Leveraging BERT to Improve the FEARS Index for Stock Forecasting

Linyi Yang[1], Yang Xu[2], Tin Lok James Ng[3], Ruihai Dong[1]
[1]Insight Centre for Data Analytics, University College Dublin, Dublin, Ireland
[2]Beijing Dublin International College, Beijing University of Technology, Beijing, China
[3]School of Mathematics and Applied Statistics, University of Wollongong, Australia
{linyi.yang, ruihai.dong}@insight-centre.org; yangxu@bjut.edu.cn; jamesng@uow.edu.au

Abstract

Financial and Economic Attitudes Revealed by Search (FEARS) index reflects the attention and sentiment of public investors and is an important factor for predicting stock price return. In this paper, we take into account the semantics of the FEARS search terms by leveraging the Bidirectional Encoder Representations from Transformers (BERT), and further apply a self-attention deep learning model to our refined FEARS seamlessly for stock return prediction. We demonstrate the practical benefits of our approach by comparing to baseline works.

1 Introduction

Efficient Market Hypothesis proposed by Fama [1965] states that stock market prices are driven by all observable information. In reality, it has been shown that investor sentiment can affect the asset prices due to the well-known psychological fact that investors with positive (negative) sentiment tend to make overly optimistic(pessimistic) judgments and decisions [Keynes, 1937]. These two classic theories open the gate of financial forecasting.

More recently, numerous empirical studies also provide consistent evidence to support the theory that investor sentiment has a significant impact on asset prices [Barber and Odean, 2007; Yuan, 2015; Da et al., 2011, 2014]. For instance, Peng and Xiong [2006] show that limited investor attention leads to category-learning behavior, i.e., investors tend to process more market-wide information than firm-specific information. Baker and Wurgler [2006] present evidence that the investor sentiment has effects on stock price movements across different stocks. They construct a novel investor sentiment index (BW index hereafter), and find that high investor sentiment predicts strongly low returns in the stock market. Vozlyublennaia [2014] investigates a link between the performances of several security indices in broad investment categories with the exception of exchange rates. He finds a significant short-term change in index returns following an increase in attention. By analyzing the effect of attention on stock market, Yuan [2015] demonstrates that investor attention is one of the factors that inherently causes individual investors in aggregate to alter their stock positions dramatically, and he also suggests that the findings have implications for other research in finance.

In an influenced study, Da et al., [2014] build a list of search terms based on Google search volume in the U.S. market for various keywords that reveal sentiment toward economic conditions. By constructing a Financial and Economic Attitudes Revealed by Search (FEARS) index as a measure of investor sentiment, they find that “FEARS” is able to predict both short-term return and temporary increases in volatility. Tetlock [2007] shows that negative terms in English language are more useful for identifying sentiment compared to positive words. For this reason, the list consists of thirty negative search terms derived from words of economic sentiment in the Harvard and Lasswell dictionary [Tetlock, 2007] which have had the largest negative correlation with the market. The constructed list includes terms such as "gold prices" and "recession", which historically have had the largest daily correlation with the stock market. Finally, the FEARS index is defined by simply aggregating the change of each term’s search volume, which implies that each term contributes equally to the FEARS index.

However, it may not be appropriate to assume that each of the search terms has the same level of contribution to stock market forecasting. Since previous works have not taken into account the semantics of the search terms in modelling their effects on the price movements. Moreover, the fluctuation of the volume of a search term may have a different effect on stock price movements on different days due to the complex dynamics of financial markets. Therefore, we argue that the current method of calculating the index is far from optimal. In this paper, instead of calculating index by simply aggregating the change of the thirty terms, FEARS index is refined by allocating different weights to different terms while the contribution is dynamic with the change of market.

