Navigating Modern Text Classification: A Conceptual, Computational, and Empirical Guide to Pre-trained Large Language Models

Martin Juan José Bucher
Marco Martini

Online Appendix

Code, data, notebooks, and documentation:

All replication materials for our paper as well as our text classification toolkit including a detailed step-by-step guidance are available under this link.

Additional implementation details for the prompt-based approaches with BART and ChatGPT:

- For BART, we use the BART Large model `facebook/bart-large-mnli` introduced by Lewis et al. (2019) together with the Zero-Shot Classification pipeline from Huggingface. We use a technique called Natural Language Inference (NLI), which prompts a model using two sentences, a *Premise* (in our case the text to be classified) and a *Hypothesis* (a possible class label for that text). The model then predicts if the hypothesis is consistent with the premise. NLI evaluates a hypothesis for each label and then selects the label with the highest confidence as output.

- For ChatGPT, we use either `gpt-3.5-turbo` or `gpt-4 -1106-preview` via OpenAI's API (https://chat.openai.com). We prompt the model as follows: The first part of the prompt describes the text classification task. This is followed by the text to be classified and the list of permissible labels. We prompt the model to only output one of the given labels and set *temperature=0.1* to induce some variance in the generated output. For some samples, the model would output a result that is not in the label list. In these cases, we rerun the prompt until the model outputs a permissible label. The prompts for ChatGPT for all datasets are provided in the supplementary materials.

References:

- Lewis, Mike, et al. "Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension." arXiv preprint arXiv:1910.13461 (2019).