CS288 Natural Language Processing
Songwen Zhao
SID: 3038663175

Due on 04/05/23 11:59 pm PST

Report - Homework #4

1 Finetuning Large Language Models

For the project report, find two sentences where the model succeeds and two sentences where the model fails. Describe what might be causing these types of failures.

- German: Das Buch ist wirklich interessant. Prediction English: The butterfly is amazing. Target English: The book is really interesting. It fails since there are too many substantive words.
- German: Der Stift ist hier. Prediction English: The shift is happening. Target English: The pen is here. It fails since German word "Stift" is very similar to English word "shift".
- German: Das ist ein Computer. Prediction English: It's a computer. Target English: It's a computer. It works.
- German: Der Mann ist im Wasser. Prediction English: The man is the ocean. Target English: The man is in the water. It almost works as the meanings of "ocean" and "water" are not much different.

For the project report, compare two translations from the GPT-2 versus LSTM model. Which one works better? LSTM works better as it's trained specifically on Multi30k. It predicts more similar translation but less fluent and with less professional words.

- German: Ein Mann schl¨aft in einem gru¨nen Raum auf einem Sofa. Target: A man sleeping in a green room on a couch. GPT-2: A man in a wheelchair in a wheelchair-accessible house. LSTM: A man is asleep in a green room with a little girl.
- German: Eine Gruppe von Männern lädt Baumwolle auf einen Lastwagen. Target: A group of men are loading cotton onto a truck. GPT-2: A group of men are repairing the engine of a truck. LSTM: A group of men are picking out of a tree.

Describe two possible improvements to your sentiment classifier. One is to use a larger or a more task-specific dataset for fine-tuning. We can also try fine-tuning other pre-trained models specifically designed for sentiment analysis tasks, such as BERT or RoBERTa.