In a nutshell, investor attention has been corroborated to be statistically and economically significant in security markets, while little research has been undertaken in the influence of the semantic information. To under the meaning of the search terms, Natural Language Processing (NLP) is leveraged. The

Proceedings of the First Workshop on Financial Technology and Natural Language Processing (FinNLP@IJCAI 2019), pages 54-60, Macao, China, August 12, 2019.

first key component in neural language understanding models is to find an approach to mathematically model words. A traditional method for representing words is the one-hot representation, where each word is represented as a binary vector with all but one entries of the vector are zero. Each integer value is represented as a binary vector that is all zero values except the index of the target word. However, there are two main shortcomings associated with such representations. First, the dimension of a vector increases accordingly when the number of words. Second, any two words represented by one-hot representation are isolated and cannot capture the information between words at the semantic level. In comparison, the use of a pre-trained word embedding allows clustering of similar words in a latent space, where semantically similar words are closer in the latent space. In recent years, language model pre-training has shown to be beneficial for improving downstream tasks of NLP [Peters *et al.*, 2017, 2018; Radford et al., 2018; Howard and Ruder, 2018].

Various extensions to word embedding have been proposed. For example, ELMo [Peters *et al.*, 2018] which is short for Embeddings from Language Models representation differ from traditional word embedding in that each token is assigned a representation, task-specific models are used to include the pre-trained representations as additional features. Besides, the Generative Pre-Trained Transformer (OpenAI GPT) [Radford et al., 2018] introduces a novel idea which involves fine-tuning the pre-trained parameters by jointly estimating task-specific parameters for the downstream tasks.

However, [Devlin, *et al.*, 2018] argue that the current techniques severely restrict the power of the pre-trained word representations. To address the limitation that the standard language models are unidirectional, Google improves the previous models of pre-training by proposing BERT: Bidirectional Encoder Representations from Transformers [Devlin, *et al.*, 2018]. They address the unidirectional constraints by proposing a new pre-training objective: the "masked language model" (MLM). Experimental results show that pre-trained representations eliminate the needs of many traditional heavily engineered task-specific models. It is one of the most representative works recently which can be seen as a milestone in the field of pre-training for language understanding.

By leveraging the pre-trained word embedding, many recent works have applied NLP techniques with multiple textual data sources to predict stock price movement [Si et al., 2013; Ding et al., 2014, 2015; Xu and Cohen, 2018]. Existing deep neural network approaches for stock price prediction have two main shortcomings. First, most of the proposed methods have focused on binary classifications of stock price movement (up or down). However, binary classification is less useful in the context of investment and financial risk management. To address this shortcoming, our developed methodology allows prediction of the return of a stock. Second, existing methods [Ding et al., 2015, Xu and Cohen, 2018, and Feng et al., 2018] typically employ the traditional approach in splitting the dataset into training and test sets in machine learning, whereby the first k% of the data is allocated to the training set, and the remaining to the test set. However, such approach is not suitable for predicting stock market return, as the financial market may encounter structure changes. Hence, we adopt the recursive forecast, which is a common method in finance [Han et al., 2018; Huang et al., 2017; Rapach et al., 2013].

In this paper, we first improve the construction of the FEARS index which represents the investor sentiment in order to get different input representations of search terms that integrating the semantic information. Then, we propose a self-attention neural network to predict the stock return using recursive training method.

The contributions of our papers are as follows:

- We propose a self-attention neural network with semantic information to predict the next short-term stock return and outperform the baseline works that only use financial index. To the best of our knowledge, semantical fears index is the first attempt to integrate sematic information with FEARS.
- We illustrate the importance of using semantic information for FEARS index to allocate different weights to different search terms.
- Unlike Si et al. [2013], Ding et al. [2014, 2015], and Xu and Cohen, [2018], we use recursive training for model estimation and prediction rather than the traditional way of splitting data into train and test sets.

2 Methodology

We introduce our model and its detailed implementation in this section. First, we provide an overview of the model architecture and the input representations. Then, we introduce our prediction model and the core innovation in this paper. Finally, the differences between our model and the classical model [Da *et al.*, 2014] are discussed in section.

