences longer than 30 words due to space limitations.

Dale, R. and Kilgarriff, A. (2011). Helping our own: The hoo 2011 pilot shared task. In *Proceedings of the 13th European Workshop on Natural Language Generation*.

Felice, M. and Yuan, Z. (2014). Generating artificial errors for grammatical error correction. In *Proceedings of the Student Research Workshop at EACL*, pages 116–126.

Foster, J. and Andersen, Ø. E. (2009). Generrate: generating errors for use in grammatical error detection. In *Proceedings of the 4th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 82–90. Association for Computational Linguistics.

Ghosh, S. and Kristensson, P. O. (2017). Neural networks for text correction and completion in keyboard decoding. *arXiv preprint arXiv:1709.06429*.

Golding, A. R. and Roth, D. (1999). A winnow-based approach to context-sensitive spelling correction. *Machine learning*, 34(1-3):107–130.

Kasewa, S., Stenetorp, P., and Riedel, S. (2018). Wronging a right: Generating better errors to improve grammatical error detection. In *Proceedings of EMNLP 2018*.

Lachowicz, D. (2018). Enchant. `https://abiword.github.io/enchant/`.

Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., and Potts, C. (2011). Learning word vectors for sentiment analysis. In *Proceedings of ACL 2011*.

McAuley, J. J. and Leskovec, J. (2013). From amateurs to connoisseurs: modeling the evolution of user expertise through online reviews. In *Proceedings of the 22nd international Conference on World Wide Web*. ACM.

Ng, H. T., Wu, S. M., Wu, Y., Hadiwinoto, C., and Tetreault, J. (2013). The CoNLL-2013 Shared Task on Grammatical Error Correction. In *Proceedings of the Seventeenth Conference on Computational Natural Language Learning: Shared Task*, pages 1–12.

Ng, H. T., Wu, S. M., Briscoe, T., Hadiwinoto, C., Susanto, R. H., and Bryant, C. (2014). The CoNLL-2014 Shared Task on Grammatical Error Correction. In *Proceedings of the Eighteenth Conference on Computational Natural Language Learning: Shared Task*, pages 1–14.

Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). Bleu: a method for automatic evaluation of machine translation. In *Proceedings of ACL 2002*, pages 311–318.

Pennington, J., Socher, R., and Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of EMNLP 2014*, pages 1532–1543.

Qiu, S., Xu, B., Zhang, J., Wang, Y., Shen, X., de Melo, G., Long, C., and Li, X. (2020). EasyAug: An automatic textual data augmentation platform for classification tasks. In *Proceedings of The Web Conference 2020*. ACM.

Rei, M., Felice, M., Yuan, Z., and Briscoe, T. (2017). Artificial error generation with machine translation and syntactic patterns. *arXiv:1707.05236*.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. *arXiv:1706.03762*.

Whitelaw, C., Hutchinson, B., Chung, G. Y., and Ellis, G. (2009). Using the web for language independent spellchecking and autocorrection. In *Proceedings of EMNLP 2009*, pages 890–899.