Literature Review

Final Project: Social Media Sentiment Analysis of Financial Subreddit Groups

Ronnie Nop
XCS224U - Natural Language Understanding
March 2022

General Problem

Social media places a mirror in front of humanity reflecting on both the light and dark aspects of human nature in how we as humans learn, provide insights, and understand the world. The opinions and emotions we express on online social media platforms provide an invaluable corpus of text in which artificial intelligent systems can process to extract general sentiment among a community. This makes sentiment analysis extremely useful as it allows us to automate the process of analyzing huge amounts of text and determine the general opinion of a community on particular topics.

The rise of deep learning has resulted in the development of advanced algorithms for solving various natural language tasks. In the area of sentiment analysis, deep learning models such as RNNs, LSTMs, and language models have been trained to classify the text sentiment of various works of literature and corpora. However, even with these recent advancements in technology, there are still challenges that researchers face when attempting to extract sentiment:

1. Lack of supervised labeled text
- While the actual text data is relatively easy to find and collect, there are actually very few sources of labeled data that can be used to train machine learning and deep learning models. This is further challenging for underrepresented communities where the data is even more limited.

2. Using social media sentiment as an indicator for predictive power
- The true influence of social media sentiment is challenging to quantify because there may be outside factors that could better explain outcomes of interest.
- Social media sentiment could either be a leading or lagging indicator.

3. Inference and serving time for deep models
- Machine learning and deep learning techniques may find it difficult to serve predictions within a reasonable amount of time making them inviable for particular tasks. This is especially true for complex architectures such as BERT.

Summaries of Articles

---

1. Araci, Dogu. "FinBERT: Financial sentiment analysis with pre-trained language models." *arXiv* https://arxiv.org/pdf/1908.10063.pdf (2019)

Summary:

This paper tackles the ineffectiveness of general machine learning models due to the lack of labeled training data in domain-specific corpora, specifically within the financial domains. Araci (2019) introduces FinBERT, a pre-trained NLP model built on top of the BERT-based model that was pre-trained on financial texts resulting in one of the first applications of transformers in the financial domain. Their research resulted in a +15% increase in accuracy for classification tasks compared to the previous state-of-the-art models at the time. One of the key ideas this paper investigates is the advantageous effects of fine-tuning the final layers of the language model to specific tasks and was found that even with significantly less training data, it still achieved state-of-the-art performance.

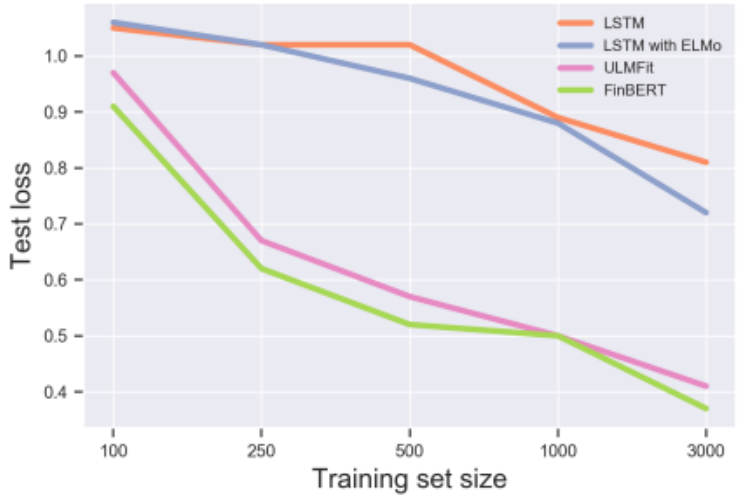

Figure 2: Test loss different training set sizes

- The chart above shows the ability of FinBERT to outperform other deep learning architectures out-of-sample even with the smallest training set size.

2. Hutto, C.J. and Gilbert, Eric. “VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text.” *eegilbert* http://eegilbert.org/papers/icwsm14.vader.hutto.pdf (2014)

Summary:

This paper introduces a rule-based model for sentiment analysis and compares its performance against machine learning classifiers such as Naive Bayes and Support Vector Machines. Researchers utilize the knowledge of sociologists, psychologists, and computer scientists to develop VADER (Valence Aware Dictionary for sEntiment Reasoning), a sentiment lexicon consisting of words and their sentiment intensity measurements for positivity, neutrality, and negativity.

One of the advantages of using VADER is its ability to handle the sentiments related to punctuation, lowercase/uppercase texts, contractions, slang, and emojis. Furthermore, the sentiment lexicons can be updated by experimenters to specific domains and communities utilizing their own distinct sentiment lexicons from other communities. Another advantage is that VADER is highly computationally efficient when compared to machine learning models such as SVMs and can avoid out-of-memory RAM issues. Lastly, it was found that VADER performed similarly to more advanced, complex techniques and even better in some cases.