2.1 Overview

The goal of our work is to leverage semantic information to improve FEARS for stock forecasting. To verify the performance of our refined FEARS index, a stock return predictive model is built in this paper. The previous state-of-the-art methods in text-based stock prediction connect the encoder and decoder through attention mechanisms [Si et al., 2013; Ding et al., 2014, 2015]. Hence, the Transformer network architecture [Vaswani *et al.*, 2017] proposed by Google, based solely on attention mechanisms is adopted in this paper for predicting the stock return. The Transformer is also known as self-attention mechanism.

Inspired by Vaswani *et al.*, 2017], we refine the FEARS index proposed by and [Da *et al.*, 2014] and test its efficiency in the task of stock return prediction. The overview of our model is shown in Figure 1. In general, our model contains four components:

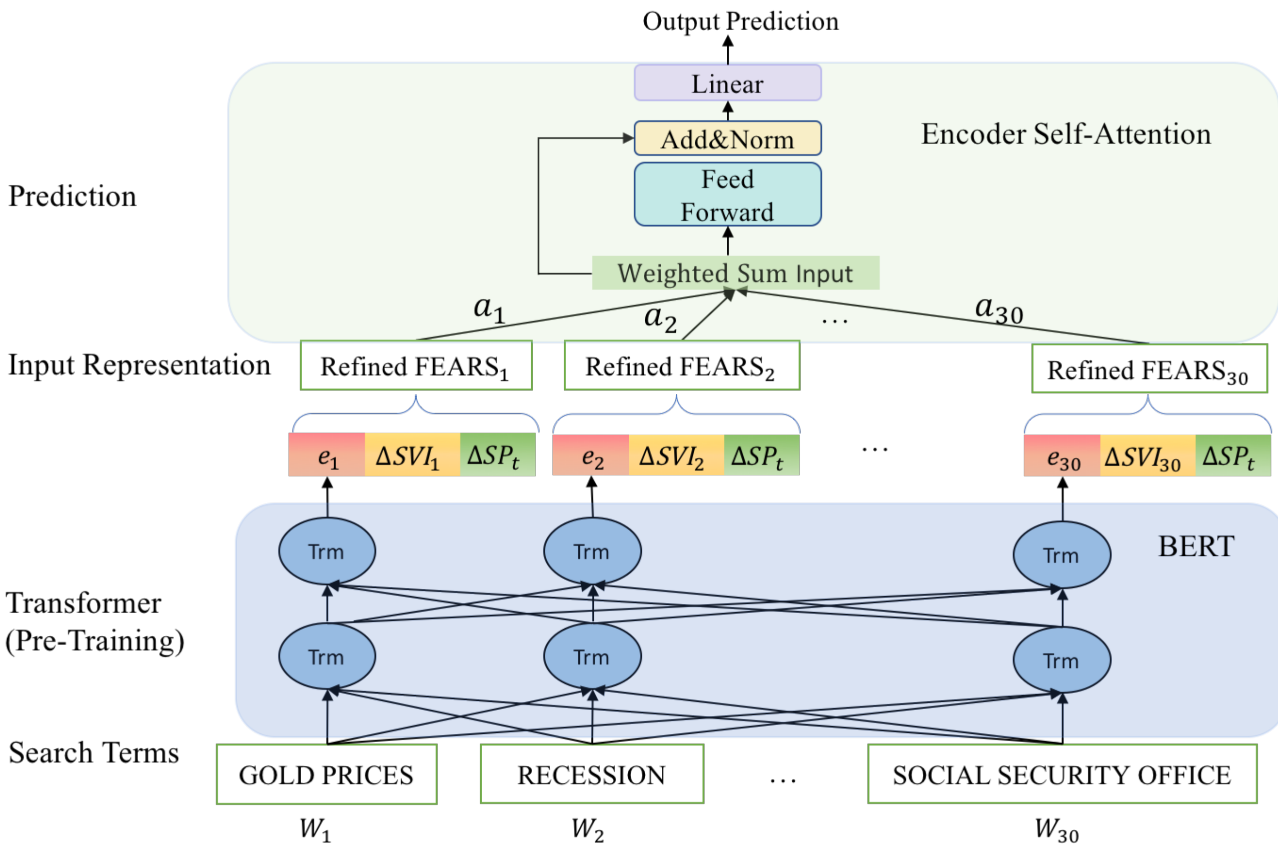

Figure 1: The architecture of our self-attention model for stock return prediction. The Pre-training model and input representation will be detailed introduced later in Section 2.2 and Section 2.3 respectively.

