**Is Sentiment of Corporate Announcements a Good Predictor Of A Share Price Change?**

**The Case of Using FinBERT On A Dataset of Joint Venture Announcements**

1. Problem Statement and Results

There exists a significant amount of research related to the stock market reaction to corporate announcements. There is an empirical prove that tonality of corporate announcements can serve a predictor of share price behaviour, although the results of the research in this field are quite mixed. Automated sentiment detection using state-of-art models can help to solve the problem. This study makes an attempt to investigate the relationship between the share price dynamics and the tonality of management announcements, using FinBERT model to analyse the dataset of 50 manually collected corporate announcements about joint venture creation. Although the experiments conducted within the framework of this study do not point to any statistically significant relationship between share price change and the sentiment of announcements, it might be helpful in identifying the potential ways of improving the user approach to the model and also the dataset collection rules.

* 1. Purpose of this research

This research aims at investigating the correlation between the sentiment and the share price change. To evaluate the sentiment, we used the pre-trained and fine-tuned FinBERT model, which is ‘the language model based on BERT for financial NLP tasks’. It its turn, BERT is a Bidirectional Encoder Representations from Transformers model initiated by Devlin et al., 2018, and it is a language model construed as a stack of encoders (Vashwani et al, 2017). FinBERT was created with a purpose of analysing the finance domain-specific texts which have

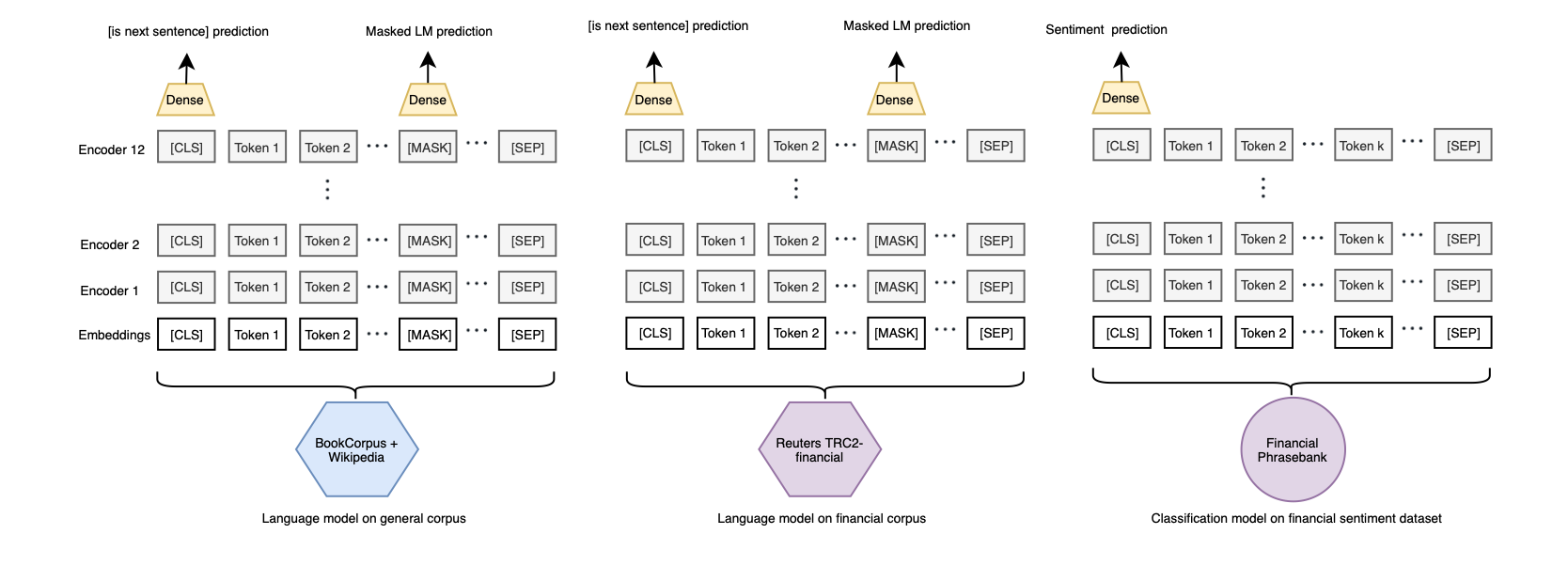
words and expressions different from generic texts. It was pretrained ‘on a large financial communication corpora of 4.9 billion tokens, including corporate reports, earnings conference call transcripts and analyst reports’, fine-tuned on 10,000 manually-annotated analyst statements, and it showed significant improvements in application on financial texts compared to the generic BERT (Yang et al, 2020).