	3-Class Classification Accuracy (F1 scores) Test Sets			
	Tweets	Movie	Amazon	NYT
VADER	**0.96**	0.61	**0.63**	**0.55**
NB (tweets)	0.84	0.53	0.53	0.42
ME (tweets)	0.83	0.56	0.58	0.45
SVM-C (tweets)	0.83	0.56	0.55	0.46
SVM-R (tweets)	0.65	0.49	0.51	0.46
NB (movie)	0.56	**0.75**	0.49	0.44
ME (movie)	0.56	**0.75**	0.51	0.45
NB (amazon)	0.69	0.55	0.61	0.48
ME (amazon)	0.67	0.55	0.60	0.43
SVM-C (amazon)	0.64	0.55	0.58	0.42
SVM-R (amazon)	0.54	0.49	0.48	0.44
NB (nyt)	0.59	0.56	0.51	0.49
ME (nyt)	0.58	0.55	0.51	0.50

Table 5: Three-class accuracy (F1 scores) for each machine trained model (and the corpus it was trained on) as tested against every other domain context (SVM models for the movie and NYT data were too intensive for our multicore CPUs with 94GB RAM)

3. Kabbani, Taylan and Enes Usta, Fatih. “Predicting The Stock Trend Using News Sentiment Analysis and Technical Indicators in Spark.” arXiv https://arxiv.org/pdf/2201.12283.pdf (2022)

Summary:

This is a recent paper that researches the challenges of predicting the stock market using sentiment analysis of financial news sources as a feature used to train machine learning models as a binary classification problem. Researchers utilized Pyspark to efficiently deal with the large corpus of news articles and various technical indicators of Apple (AAPL), Amazon (AMZN), and Netflix (NFLX) stock between January 2016 to April 2020.

Researchers performed sentiment analysis using the VADER sentiment model processing 2.7 million news articles that we pre-processed and tokenized into word vectors. The overall daily sentiment score was computed by aggregating the sentiment scores of all the articles released on that day. This aggregated sentiment score was used as a feature to train the machine learning model.

Models were evaluated on the accuracy, precision, recall, and f1 score, k-fold Cross-Validation, and was found that the Random Forest performed the best with an accuracy of 0.6358 out-of-sample.

Classification performance of algorithms on testing data set of Netflix (NFLX) in terms of overall accuracy, precision, recall, and F-measure				
Classes	**Metrics**	**Algorithms**		
		LR	**RF**	**GBM**
	Overall Accuracy (%)	**0.61**	**0.63**	**0.61**
Uptrending	**Precision (%)**	**0.58**	**0.61**	**0.59**
	Recall (%)	**0.81**	**0.8**	**0.74**
	F-measure (%)	**0.62**	**0.64**	**0.61**
Downtrending	**Precision (%)**	**0.66**	**0.68**	**0.63**
	Recall (%)	**0.4**	**0.46**	**0.47**
	F-measure (%)	**0.59**	**0.62**	**0.60**

4. Liu, Yinhan, et. al. “RoBERTa: A Robustly Optimized BERT Pretraining Approach.” *arXiv* https://arxiv.org/pdf/1907.11692.pdf (2019)

Summary:

This paper introduces a variant of the large BERT language model where researchers looked to modify the model design and optimize the hyperparameter search that was not investigated in the original BERT paper. One of the major differences in RoBERTa is that it utilizes dynamic masking where a masking pattern is generated at each instance a sequence is fed into the model whereas BERT utilized static masking that performed more upstream during the data pre-processing step.

Researchers also found that removing the Next Sentence Prediction (NSP) task in the original BERT resulted in matching or slightly better performance and concluded to remove it from the RoBERTa model.

RoBERTa is also trained on a larger corpus than BERT along with a larger batch size with half as many optimization steps during training which resulted in better performance. The researchers implemented more advanced parallel processing techniques to explore their wider hyperparameter space compared to BERT.

Model	**data**	**bsz**	**steps**	**SQuAD** (v1.1/2.0)	**MNLI-m**	**SST-2**
RoBERTa						
with BOOKS + WIKI	16GB	8K	100K	93.6/87.3	89.0	95.3
+ additional data (§3.2)	160GB	8K	100K	94.0/87.7	89.3	95.6
+ pretrain longer	160GB	8K	300K	94.4/88.7	90.0	96.1
+ pretrain even longer	160GB	8K	500K	**94.6/89.4**	**90.2**	**96.4**
$\text{BERT}_{\text{LARGE}}$						
with BOOKS + WIKI	13GB	256	1M	90.9/81.8	86.6	93.7
$\text{XLNet}_{\text{LARGE}}$						
with BOOKS + WIKI	13GB	256	1M	94.0/87.8	88.4	94.4
+ additional data	126GB	2K	500K	94.5/88.8	89.8	95.6

5. Sanh, Victor, Debut, Lysandre, Chaumond, Julien, and Wolf, Thomas. "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter." arXiv https://arxiv.org/pdf/1910.01108.pdf (2019)

Summary:

This paper was published by the team at Hugging Face looking to propose a smaller language model, DistilBERT, as a faster and cheaper alternative to BERT while still achieving similar performances. There are many key advantages to researching smaller alternatives while maintaining performance: decreased environmental costs, computational costs, training costs, and serving time.