1) Data Source: Previously, Da [*et al.*, 2014] observe the stock return has the strong negative effect associated with investor sentiment that can be presented by FEARS and last week's stock return because of the return reversals in a short-term. Hence, our input contains both the FEARS and last week's stock return.
2) Embedding layer: First, the thirty-selected search terms [Da *et al.*, 2014] that have the largest negative influence for the market are pre-trained into embeddings based on the architecture of BERT. As a result, semantically similar search terms are mapped into similar locations in a latent space.
3) Input Representation: We obtain our input representation by combining the embeddings of the search terms with stock return and change in search volume.
4) Prediction Layer: The self-attention mechanism is used to allocate different weights to different search terms for stock return prediction. We then output the prediction of the next week's stock return using a dual-layer feed forward neural network and the mean squared error is used as loss function to train the entire model.

Our model contains a bidirectional Transformer encoder-based on the original implementation described in [Vaswani *et al.*, 2017] and a self-attention prediction model. We will first introduce the transformer model for word embedding, then cover the model for prediction using self-attention mechanism.

2.2 Transformer for search terms embedding

Firstly, BERT is used as a term encoding service to map our variable-length selected search terms to a fixed-dimension vector. BERT is a language representation model developed by Google. It leverages an enormous amount of public textual data on the web and is trained in an unsupervised fashion. Pre-training a BERT model is fairly expensive and time-consuming process. Hence, in this paper, a pre-trained model that contains 110M parameters is used to obtain representations of our search terms. The pre-trained model can be downloaded from Google[1].

We use BERT as a terms encoder and hosts it as a service via ZeroMQ [Hintjens, 2013] to map our search terms into fixed-dimension vectors E = $\{e_1, e_2 \dots e_{30}\} \in \mathbb{R}^{30\times D}$, where D is the dimension of the phrases embedding of the search terms

[1] https://github.com/google-research/bert#pre-trained-models

in the current timestamp, and the length of search terms is 30. We apply PCA [Jolliffe, 2011] to the embedded vectors in order for visualization, and the results are shown in Figure 2.

As a preliminary step, we examine if the word embeddings obtained from the pre-trained model can reasonably represent the semantic relatedness of the words. Following [Da *et al.*, 2014], we just select the most influenced 30 negative search terms without any positive terms. Since in Tetlock (2007) it appears that negative terms in English language are most useful for identifying sentiment. As illustrated in Figure 2, terms with similar economic interpretations are closer in the projected two-dimensional space, and vice versa. We observe the clustering of the search terms “gold”, “gold prices” and “price of gold” that are all related to the precious metal gold, which is normally perceived as “safe heaven” of the capital market. Intuitively, capital inflows to gold market dramatically increase when equity markets experience bearish condition.

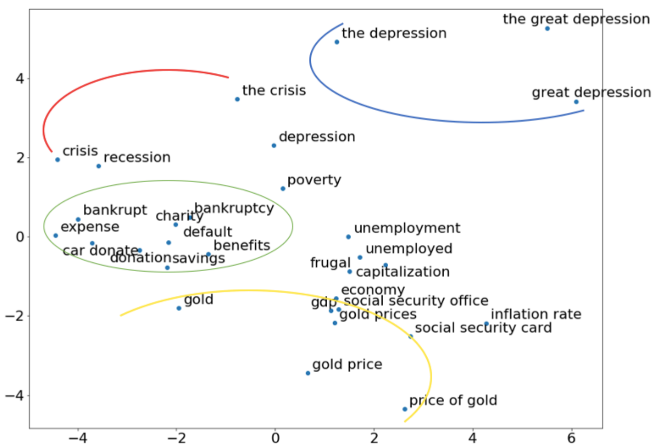

Figure 2: The visualization of our thirty selected search terms after embedding.

