Forecasting stock volatility following earnings calls release of SP1500 companies

1st Naman Goenka
Department of Computer Science and Information Systems
Birla Institute of Technology and Science, Pilani
Rajasthan, India
f20180398@pilani.bits-pilani.ac.in

***Abstract*—Improved performance due to multimodal application has led to the same applied in financial domains too. Stock Volatility , defined later, is been increasingly used as a measure of financial risk associated with stock of a company. Earning call of a publicly listed company is considered to be an important event by analysts and investors for deciding their faith in a stock for that firm by referring to discussions made by key-decision makers of the organisation. Here, stock volatility forecast is carried out following an earning call based on multimodal vocal and verbal features from earning call.**

***Index Terms*—vocal cues, multimodal, stock volatility, earnings calls**

I. Summary of approach

A. Problem Statement

Stock Volatility Forecasting is taken as a univariate regression task to predict stock volatility following earning call release of SP1500 companies with τ= 3 days, refering to equation 1 in [4]. Return prices for the next 4 days is needed to estimate average volatility for τ= 3 days. Aligned textual and audio features corresponding to each clip in earning call is available as input for each call given by [4].

B. Data Preprocessing

- A small subset of 120 earning calls is taken for analysis due to computational constraints. Corresponding to a call, based on company code and date of earning call, closing price for 5 **working days** starting from day of call is scraped from Yahoo Finance. return prices for next four days is calculated for calculating stock volatility with τ= 3 days to act as true target value. Process of scraping is prone to errors and datapoints with erroneous labels are dropped off.
- Low level audio features are utilised due to computational constraints, missing values are substituted and data is rounded and standardized.
- Audio Features are padded to the maximum number of clips in earning call for subset of dataset with is found to be 409. Similarly textual features are padded to a smaller length ,128, after being appropriately tokenized since they are considered at a earning call level not at sentence level.

C. Architecture Summary

- State-of-the-art dynamic contextual model BERT is used for encoding textual features. Following bert layer, dense layer is added to fine tune preloaded weights as in any other transfer task.
- Drawing insipration from [1], biLSTM is used for extracting sequential information corresponding to aligned low level audio features, which is a (409 * 29) shaped vector for each earning call. Here, second dimension corresponding to 409 acts as timestamps and third dimension acts as features.
- Textual and vocal encodings are concatenated and passsed through a fully connected network for further processing towards a predicted value of stock volatility.
- As mentioned in deep learning literature, care has been taken to gradually reduce the feature size, and increment it whenever necessary as in case of biLSTM hidden outputs.

II. Experimental Setup and Results

As earning call happen sequentially, data is divided into test and train without shuffling with top 80 examples for train and remaining 30 for test. Mean Squared Loss is used along with Adam optimizer as suggested by univariate regression literature. EarlyStopping callback in keras is utilized to prevent overfitting with patience of 4 epochs and mse being monitored quantity. Finally, model is fitted to training set with 10 % validation for 15 epochs only due to computational constraints.

	Min Train Loss	Min Val Loss	Test Loss
Value	0.774	0.429	0.682
Epoch	13	10	-

III. Conclusions and Inferences

- Initially a problem of exploding gradients was observed, which was corrected by changing activation from Relu to LeakyRelu and standardizing audio features.
- Normalizing stock volatility labels were tried but they deteriorated the performance as suggested by some literature and caused slow convergence of model.
- Test loss is found comparable to training and validation loses,, indicating no overfitting. Slightly higher value of training loss compared to test loss indicates towards mild possibility of underfitting.

- Dropout rate was tuned to 0.15 keeping all dropouts except BERT layer to be same. BERT layer dropout was set to 0.5 suggested value. ADAM performance was found to be getter than RMSprop optimizer.
- Although task at hand was not exploited to the maximum extent computationally but results seen to be close to state-of-the-art studies [1]–[5] in the literature.
- From [2], during the experiment, we find that training with audio data is more prone to overfitting, hence we overcome it by earlystopping and dropout layers.

IV. Future Scope

- Foremost possibility is to exploit task computationally by using larger dataset, traning for more epochs, using better version of BERT model and finally using high-level MFCC audio features dataset of huge size.
- Person encodings available along with dataset can be employed to include only CEO's vocal and verbal cues often encoded as person 1 in analysis. One may try supplying speaker information along with audio features to dataset to determine importance of speaker encoding for the task at hand.

References

[1] Yu Qin and Yi Yang. 2019. What You Say and How You Say It Matters: Predicting Stock Volatility Using Verbal and Vocal Cues. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Association for Computa-tional Linguistics, Florence, Italy, 390–401. https://doi.org/10.18653/v1/P19-1038

[2] Ramit Sawhney, Puneet Mathur, Ayush Mangal, Piyush Khanna, Rajiv Ratn Shah, and Roger Zimmermann. 2020. Multimodal Multi-Task Financial Risk Forecasting. In Proceedings of the 28th ACM International Conference on Multimedia (MM '20), October 12–16, 2020, Seattle, WA, USA. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3394171.3413752

[3] Sawhney, R., Khanna, P., Aggarwal, A., Jain, T., Mathur, P., Shah, R. (2020, November). VolTAGE: Volatility Forecasting via Text-Audio Fusion with Graph Convolution Networks for Earnings Calls. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 8001-8013).

[4] Li, J., Yang, L., Smyth, B., Dong, R. (2020, October). MAEC: A multimodal aligned earnings conference call dataset for financial risk prediction. In Proceedings of the 29th ACM International Conference on Information Knowledge Management (pp. 3063-3070).

[5] Ahmad, N., Zinzalian, A. (2010). Predicting stock volatility from quarterly earnings calls and transcript summaries using text regression. CS224N Final Report.