Summarizing the existing studies on FinBERT, the three key elements of this model (also reflected in its name) are:

* Bi-directionality (the model looks both backs and forward). As pointed in the original article on BERT, this model can ‘pretrain deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context of all layers’ (Delvin et al, 2019). This allows to finetune the model using only one additional output layer, and also favorably compares BERT from one-directional models as it allows to better understand the context.
* ‘Transformers’ model, which is ‘relying entirely on an attention mechanism to draw global dependencies between input and output (Vaswani et al, 2017).
* Pre-trained contextualized word embeddings.

Below we present the chart showing the sentiment classification model architecture. The structure assumes a dense (non-linear) layer being imposed after the last hidden state, and this layer is used for the sentiment classification task (Chart 1).

Chart 1. Sentiment Classification Model Architecture.



Source: <https://openreview.net/pdf?id=HylznxrYDr>

* 1. Research question

The key research question is whether sentiment of an announcement can be a predictor of share price movement after a joint venture announcement.

1. Model and experimental design
   1. Model

For this experiment we used the finance-uncased pre-trained FinBERT. The model was sourced from <https://github.com/yya518/FinBERT>. The uncased model was chosen as the existing research points to better performance of uncased models compared to cased one as pointed to by Yang et al., 2020.

* 1. Dataset

The dataset was collected manually and consisted of 50 joint venture announcements. These announcements were sourced from the companies’ websites and from the newswire. In the latter case, we selected only those announcements which were mirroring the companies’ announcements without any changes, in order to preserve the style and vocabulary of the management and to avoid an additional layer of interpretation and sentiment added by third parties – journalists, analysts or investors. All of these announcements belong to different companies and different underlying transactions. All reporting companies had to be listed on the US stock exchange and registered in the US. This was done to ensure that language differences and corporate disclosure standards, which may differ across the countries, do not create biases.

Dataset pre-processing involved some manual manipulations. Firstly, as part of pre-processing, we removed the titles and the summary bullet points which generally repeat and summarize the main context of the text. We also removed the regulatory disclosures and companies’ descriptions in the end of texts, as well as details of investor calls where relevant.

Secondly, we removed spaces between the paragraphs. This was done to ensure that the model does not create expedient neutral-colored tokens for these spaces which might distort the results. (Of note, this step is potentially to be automated).

The next step was to label the dataset into several categories, based on the share price change on the date of announcement. For this, we used the yahoo-fin module which imports the historical share price data from Yahoo Finance website. At first, we split the dataset into two group: with a positive share price change and with a negative share price change. Additionally, we split the same dataset into three samples: with a strongly pronounced share increase on the day of announcement (above 0.8%), with limited share price increase (-0.8% to 0.8%) and with a strong negative change (more than -0.8%). The level of the price change was chosen by trial and error to have three relatively balanced groups within the dataset. The split of dataset using the share price change mirrored the approach used by Jaggi et al., 2021.

|  |  |
| --- | --- |
|  |  |

* 1. Experimental design
     1. Difference between the groups

As discussed above, for this analysis we split the dataset into two and three groups based on the share price change. For the experiment we used two non-parametric tests – Mann-Whitney test for the experiment with two group and Kruskal-Wallace test for the experiment with three groups. The choice of tests was defined by non-Gaussian distribution of the key variables and by the number of groups (two and three in the first and the second experiment, respectively).

For the assessment of the sentiment, we used the coefficient of 2 for positive sentences and negative sentences and the coefficient of 0.5 for neutral sentiment to contrast them. We also normalized the sentiment measures by dividing them by the total number of paragraphs in a text.

* + 1. Classification

The second part of the experiment was based on the classification models. To enrich the dataset, we added an additional feature -- the dominant topic -- extracted using the Latent Dirichlet Allocation (LDA) unsupervised model. Given the small and relatively homogenous dataset, we have manually chosen the number of topics at 5. We included these topics after having converted them into dummy variables. After including these features along with the sentiment measures, we used the min-max feature scaling method to minimize the scale effect. As the target variable, we used the labeled share priced change groups.

For the classification problem, we used the logistic regression, decision tree, random forest, KNN, LDA, Gaussian NB and SVC models. All models showed low performance on the test sample but the decision tree classifier performed slightly better than others. The results are presented in the next sections of the study.

1. Limitations and problems

The key limitation is the small size of the dataset. In the present study, we use the pre-trained and fine-tuned FinBERT model. As the dataset is quite small, it would not be possible to use it to fine-tune the model and to make it a better fit for the assessment of JV announcements as opposed to more generic financial texts. Also, a larger dataset would increase the robustness of the statistical tests.