2.3 Input Representation

Da et al. [2014] found strong negative association between FEARS index and S&P 500 index daily return and claim that FEARS can be used as a proxy for investor sentiment. The calculation of FEARS involves only averaging the change in the search volume of the thirty selected terms. However, as explained in the previous section, such approach ignores the semantics of the search terms which may result in inferior predictions.

To address this issue, we propose a novel method to refine FEARS index by integrating the semantic information with the original calculation. Consequently, different weights will be assigned to each of the search term separately using a self-attention mechanism., In addition, the weights will be dynamically adjusted as the financial market evolves over time. We define the FEARS index corresponds to search term i term on day t as:

$$\boldsymbol{FEARS_{i,t}} = \boldsymbol{e_i} \times (\boldsymbol{\alpha \Delta SVI_{i,t}} + (\mathbf{1} - \boldsymbol{\alpha})\boldsymbol{\Delta SP_t}) \quad (1)$$

where e_i is the term embedding trained by BERT $e_i \in \mathbb{R}^{768}$, $\boldsymbol{\Delta SVI_{i,t}}$ represents the weekly change in search volume for the search term *i*:

$$\boldsymbol{\Delta SVI_{i,t}} = (\boldsymbol{SVI_{i,t}} - \boldsymbol{SVI_{i,t-1}})/\boldsymbol{SVI_{i,t-1}} \quad (2)$$

similarly, the $\boldsymbol{\Delta SP_t}$ is the S&P500 index return in trading day t, it can be calculated as:

$$\boldsymbol{\Delta SP_t} = (\boldsymbol{SP_t} - \boldsymbol{SP_{t-1}})/\boldsymbol{SP_t} \quad (3)$$

Finally, we generate our input representations in every timestamp to predict the next timestamp’s S&P500 return later:

$$\boldsymbol{FEARS} = \{\boldsymbol{FEARS_1}, \boldsymbol{FEARS_2}, \ldots \boldsymbol{FEARS_l}\} \quad (4)$$

where $\boldsymbol{FEARS} \in \mathbb{R}^{l \times 30 \times 768}$, and l, 30, and 768 represent the length of the dataset, the thirty selected terms and the dimension of the latent space, respectively.

We adjust the fine-tuning procedure of the original BERT model [Devlin, et al., 2018] for our prediction task. Different from the original model where all parameters of BERT and the additional layer are fine-tuned jointly to minimize the loss, we fix the pre-trained word embeddings where are used to form our input representation. Keeping the embeddings fixed allows us to speed up our model training. Besides, we deprecate the position embedding since our input is not a sentence, in contrast to classical NLP tasks.

2.4 Predictive Model

The calculation of the prediction stock return *r* is shown below:

$$\boldsymbol{u_{t,i}} = \boldsymbol{ReLU}(\boldsymbol{W} \cdot \boldsymbol{FEARS}_{\boldsymbol{t,i}} + \boldsymbol{b}) \quad (5)$$

$$\boldsymbol{a_{t,i}} = \boldsymbol{exp}(\boldsymbol{u_{t,i} u^T}) / \textstyle\sum_{i'=1}^{30} \boldsymbol{exp}(\boldsymbol{u_{t,i'} u^T}) \quad (6)$$

$$\boldsymbol{v_t} = \textstyle\sum_{i=1}^{30} (\boldsymbol{a_{t,i} FEARS_{t,i}}) \quad (7)$$

$$\boldsymbol{r_t} = \boldsymbol{FFN}(\boldsymbol{v_t}) \quad (8)$$

where $u \in \mathbb{R}^{30 \times h}$ is the query vector, and equals to key and value vector in the self-attention mechanism; h denotes the number of hidden units; the weight matrix $W \in \mathbb{R}^{D \times h}$, D is the dimension of input; $v_t \in \mathbb{R}^D$ is the weighted sum of inputs. FFN is the short for feed forward neural network.

