Software Project: Adapting Large Language Models to Human Feedback (w/ and w/o Reinforcement Learning)

Stefan Riezler

SoSe 2023
Large Language Models (LLMs)

- chatGPT and friends - Revolution not only in societal impact, but also paradigm shift in research
  - Pre-trained large language models learn from few-shot demonstrations, specified via text interactions with the model (GPT-3 [Brown et al., 2020])
  - Additional fine-tuning by reinforcement learning from human feedback (chatGPT [Christiano et al., 2017, Kreutzer et al., 2018, Ouyang et al., 2022])
Reinforcement Learning from Human Feedback via Text Interactions

- Human feedback is input as text sequence, fine-tuning only by few-shot prompting / in-context learning
  [Liu et al., 2023, Madaan et al., 2023]
Project Idea

- Choose a text generation task
  - Example: Machine translation by prompting LLMs (Bloom\(^1\), mT5\(^2\)) [Vilar et al., 2022]
- Collect feedback on model output
  - Example: Human feedback on machine translations [Kreutzer et al., 2018], or simulated by metrics like COMET\(^3\)
- Fine-tune model on textual feedback to model outputs
  - Option 1: Fine-tune with RLHF [Ouyang et al., 2022]\(^4\)
  - Option 2: Fine-tune with Cross-Entropy [Liu et al., 2023]\(^5\)
  - Option 3: Fine-tune with few-shot prompting/in-context learning (apply [Madaan et al., 2023] to human feedback)

1 https://huggingface.co/docs/transformers/model_doc/bloom
2 https://huggingface.co/docs/transformers/model_doc/mt5
3 https://huggingface.co/Unbabel/wmt22-comet-da
4 https://github.com/CarperAI/trlx
5 https://huggingface.co/blog/peft
Project Goal

- Learn to work with pre-trained large language models
- Provide small amounts of feedback yourself
- Understand how large language models learn from your feedback
References


Self-refine: Iterative refinement with self-feedback.


Training language models to follow instructions with human feedback.
In Advances in Neural Information Processing Systems (NeurIPS), New Orleans, Louisiana, USA.