1. Results

The initial observations point to the predominately neutral sentiment of all announcements, with a median normalized neutral sentiment of around 0.53 across the dataset. The median normalized positive sentiment is around 0.46, and negatively labeled parts of the dataset occurs very rarely – in 8% of the texts and only once per each text). The intuition behind it is that generally the tone of corporate announcements is deliberately impartial as part of corporate communication culture. That being said, another explanation to this might be incapacity of FinBERT to distinguish between the truly positive and truly neutral statements, with 73% of FinBERT’s misclassifications being between neutral and positive, and only 5% between negative and positive (<https://medium.com/prosus-ai-tech-blog/finbert-financial-sentiment-analysis-with-bert-b277a3607101>).

The results of group comparison point to absence of statistically significant difference between groups, with p value significantly above 0.05 both in two-group and three-group comparison.

The results of the classification problem shown below point to a weak power of the model, largely due to weak degree of correlation between the variables. The accuracy on the train set was high for some methods (in particular, the random forest method which yields relatively good results on non-Gaussian and small datasets) but the accuracy on the test set was low for all methods pointing to the overfitting problem (Tables 1, 2)

Table 1. Classification problem results

|  |  |  |
| --- | --- | --- |
| Model | Accuracy on the train set | Accuracy on the test set |
| Logistic Regression | 0.69 | 0.60 |
| Decision Tree | 0.94 | 0.47 |
| Random Forest | 0.94 | 0.47 |
| KNN | 0.63 | 0.33 |
| LDA | 0.71 | 0.53 |
| Gaussian LB | 0.54 | 0.40 |
| SVC | 0.66 | 0.53 |

Table 2. Classification report for the decision tree / random forest classifier.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1-score | Support |
| Price down | 0.33 | 0.20 | 0.25 | 5 |
| Price neutral | 0.50 | 0.20 | 0.29 | 5 |
| Price up | 0.50 | 1.00 | 0.67 | 5 |
| Accuracy |  |  | 0.47 | 15 |
| Macro Avg | 0.44 | 0.47 | 0.40 | 15 |
| Weighted Avg | 0.44 | 0.47 | 0.40 | 15 |

1. Improvements

The improvements to the analysis can be multifold, and can be broadly separated into two groups:

* 1. Dataset improvements
     1. Increasing the dataset size. As mentioned above, the small size of a dataset is amongst the key limitations of the experiment. Increasing the dataset could improve the robustness of the model and make it more generalizable.
     2. Using a different time window for estimating a share change movement. The dependent variable for this study is a share price change for only one day. The experiment could yield better results if the experiment were to be conducted on a large window, given the imperfect market efficiency of the real world and possible lower relevance of JV announcements as compared, for instance, to earnings announcements, which are priced in an assumingly more efficient manner.
     3. Using market models to estimate the share price change (as the current approach ignores the risk-free rate changes and stock betas).
     4. Including more variable in a model. This study is lop-sided in a way that the key variables are textual attributes. The model would greatly benefit from inclusion of company-related variables based on the rich empirical research pointing to a relationship between companies’ characteristics and the stock market reaction to corporate announcements.

* 1. Model improvements
     1. Fine-tuning the model on a relevant annotated dataset, as ‘data from the direct target might provide better target domain adaptation’ ( <https://openreview.net/pdf?id=HylznxrYDr>). This is a resource-intensive task (including manual labeling of the dataset to be used for training) and it is not practical for the current study; however, it could improve the weights and consequently the model’s classification accuracy.
     2. Changing weights of paragraphs. In the current study, we assign the same weights to sentiment of all paragraphs in a text. Nevertheless, looking at the sequence of sentiments across the dataset we might conclude that introductory and concluding paragraphs are largely neutral, and the main body of texts is more sentiment-loaded. Consequently, we could assign lower weights to the introductory and conclusive sentences to make the sentiment-colored paragraphs more meaningful in the model. Another approach, a more complex one, could imply imposing an additional step of summarizing the texts and conducting the analysis on this smaller but more relevant dataset. The idea is sourced from Ding el all, 2020 who compare BERT with human working memory, which employs not all messages but the key once for memorizing the overall message.
     3. Experimenting with truncation strategies. The current model uses the sequences defined by the max length hyperparameter. This means that for long sequences only part of the text is taking into account, and this is the beginning of the sequence. In many cases, the key message may hide in the middle of the sentence. So, moving the window to the middle of the sequence may add value.

1. Conclusion

To conclude, use of sentiment – complemented by other textual attributes of corporate announcements can be a promising way to predict the share price; however, achieving high accuracy requires larger datasets and fine-tuning the model.

References

(TBA).