The use of a self-attention mechanism allows allocating different weights to the words when making out of sample predictions, it decides which word should be paid more attention by calculating the similarity between the query vector and the key vector. Then multiply the value vector with the score of the similarity after softmax. In the encoder self-attention mechanism, the query vector, key vector and value vector are all itself. We adopt the dropout strategy for training our model. Finally, we get our prediction for the return of S&P500 index with a dual-level FFN in this paper.

2.5 Recursive Training

The standard approach [Ding et al. 2015; Xu and Cohen, 2018; and Yang et al. 2019] in splitting the data into training and test sets will generally not work well in financial applications, due to the difficulty in predicting stock returns using data from years ago. On the other hand, updating trading strategy periodically is a common approach in quantitative finance [Han et al., 2018; Huang et al., 2017; Rapach et al., 2013].

Hence, we apply the online-learning (recursive training) method to build our model. The parameters will be updated with the loss in last training period. In this paper, we set the training length as 8 weeks while the testing length is 4 weeks. It means that we make use of the last 8 weeks' data to train our model while the next 4 weeks' data for testing. The stock return of S&P500 will be recursively predicted by repeating this process.

3 Experiments

Our experiments aim to demonstrate that the semantic information integrating with search volume is beneficial to predict the stock return. In this section, we first introduce the procedure of the collecting the weekly search index and S&P500 return. Secondly, we discuss the loss function used in this paper. Next, we will specify the hyper-parameters in Section 3.3. Finally, we compare the performance of our method on S&P 500 index prediction to demonstrate the effect of self-attention mechanism with semantic information.

3.1 Data Collection

We use S&P 500 market index as the proxy for the US equity market and the historical prices are obtained from Quandl[2]. Stock returns are computed as the change in end-of-week settlement prices.

To construct FEAR index for the US stock market, we use the public Search Volume Index (SVI) from Google Trends[3] as attention proxies, following Da et al. (2014) and Han et al. (2018). The numbers present search probabilities of a given keyword at a given time. We consider the 30 terms that have been proven to be effectively associated with security prices from Da et al. (2014). These terms are suggested to contain information on financial markets and useful to predict future stock prices. All attention data cover a weekly period of 2004:01-2015:12. We work in logarithms of search terms probabilities for ease of exposition and notation.

3.2 Evaluation Metric

We use the Mean Square Error (MSE) to evaluate our model in stock return prediction. MSE is calculated as:

$$\mathbf{MSE} = \mathbf{1/n} \times \sum_{t=1}^{n} (\boldsymbol{p_t} - \widehat{\boldsymbol{p_t}}) \tag{9}$$

Where n denotes the length of total test sets, $\boldsymbol{p_t}$ is the true value of the S&P500 index while $\widehat{\boldsymbol{p_t}}$ represents the output of our model at timestamp t.

3.3 Experiment Setup

Hyper-parameters for BERT. The hyper-parameters are shown in Table 1 and are the same as in the model BERT-BASE.

Settings	
Embedding Dimension	768
Number of Layers	12
Hidden Size	768
Attention Heads	12

Table 1: BERT-Base Hyper-parameters

Hyper-parameters for Prediction. Since the BERT-Base model we applied has 110M parameters. Hence, we change the terms embedding to non-trainable variables in our model. That is, we train our two models separately in order to speed up the training process. Experimental hyper-parameters of the prediction model are shown in Table 2.

Settings	
Input Size	768
Number of Layers	2
Hidden Units	256
Epochs	6
Batch Size	2
Optimizer	Adam
Dropout Probability	0.6

Table 2: Hyper-parameters of Prediction Model

3.4 Experimental Results

In this section, we demonstrate the efficiency of our proposed model based on our experimental results. We first reproduce the baseline work of [Da et al., 2014], then we compare different ways of integrating the semantic information with the baseline work in terms of their performance on the weekly dataset we collected. We evaluate our model using the online-training strategy. Since there are no previous attempts on adopting non-linear method based on the FEAR index, we just compare our method with the original strategy proposed by [Da et al., 2014] in experiments.

