Research proposal – audio transcript

Transfer learning: live sentiment analysis for market prediction

Significance (slide 2)

There are few transformers-based models, which are state of the art for NLP deep learning, that have been tuned to understand financial jargon, most notably FinBERT, FinALBERT that I’ve listed in the reference list. These are very accurate models at interpreting sentiment of past data, they are less suitable for live incoming data due to their size and inference times. Furthermore, they need many resources to be trained and to run properly. There is no model in the financial sentiment analysis domain that is tuned for performance and speed and in some markets, such as the forex, a matter of seconds can cost vast amounts of money. The market environment is one where speed and efficiency matters the most due to its constant evolving nature, so a fast and reliable model that interprets data live could be extremely important (Mukul J. et al.,2021).

Another aspect that goes in favour of such approach is carbon footprint and energy consumption that are far less when using a smaller fine-tuned model instead of a larger one (Lacoste et al., 2019).

Question and Objective (slide 3/4)

The problem this research is asking is: is it possible to create a model that is fast enough, uses less resources and less memory to analyse incoming financial data and give tempestive and useful insights to investors? Big models could quickly become slow and costly to the point that they are not anymore worth using in the price prediction, factoring that the amount of data can become enormous, so we can see there is a trade-off between performance and accuracy.

The aim of the proposed research is to generate a state of the art, transformers-based model using transfer learning techniques, leveraging one of the smaller BERT publications, that I’ll mention in the next slide.

(Slide 5)

Many smaller and faster models have been developed starting from the revolutionary open-source transformers-based model BERT published by Google. Some of them achieve incredible performances up to ten times smaller and twenty times faster sacrificing a reasonable amount of accuracy, they suffer in harder NLP tasks but not notably in the sentiment analysis tasks which is an advantage (Xiaoqi et al., 2020). The other transformers-based models such as the GPT by OpenAI will not be used since they are far too large to be utilized without sufficient resources and the distilled versions still maintain a large number of parameters (109M for distilled GPT-2 against 14M for TinyBERT4) (Xiaoqi et al, 2020) (Huggingface, 2020).

In this slide we can see the effects of model distillation, a famous technique used to reduce the size of models where the bigger model (Teacher) is used as a reference for the smaller model (student) (Sanh et al., 2019). The smaller model outputs are compared with the teacher’s one and used for training. In this way are created most of the smaller and faster models such as TinyBERT and DistilBERT (Sanh et al., 2019).

Key literature (slide 6)

The funding base of every modern transformer-based model is the paper “Attention is all you need” by Vaswani et al., 2017. This study introduced the self-attention mechanism that is at the heart of today’s models like BERT and GPT. Attention overcame a problem of “short memory”, what memory does is allowing the decoder, which is the second part of the model, to keep track of all the encoder past states, therefore extracting information from the whole sentence.

The studies of the smaller BERT models will then be fundamental since one of these models will be the base for our research. Two of the most diffused models are DistilBERT by Sanh et al., 2020, Huggingface researchers and TinyBERT by Xiaoqi J. et al., 2020, researchers at Huawei and Huazhong university.

Finally, we have the papers about transfer learning techniques, which are fine tuning and feature extraction. One of the fundamental papers in this domain I want to cite is UMLFit by Howard J. & Ruder S., 2018 where they introduce transfer learning techniques for NLP specific models. Apart from UMLFit the most important papers about transfer learning for sentiment analysis and in particular in the financial sector will be examined, including the ones that used large models such as FinBERT (Aract D., 2019) or FinALBERT (Mukul J. et al., 2021).

Methodology (slide 7)

My proposed methodology for this research on transfer learning will be primarily quantitative and experimental in nature, with careful observation of the results obtained.

To begin, I will collect a dataset of financial news articles and social media posts related to various companies and their stocks or evaluate existing and open-sourced ones, along with their corresponding sentiment labels. I will then train or tune a fast and lightweight model on this data.

Next, I will conduct a series of experiments to evaluate the performance of the transfer learning model compared to other standard deep learning models, such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs), on the same dataset. These experiments will involve varying hyperparameters, such as learning rate and batch size, and comparing various performance metrics, such as accuracy and F1 score.

