Homework 4

Fine Tuning for Sentiment Analysis on Financial News

Please refer to the dataset and description of the problem at <https://www.kaggle.com/datasets/ankurzing/sentiment-analysis-for-financial-news/data>

I will reproduce key elements from there.

1. **Context**

This dataset (FinancialPhraseBank) contains the sentiments for financial news headlines from the perspective of a retail investor.

1. **Content**

The dataset contains two columns, "News Headline" and “Sentiment”. The sentiment can be negative, neutral or positive. The data is labeled by chosen annotators as explained below.

The objective of the phrase level annotation task was to classify each example sentence into a positive, negative or neutral category by considering only the information explicitly available in the given sentence. Since the study is focused only on financial and economic domains, the annotators were asked to consider the sentences from the viewpoint of an investor only; i.e. whether the news may have positive, negative or neutral influence on the stock price. As a result, sentences which have a sentiment that is not relevant from an economic or financial perspective are considered neutral.

2. Data

This release of the financial phrase bank covers a collection of 4840 sentences. The selected collection of phrases was annotated by 16 people with adequate background knowledge on financial markets. Three of the annotators were researchers and the remaining 13 annotators were master’s students at Aalto University School of Business with majors primarily in finance, accounting, and economics. To provide an objective comparison, we have formed 4 alternative reference datasets based on the strength of majority agreement:

(i) sentences with 100% agreement [file=Sentences\_AllAgree.txt],

(ii) sentences with more than 75% agreement [file=Sentences\_75Agree.txt],

(iii) sentences with more than 66% agreement [file=Sentences\_66Agree.txt], and

(iv) sentences with more than 50% agreement [file=Sentences\_50Agree.txt]

Examples:

1. Jan. 6 -- Ford is struggling in the face of slowing truck and SUV sales and a surfeit of up-to-date , gotta-have cars [.@negative](mailto:.@negative)
2. According to Gran , the company has no plans to move all production to Russia , although that is where the company is growing .@neutral
3. For the last quarter of 2010 , Componenta 's net sales doubled to EUR131m from EUR76m for the same period a year earlier , while it moved to a zero pre-tax profit from a pre-tax loss of EUR7m [.@positive](mailto:.@positive)

**Your Assignment**

1. First do 50 to 60 (equally distributed between neutral, positive, and negative), zero-shot prompting to determine accuracy either using an API (chat completion) or a chatbot. Get a baseline accuracy.
2. Then do 12-shot prompting, with 4 each for positive, negative, and neutral with the sentiment labels as input and then do an in-context testing on 50 to 60 news items for predicting the sentiment. This is probably easier done with prompt engineering. Technically, the accuracy should be better than zero-shot. But report the result as is.
3. In both of the cases above use the data files AllAgree.txt or 75Agree.txt. For this assignment we will not use 66Agree or the 50Agree data files at all.
4. Then fine-tune an LLM with 450 from AllAgree and 75Agree datasets with 150 news items for each sentiment. Then use the fine-tuned model to predict a total of 500 (not including the dataset for training). Here, chose the news items randomly (they do not have to be distributed equally).
5. Tabulate the results with a confusion matrix using True positive, True negative, False positive, and False negative for the fine-tuning experiment. Read Wikipedia or other sources for learning about the confusion matrix.
6. Lastly, go to finance.yahoo.com or other financial news websites, get 10 news items (perhaps longer than a sentence) on different companies and test them with the fine-tuned model. Then do a human level accuracy check and report the results.