The researchers accomplish this task by utilizing a compression technique called knowledge distillation where a compact 'student' model is trained to reproduce the habits of a larger 'teacher' model. The training objective of the student model becomes a combination of the distillation loss, a standard supervised classification loss, and an added cosine embedding loss. The final DistilBERT student model retains about 97% of BERT's performance while being 40% smaller. Another key advantage is that inference is also 60% faster which is important for serving purposes.

Table 1: **DistilBERT retains 97% of BERT performance.** Comparison on the dev sets of the GLUE benchmark. ELMo results as reported by the authors. BERT and DistilBERT results are the medians of 5 runs with different seeds.

Model	**Score**	CoLA	MNLI	MRPC	QNLI	QQP	RTE	SST-2	STS-B	WNLI
ELMo	68.7	44.1	68.6	76.6	71.1	86.2	53.4	91.5	70.4	56.3
BERT-base	79.5	56.3	86.7	88.6	91.8	89.6	69.3	92.7	89.0	53.5
DistilBERT	77.0	51.3	82.2	87.5	89.2	88.5	59.9	91.3	86.9	56.3

Table 3: **DistilBERT is significantly smaller while being constantly faster.** Inference time of a full pass of GLUE task STS-B (sentiment analysis) on CPU with a batch size of 1.

Model	# param. (Millions)	Inf. time (seconds)
ELMo	180	895
BERT-base	110	668
DistilBERT	66	410

Compare and Contrast

The papers chosen for the literature review focused on the different subtasks of sentiment analysis for social media and approaches to tackle the problem holistically.

1. Lack of supervised labeled text
- FinBERT and RoBERTa papers suggest fine-tuning a pre-trained language model to achieve state-of-the-art performance.

2. Using social media sentiment as an indicator for predictive power
- Kabbani, Taylan, and Enes Usta (2022) suggest a combination of technical indicators and sentiment as features to train a machine learning model for predicting the direction of a stock.

3. Inference and serving time for deep models
- DistilBERT paper suggests implementing a smaller language model that achieves similar predictive performance while decreasing training costs and inference and serving time
- VADER proposes a rule-based model that achieves similar performance to that of more complicated machine learning models while being highly computationally efficient resulting in the fastest inference and serving time.

Future Work

It seems indicative that the direction of future research will be exploring ways to reduce the size and computational costs of these large language models while maintaining similar or better performance. The advantages include reducing our environmental footprint of training such large models that seemingly only result in small improvements in performance. Reducing the model size also democratizes them to allow researchers across the world with less access to powerful computer systems to utilize them for their own specific tasks.

Within the finance domain, the application of alternative data (e.g. sentiment analysis) as a data-driven approach to quantitative trading has been increasingly popular with the emergence of machine learning and deep learning. It would be interesting to research the approaches of hedge funds utilizing sentiment as part of their trading strategies and their overall trading performance and realized profits.

For the final project, it may be interesting to explore implementing simpler models such as Naive Bayes and SVMs, then transitioning to complex language models such as FinBERT and RoBERTa, as well as a rule-based model such as VADER.

When computing sentiment, it will be important to keep in mind how the extracted sentiment themselves can be useful to generate actionable insights. This could mean constructing them as an important feature to train a supervised machine learning model or a standalone indicator to generalize the overall opinion of a specific community.

Lastly, it may be important to think about model inference and serving depending on the overall goal of the sentiment analysis task. Language models such as FinBERT and RoBERTa may provide great performance but if the task relies on real-time inference, then these models may not be viable.

References

Araci, Dogu. "FinBERT: Financial sentiment analysis with pre-trained language models." *arXiv* https://arxiv.org/pdf/1908.10063.pdf (2019)

Hutto, C.J. and Gilbert, Eric. “VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text.” *eegilbert* http://eegilbert.org/papers/icwsm14.vader.hutto.pdf (2014)

Kabbani, Taylan and Enes Usta, Fatih. “Predicting The Stock Trend Using News Sentiment Analysis and Technical Indicators in Spark.” *arXiv* https://arxiv.org/pdf/2201.12283.pdf (2022)

Liu, Yinhan, et. al. “RoBERTa: A Robustly Optimized BERT Pretraining Approach.” *arXiv* https://arxiv.org/pdf/1907.11692.pdf (2019)

Sanh, Victor, Debut, Lysandre, Chaumond, Julien, and Wolf, Thomas. “DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter.” *arXiv* https://arxiv.org/pdf/1910.01108.pdf (2019)