Baseline:

- FEARS and Asset prices [Da et al., 2014]: They use daily Internet search volume from millions of households to reveal market-level sentiment. Then

[2] https://www.quandl.com

[3] http://www.google.com/trend

the volume of queries in U.S. are simply aggregated to construct FEARS. They finally use FEARS to predict short-term stock return by linear regression.

Our Method:

- We propose a novel model that integrates semantic information with the traditional financial index to predict the return of S&P500 index.

We test our methods and the baseline model using recursive forecast. The experimental results are shown in Table 3.

Method	MSE
Linear Regression (FEARS) [Da et al. 2014]	0.001094
Linear Regression (Optimal $\alpha = 0.6$)	0.000809
Transformer (FEARS)	0.001034
Transformer (Embedding$\times$ ΔSP; $\alpha = 0$)	0.000941
Transformer (Embedding$\times$ ΔSVI; $\alpha = 1$)	0.000678
Transformer (Optimal $\alpha = 0.6$)	**0.000585**

Table 3: Prediction Model

Since stock return prediction is a challenging task and a minor improvement usually leads to large potential profits. From Table 3, we can find that our model outperforms the baseline work in terms of the MSE loss.

Effects of Terms Embedding. First, in comparison with the performance of two works using linear regression model, we find that the MSE decreases if we construct the input representation with embedding. In addition, best predictive performance cannot be attained if only weekly aggregated search frequency (SVI) is used. These demonstrate the benefits of including the semantic information in the model, especially, the embedding of the search terms.

Effects of Self-Attention Mechanism. Second, the first two baseline methods simply take the sum of thirst search terms' values as input which are unable to capture the fact that they have different contributions on different days. The performance of the proposed model with self-attention mechanism show that transformer model architecture is useful to predict stock return. It is mainly because our proposed model can allocate different weights to different search terms in terms of their importance for prediction on different trading days during the online-training and testing.

Finally, inspired by two main conclusions in [Da et al. 2014], namely, 1) FEARS has negative effect on the market, 2) short-term stock return predictability is reflected in the contrarian effect, we investigate the relative importance of the change in FEARS and last week's S&P500 return on the performance of prediction.

The parameter α defined in Eq. (1) represents the tradeoff between the FEARS index and last week's stock return. The performance of the stock return prediction with a range of values for α is shown in Figure 3.

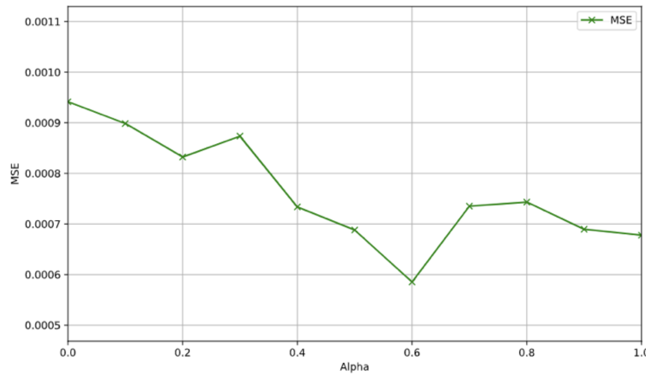

Figure 3: Performance of S&P500 weekly return prediction with varied α, see Eq. (1).

As shown in Figure 3, the best performance is achieved at α = 0.6. The MSE loss curve gradually decreases as α increases, before reaching its minimum at α = 0.6. It then ascends abruptly. The MSE loss curve finally remained relatively stable towards α = 1.

