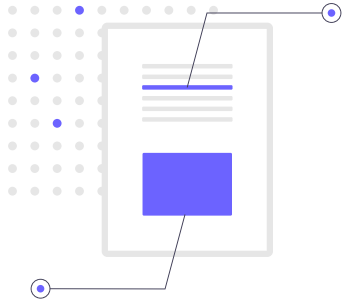
PD03 Plagiarism detection with a multi-task perspective

Background

The recent success in NLP can be attributed to self-supervised learning on massive text corpora. Through self-supervision, language models learn a broad set of skills and pattern recognition abilities. Also in plagiarism detection, language models have experienced great success. However, many of the required skills to solve a single task (e.g., author identification) are also present in related tasks (e.g., originality) for which labeled data exists. Therefore, a recent trend in improving language models prior to fine-tuning has become multi-task learning which leverages labeled training data to learn many skills simultaneously. As plagiarism has different forms, such as paraphrasing, idea plagiarism, author similarity, it is an intuitive candidate to perform multi-task learning



Goal

- Explore multi-task learning for plagiarism with neural language models

Tasks

- Train detection models based on previous state-of-the-art work
- Propose training architectures and paradigms.
- Evaluate with human studies and automated metrics

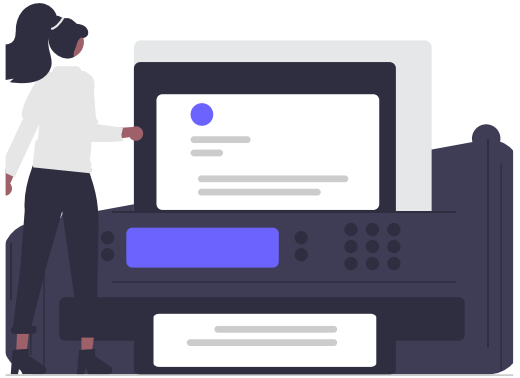
Jan Philip Wahle
wahle@gipplab.org

Terry L. Ruas
ruas@gipplab.org

PD04 Generating paraphrased plagiarism

Background

To identify neural-paraphrased plagiarism, we need data to learn which features make paraphrased examples so convincing. We ideally seek for automated solutions as they are scalable. With data paraphrased by multiple techniques, we can optimize detection methods that are robust and generalize well to unseen scenarios. We further assume generative language models paraphrase similar to humans. If we can confirm this hypothesis, and generate training data automatically, we can increase the performance of detection methods without the tedious process of finding real-world plagiarized examples.



Goal

- Explore the generation of machine-paraphrased plagiarism with neural language models.

Tasks

- Use existing datasets and find their weak spots to extend them to be more robust.
- Propose paraphrasing methods and test them using human studies.
- Evaluate whether neural-paraphrasing is close to how humans paraphrase text.

Jan Philip Wahle

wahle@gipplab.org

Terry L. Ruas

ruas@gipplab.org

NII

PD05 Identifying neural-paraphrased plagiarism

Background

Machine-paraphrasing has become a concerning problem for research institutions, publishers, and schools, as anyone can obtain access to free tools that generate convincing plagiarism. Large auto-regressive language models with more than a hundred billion parameters, such as GPT-3, can generate text indistinguishable from human writing which makes plagiarism effortless, yet extremely difficult to spot. In the near future, when large language models become more accessible, the number of plagiarized texts increases dramatically. Therefore, we need automated plagiarism detection solutions now before models are widely misused for plagiarism

Goal

- Explore machine-paraphrased plagiarism with neural language models

Tasks

- Build detection models based on previous state-of-the-art work.
- Propose training architectures and paradigms.
- Evaluate with human studies and automated metrics.

Jan Philip Wahle

wahle@gipplab.org

Terry L. Ruas

ruas@gipplab.org

NII