Prompt Engineering Lab

21/03/2024

Instructions

In this lab you will be working with prompts and apply the learnings from the presentation. Prompt engineering techniques will be used to analyse data for both of the cases below and generate insights. The deliverable for this lab will be the final presentation with the result of your analysis, python notebook, and the document with screenshot of the different prompts and their corresponding output. The details on how the presentation will be graded and it's content will be discussed in the coming weeks.

1. Dataset

For this lab, we will use the ***'amazon_data'*** and the ***'amazon_alexa_review'*** datasets which can be downloaded from canvas. Only the **top 50 rows** for both of the datasets should be used for analysis in order to reduce the computation time of the large GPT model.

2. Model

The model you will be working with is the GPT2 model from Hugging face. **Optional**: To use different GPT models (GPT2 vs GPT2-XL). You can use the following code in your google collab notebooks to start with the analysis. You will need to optimize the code to make it suitable to the dataset and case.

```
prompt = """
Q: What is the capital of France?
A: Paris


Q: What is the capital of England?
A: London

Q: What is the capital of Japan?""


generator = pipeline('text-generation', model='gpt2')
generator(prompt, max_length=70)
print(model_output)
```

When working with the model parameters (*temperature, top_p, max_length*) document the different values of the parameters you have worked with in your notebook. Include it in the document with a screenshot.

3. Prompt

While prompting with few shots' technique, you can include examples through analyzing the top 10 or 20 rows of the dataset to reduce the processing time. You can choose one of the cases mentioned below and perform prompt engineering for the questions. Additionally, in your group you need to come up with about 2 - 4 analysis questions that the business case can use to improve. You can use the dataset created by prompting.

While you're experimenting with different prompts and their generated text, store the outputs generated in a document with a screenshot of the output, input prompt and the parameters. Make sure to document your notebook well and analysis derived from your data. Enjoy!

Cases

Case 1 : Customer Feedback Analysis

TechTrendz is a leading provider of electronic gadgets. TechTrendz is receiving a large volume of customer feedback through various channels like emails, social media, and online reviews. The company wants to leverage this feedback to improve its products and services. For this, they use the dataset of Amazon Alexa review. The reviews are stored in the ***verified_reviews*** column.

Questions:

1. Using zero and few shot prompt engineering, how would you formulate a prompt to identify common themes (e.g., aspect of the product) mentioned in customer feedback?
2. With zero and few shot prompt engineering, how would you construct prompts to classify feedback into positive, negative, or neutral sentiments?
3. Apply chain of thought prompt for extracting specific suggestions or feature requests from the feedback data. If no feature requests are mentioned include 'no request mentioned'.

In your notebook, create a section in which you enter the review and get answer for each of the customer review from the dataset. You can include visualizations (e.g., bar chart, pie chart) to illustrate the insights and trends extracted from the feedback.

Case 2: Information Extraction

StyleSavvy is a fashion retail company facing challenges in accurately analyzing their current sales data for its historical and upcoming collections. The company wants to leverage prompt engineering techniques to improve the accuracy of its sales forecasts. They are using the dataset of consumer review of clothing product. The reviews are stored in the column ***Review*** in an unstructured manner. StyleSavvy wants to optimize the process of labeling their reviewed products collections so they can perform analysis. The dataset *data_amazon* can be downloaded from the canvas page.

Questions:

1. Using zero and few shot prompt engineering, how would you formulate prompts to extract the product name from the review? For comparing the results of the product name, you can use the ***cloth_class*** column.
2. Apply chain of thought prompting to generate prompts for identifying factors influencing sales fluctuations.
3. Analyze the dataset created through Q1 to find out which product has the highest/lowest profitability and the effect of ratings given.

Store the dataset created through Q1. In your notebook, create a section in which you enter a new review, and extract the product name. You can include visualizations (e.g., bar chart, pie chart) to illustrate the insights and trends extracted from the feedback.