UBIQUITOUS KNOWLEDGE PROCESSING

Bachelor/Master Thesis

TECHNISCHE UNIVERSITÄT DARMSTADT

Improving the Robustness of NLP Models

Motivation

The majority of NLP datasets contain spurious patterns that are associated with the target label and are easy to learn. Models tend to focus on learning these dataset-specific spurious patterns instead of learning more generalizable patterns to solve the underlying task. As a result, while models that are trained on such datasets achieve high performances on the same data distribution, they fail on out-of-domain data distributions. In this regard, we explore innovative approaches to improve the robustness of NLP models across various datasets, tasks, and data distributions.

Possible Tasks

- Analysing existing NLP datasets for their potential artifacts
- Novel techniques to improve robustness and zero-shot performances of NLP models

References

- Prasetya Ajie Utama, Nafise Sadat Moosavi, Iryna Gurevych. 2020. “Mind the Trade-off: Debiasing NLU Models without Degrading the In-distribution Performance”. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)
- Prasetya Ajie Utama, Nafise Sadat Moosavi, Iryna Gurevych. 2020. Towards Debiasing NLU Models from Unknown Biases. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)

Analysis	■■■■□
Programming	■■■□□
Literature	■■■□□

Contact

Prof. Dr. Iryna Gurevych

Dr. Nafise Sadat Moosavi

thesis@ukp.informatik.tu-darmstadt.de