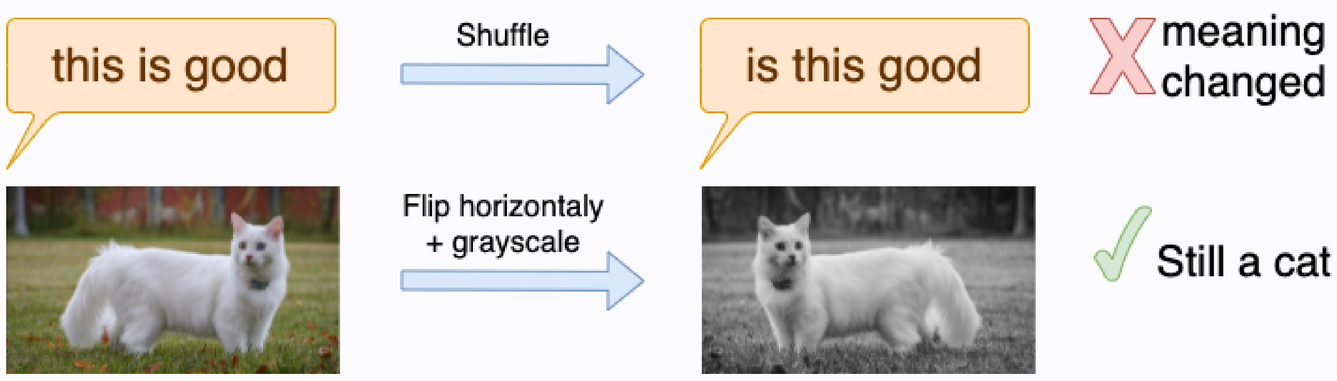
Philipps Universität Marburg

Size matters!
Data augmentation for stance detection

Master/Bachelor Thesis

Motivation

Data augmentation techniques are used to generate additional, synthetic data using already existing data. While images can be augmented easily by transformations such as rotations or changes of the RGB channel without affecting the classification models, text augmentation is much more difficult because there are no universal rules for automatic textual data transformations that can be applied while maintaining the meaning of the text. Therefore this thesis aims to investigate various natural language generation techniques for paraphrasing opinionated text and evaluate the quality of the generated data on opinion detection.



Difficulty

Analysis

Programming

Literature

Task Description

- Researching different textual data augmentation methods
- Performing comparative analysis of augmentation techniques with respect to their performance on stance detection
- Evaluating the best performing data augmentation methodology on other relevant tasks such as sentiment analysis and fake news detection
- Performing error analysis

Contact

Prof. Dr. Lucie Flek &
Dr. Flora Sakketou
f.sakketou@gmail.com
@CAISAlab

References

[1] Markus Bayer, Marc-André Kaufhold, and Christian Reuter. A survey on data augmentation for text classification. *CoRR*, abs/2107.03158, 2021.

[2] Thien Ho Huong and Vinh Truong Hoang. A data augmentation technique based on text for vietnamese sentiment analysis. In *Proceedings of the 11th International Conference on Advances in Information Technology*, IAIT2020, New York, NY, USA, 2020. Association for Computing Machinery.

[3] Tomas Liesting, Flavius Frasincar, and Maria Mihaela Trusca. Data augmentation in a hybrid approach for aspect-based sentiment analysis. *CoRR*, abs/2103.15912, 2021.