4 Conclusion

This paper proposes a novel method for refining the FEARS, which can leverage the embedding of search terms to better represent the investor sentiment. Also, a prediction model based on self-attention mechanism is introduced for stock return prediction. It aims to automatically allocate different weights to different search terms considering their contribution to the target trading day. The experimental results on our weekly dataset illustrate that the semantic information benefits the task of stock return prediction, while a trade-off between the price data and search volume data is useful to improve the performance.

In the future, there are two potential extensions of this work: 1) The dictionary of top 30 search terms is fixed in this work. It might be beneficial to dynamically update the search terms used for prediction of stock return for capturing some fresh significate keywords. 2) The trade-off parameter α now is fixed at 0.6 in this work, by allowing α to vary across time, we may achieve better performance at stock return prediction.

Acknowledgments

This research was supported by Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289.

References

[Keynes, 1937] The general theory of employment. *The quarterly journal of economics*, 51(2): 209-223, 1937.

[Fama, 1965] The behavior of stock-market prices. *The journal of Business*, 38(1): 34-105, 1965.

[Barber and Terrance, 2007] All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The review of financial studies 21*, no. 2: 785-818, 2007.

[Yuan, 2015] Market-wide attention, trading, and stock returns. *Journal of Financial Economics*, 116(3): 548-564, 2015.

[Da et al., 2011] In search of attention. *The Journal of Finance,* 66(5): 1461-1499, 2011.

[Vozlyublennaia, 2014] Investor attention, index performance, and return predictability. *Journal of Banking & Finance 41*: 17-35, 2014.

[Baker and Wurgler, 2006] Investor sentiment and the cross-section of stock returns. *The journal of finance.* 61(4): 1645-1680, 2006.

[Da et al., 2014] The sum of all FEARS investor sentiment and asset prices. *The Review of Financial Studies*, 28(1): 1-32, 2014.

[Tetlock, 2007] Giving content to investor sentiment: The role of media in the stock market. *The Journal of finance*, 62(3): 1139-1168, 2007.

[Peters et al., 2017] Semi-supervised sequence tagging with bidirectional language models. *In Proceedings of the 55th Annual Meeting of the 2017 Association for Computational Linguistics (Volume 1: Long Papers)*, pages: 1756-1765, 2017.

[Peters et al., 2018] Deep contextualized word representations. *In Proceedings of NAACL-HLT*, pages: 2227-2237, 2018.

[Radford et al., 2018] Improving language understanding with unsupervised learning. Technical report, OpenAI, 2018.

[Howard and Ruder, 2018] Universal Language Model Fine-tuning for Text Classification. *In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages: 328-339, 2018.

[Devlin et al., 2018] Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[Si et al., 2013] Exploiting topic based twitter sentiment for stock prediction. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages: 24-29, 2013.

[Ding et al., 2014] Using structured events to predict stock price movement: An empirical investigation. *In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages: 1415-1425, 2014.

[Ding et al., 2015] Deep learning for event-driven stock prediction. In *Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI),* 2015, pages: 2327-2333), 2015.

[Xu and Cohen, 2018] Stock movement prediction from tweets and historical prices. *In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages: 1970-1979, 2018.

[Vaswani et al., 2017] Attention is all you need. *In Advances in neural information processing systems 30 (NIPS 2017)*, pages: 5998-6008, 2017.

[Hintjens, 2013] ZeroMQ: messaging for many applications. *O'Reilly Media, Inc*., 2013.

[Jolliffe, 2011] Principal component analysis. *Springer Berlin Heidelberg.*, 2011.

[Yang et al., 2019] Explainable Text-Driven Neural Network for Stock Prediction. *arXiv preprint arXiv:1902.04994*, 2019.

[Han et al., 2018] Forecasting the CNY-CNH pricing differential: The role of investor attention. *Pacific-Basin Finance Journal 49*: 232–247, 2018.

[Huang et al., 2016] Forecasting stock returns in good and bad times: The role of market states. *In 27th Australasian Finance and Banking Conference*, 2016.

[Rapach et al., 2013] International stock return predictability: what is the role of the United States? *Journal of Finance 4*:1633–1662, 2013.