
Natural Language Processing – Assignment #3

Assignment description

This assignment deals with the similarities and the differences between human- and models-generated language, and it will be implemented in the attached notebook. Contemporary large language models (LLMs) show astonishing capabilities in generating language that resembles the language of humans to the extent that makes it difficult (almost impossible) to distinguish between the two.

Considering different NLP tasks we have studied so far, we would like to apply several metrics for measuring the differences between the two. A large dataset of human replies to various [Reddit](#) posts is attached to this assignment, as well as replies by several LLMs to the same set of questions:

human_written_comments_reddit.csv – posts and comments by human Redditors.

generated_comments_<model-name>_<temperature>.csv – comments generated by a model with a specific temperature setting (0.7), in reply to roughly the same set of (human-written) posts. We are using two open (Llama, Qwen) and two closed (GPT4.1-mini, Claude-Sonnet) LLMs, with the columns:

- post_id – the unique ID of the post
- post_title – the title of the post (human-written)
- post_text – the body of the post (human-written)
- comment – written by humans or generated by models

Comment: Occasionally the data is noisy and improperly encoded (happens due to OS and charset differences). Use “utf-8” encoding for reading – that will help you overcome the issues.

Get familiar with the data – read through posts and comments; that will help you to come up with ideas about measures you would like to apply. Implement a few metrics that potentially show (similarities or) differences between the two types of comments and report your final results.

Consider at least three metrics: (1) mean sentence length, (2) type to token ratio (TTR, think if the text length affects the metric) and (3) mean parse tree depth. We strongly encourage you to think about additional measurements along the lines of what we have seen in the lectures. We can use external libraries for this task, when in doubt – ask on the forum.

Report your metrics for each comment type – human-generated or generated by a model.

Keep your code readable – implement each metric in a separate cell and report your final results towards the end. Clean and easily readable solutions will gain higher scores.

Write down your impressions and conclusions in a pdf (about one page).

P.S. If willing (and running out of reading materials :-), you may want to read through [this recent study](#) by our graduate student. Among others, the paper describes the dataset collection.

Submission

Before submission: restart kernel and make sure the whole notebook runs smoothly.

Submit a single zip file – assignment3_XXXXXXXX_XXXXXXXX.zip , where “XXXXXXXX” stands for a student id. Please specify two student ids (your and your partner’s). It should include two files:

1. Your implementation for the task -- human-and-models-language.ipynb.
2. A document (pdf) with your impressions and conclusions.

Grading criteria include: correctness, creativity, code design, readability and documentation.

Good Luck!