Throughout these experiments, I will make careful observations of the behavior of the transfer learning model, particularly its ability to adapt to the specific task of financial sentiment analysis despite being pre-trained on a different corpus of text data.  
Finally, using twitter APIs, I will test the model on live streaming data to assert its performance and capabilities to keep up with an incoming flow of data and produce usable insights for market investors.

Ethics (slide 8)

We can see on the slide three main focuses: data, society and laws.

As a researcher the ethics behind any research is fundamental. While this project does not involve any external participant, there are still ethical considerations and concerns to address.

I will ensure that the data used for this project is obtained ethically and legally. I will obtain the data from publicly available sources and ensure that it is properly labeled for sentiment analysis. Additionally, I will take measures to protect the privacy of individuals mentioned in the data by removing any personally identifiable information (GDPR, 2023).

Secondly, I will ensure that the results of this research are reported truthfully and accurately, without exaggeration or misleading claims. It is important to be transparent about the limitations and potential biases of the research, and to present results in a way that is clear and understandable.

Furthermore, I will consider the potential impact of this research on society and take steps to minimize any negative consequences (UoE, 2020). For example, if the research leads to the development of a financial sentiment analysis tool, it will be important to ensure that the tool is not used to manipulate financial markets or harm investors.

Finally, I will ensure that this research is conducted in compliance with all relevant laws and regulations, including those related to data privacy and intellectual property. Any potential conflicts of interest or biases will be disclosed and addressed appropriately (UoE, 2020).

Timeline and final artifact (slide 9)

The proposed timeline for this research starts with the literature review of state-of-the-art models and datasets. A thorough research of these two fundamental parts will be essential for the quality of the final output. After the initial review a I will choose a model that is as lightweight and fast as possible without sacrificing accuracy. Quality of datasets is an incredibly important part of machine learning. Then a phase of testing will start, the model will undergo different types of transfer learning optimizations following what was learned from the key literature.

After the testing phase the most performant model will be tested against other models with the same testing datasets to compare results in performance metrics such as accuracy and F1 score.

Finally, the model will undergo a data live streaming test and try to produce live useful insights on determined market areas sentiment to find out if it can perform well enough to be useful in real world scenarios. Depending on the performance a web application could be a good final artifact.

References

Aract D., (2019). FinBERT: Financial Sentiment Analysis with Pre-trained Language Models. University of Amsterdam. Available at: <https://arxiv.org/pdf/1908.10063.pdf> [Accessed: 22/03/2023]

Howard J. & Ruder S., (2018). Universal Language Model Fine-tuning for Text Classification. University of San Francisco, NUI Galway Aylien Ltd. Available at: <https://arxiv.org/pdf/1801.06146.pdf> [Accessed: 30/03/2023]

Huggingface, (2022). Distilgpt2. Huggingface. Available at: <https://huggingface.co/distilgpt2> [Accessed: 25/03/2023]

Lacoste a. et al., (2019). Quantifying the Carbon Emissions of Machine Learning. Montreal University. Available at: <https://arxiv.org/pdf/1910.09700.pdf> [Accessed: 29/03/2023]

Mukul J. et al., (2021). Text Mining of Stocktwits Data for Predicting Stock Prices. MDPI Applied system innovation. Available at: <https://arxiv.org/ftp/arxiv/papers/2103/2103.16388.pdf> [Accessed: 30/03/2023]

Sanh V. et al., (2020). DistilBERT, a distilled version of BERT: smaller,

faster, cheaper and lighter. Huggingface. Available at: <https://arxiv.org/pdf/1910.01108.pdf> [Accessed: 29/03/2023]

University of Essex, (2020). Code of Good Research Practice. UoE. Available at: <https://www.essex.ac.uk/governance-and-strategy/research-integrity> [Accessed: 31/03/2023]

Vaswani A. et al., (2017) Attention is all you need. 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA. Available at: <https://arxiv.org/pdf/1706.03762.pdf> [Accessed: 23/02/2023]

Xiaoqi J. et al., (2020). TinyBERT: Distilling BERT for Natural Language Understanding. Huawei Noah’s Ark Lab 3Huawei Technologies Co., Ltd. Available at: <https://arxiv.org/pdf/1909.10351.pdf?trk=public_post_comment-text> [Accessed: 25/03/2023]