Task1 - Research

November 2024

1   FinBERT

FinBERT is a pre-trained NLP model to analyze sentiment of financial text. It is built by further training the BERT language model in the finance domain, using a large financial corpus and thereby fine-tuning it for financial sentiment classification. Financial PhraseBank by Malo et al. (2014) is used for fine-tuning.

- **pre-processing**: there's no need to perform any pre-processing on the text before feeding it to the model

- **output**: the raw output from the model is represented as logits—numerical values indicating the model's "confidence" for each class. These logits are not direct probabilities but can be converted into probabilities using an activation function like softmax.
  We could calculate the expected value if +1 represents positive, -1 represents negative and 0 neutral

- **input**: the input is a dataframe containing the text, the csv file of speeches and documents will hopefully be given in different files.

```python
import pandas as pd
import scipy
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.metrics import accuracy_score,
    classification_report, confusion_matrix, roc_auc_score
import torch
from transformers import AutoTokenizer,
    AutoModelForSequenceClassification

tokenizer = AutoTokenizer.from_pretrained("ProsusAI/Finbert"
    )
model = AutoModelForSequenceClassification.from_pretrained("
    ProsusAI/Finbert")


```

```
13  def finbert_sentiment(text:str) -> tuple[float, float, float
        , str]
14      with torch.no_grad(
15              inputs = tokenizer(...)
16          ...
17          )
18  ...
```

2 Analysis

We want to measure the impact of central banks' speeches on the market prices of the followings:

- Euro/Dollar exchange rate;

- US treasury / German-Italian bonds futures (2y, 10y)

- ETF related to S&P500, Dow Jones, EURO STOXX, FTSE MIB

(Datas are available on yahoo finance and investing.com)

3 Proposals

1. Splitting the sentences into categories to have multidimensional outputs and better measuring individual impacts.

2. Measuring the speeches' unpredictability to have an information advantage.

3. Do we want to create a live instrument (minute x minute trading)?

References

[1] D Araci. "FinBERT: Financial Sentiment Analysis with Pre-trained Language Models". In: *arXiv preprint arXiv:1908.10063* (2019).

[2] Moritz Pfeifer and Vincent P Marohl. "CentralBankRoBERTa: A fine-tuned large language model for central bank communications". In: *The Journal of Finance and Data Science* 9 (2023), p. 100114.

[3] Maria Paola Priola, Piero Lorenzini, Giacomo Tizzanini, et al. "Measuring central banks' sentiment and its spillover effects with a network approach". In: (2021).