Improving Hybrid Attention Network (HAN) for Stock Movement Prediction using FinBERT-enhanced Embeddings

Anurag Patel (230173)   Shlok Misar (230653)

Abstract

Financial news plays a crucial role in shaping short-term stock market movements, yet extracting meaningful signals from textual data remains a challenging task. This project investigates the effectiveness of a **Hybrid Attention Network (HAN)** [2] for stock movement prediction and introduces an improved variant that incorporates FinBERT [1], a domain-specific language model pretrained on large-scale financial corpora. The baseline HAN uses Word2Vec embeddings and a hybrid attention structure that combines content-level (news) and temporal attention to aggregate information from multiple days of news. However, static embeddings and simple averaging can miss important contextual nuances.

To address this limitation, we propose **FinBERT-HAN**, a modified architecture that replaces word-level static embeddings with contextual FinBERT [CLS] representations. Both models are trained and evaluated on a curated dataset aligned with stock price movements. Experimental results indicate that FinBERT-HAN consistently improves predictive performance on key metrics, demonstrating the benefit of domain-specific contextual embeddings when combined with hybrid temporal and news-level attention mechanisms. We also discuss computational trade-offs, ablation ideas, and practical considerations for deployment.

1 Introduction

1.1 Motivation

Predicting short-term stock movements using financial news remains a challenging task due to the nuanced and domain-specific nature of financial language. Traditional word embedding methods often fail to capture the subtle semantic and sentiment cues present in market-related text. In contrast, domain-adapted language models such as FinBERT provide richer and more contextually relevant representations, making them well-suited for financial prediction tasks.

1.2 Problem Statement

The objective of this work is to perform a **three-class stock movement classification** task, where the model predicts whether the stock price will increase, decrease, or remain neutral based on associated financial news articles.

1.3 Work Done

1. **Implementation and Evaluation of the Original HAN Model:** We first reproduced and evaluated the baseline **Hybrid Attention Network (HAN)** using the CMIN-US dataset, which contains financial news paired with stock movement labels. This established a reference performance against which improvements could be measured.

2. **Development of the FinBERT-HAN Model:** We then designed an enhanced variant of the model, termed **FinBERT-HAN**. The primary modification involves replacing the original word-level encoder with FinBERT, a transformer-based model pretrained specifically on financial text. Instead of relying on generic embeddings, we utilize FinBERT's `[CLS]` representation as a semantically rich summary of each news sentence, enabling the model to capture financial sentiment and contextual meanings more effectively.

3. **Comparative Analysis:** The performance of the Original HAN and FinBERT-HAN was systematically compared across multiple dimensions:

   - **Accuracy:** Evaluation of predictive performance to assess relative improvements.
   - **Computational Efficiency:** Analysis of training time and resource utilization.
   - **Practical Applicability:** Assessment of model suitability for real-world production environments.

   This comparative study highlights the practical benefits and limitations of integrating domain-specific language models into hierarchical attention frameworks.

2 Dataset [3]

- **Source:** US stocks news and their prices from the cited source.
- **Time Range:** 01-01-2018 to 31-12-2021
- **Splits:**
  - **Train:** 01-01-2018 → 30-04-2021
  - **Dev:** 01-05-2021 → 31-08-2021
  - **Test:** 01-09-2021 → 31-12-2021
- **Stocks Selected:** Apple (AAPL), AbbVie (ABBV), Abbott Laboratories (ABT), Accenture (ACN), Adobe (ADBE), Automatic Data Processing (ADP), Amazon (AMZN), Alphabet – Google (GOOG), JPMorgan Chase (JPM), Morgan Stanley (MS), Microsoft (MSFT), Netflix (NFLX), NVIDIA (NVDA), RTX Corporation (RTX), Tesla (TSLA), Welltower (WELL), Wells Fargo (WFC), Walmart (WMT), Xcel Energy (XEL), Exxon Mobil (XOM).
- **No. of Samples:**
  For 10 stocks:

- **Train:** 3334
- **Dev:** 266
- **Test:** 291

For 20 stocks:

- **Train:** 6934
- **Dev:** 578
- **Test:** 674

- **Labeling:** UP (Price change $\geq$ 0.87%) ; DOWN (Price change $\leq$ -0.41%); PRESERVE(-0.41% $\leq$ Price change $\leq$ 0.87%)

3 Preprocessing

3.1 Original HAN (Hybrid)

- Tokenization using a custom tokenizer (sentence / word-level splitting as needed)
- Word2Vec embedding pretraining using `gensim` on the corpus of news text
- Sequence shape used by the model: (batch, days, news_items, words)
- For each news item we build a fixed-length word-index vector (padding/truncation)

3.2 FinBERT-HAN

- Tokenization via `AutoTokenizer.from_pretrained()` (FinBERT WordPiece tokenizer)
- Input shape per sample: (batch, days, max_tweets, 3, max_tokens) (channels: `input_ids`, `token_type_ids`, `attention_mask`)
- For each news item we precompute / dynamically compute the FinBERT output and take the `[CLS]` embedding (256-d)

4 Model Architectures

4.1 Original HAN (Baseline) – Hybrid Attention Network

Overview: The project's original model is a **Hybrid Attention Network** that can be divided 4 parts:

1. **News embedding:** contains Word2Vec layer which calculates embeddings for each news article.
2. **News-level attention:** assigns importance weights to different news items within the same day (which news matters most).

3. **Sequential modelling:** contains a Bi-directional GRU layer which processes the sequence of daily news representations in both forward and backward directions, capturing past and future contextual dependencies. This produces a richer hidden representation for each day by combining information from preceding and succeeding days.

4. **Temporal attention:** models sequential importance across multiple days (which day in the lookback window matters more).

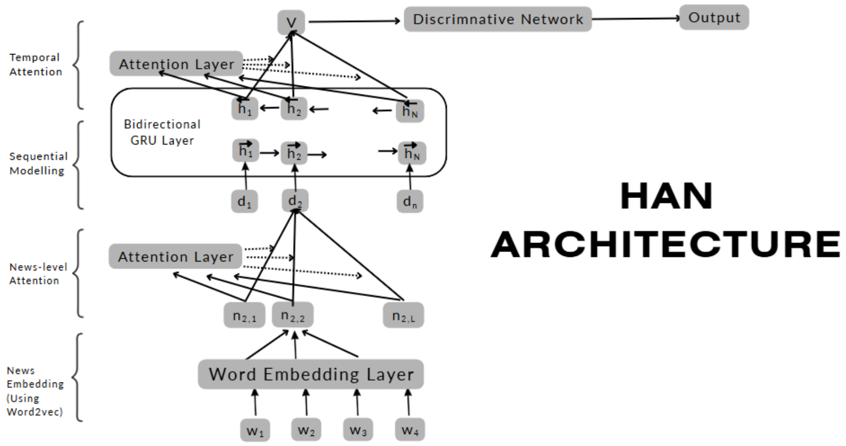

Figure 1: HAN Architecture

4.2 FinBERT-HAN (Modified)

Overview: Here we have replaced the word-embedding layer(Word2Vec), which generates static embeddings, with FinBERT, which gives contextual embeddings for each news article. Rest of the architecture is same as the original HAN. The figure 2 below shows the high level architecture of FinBERT.

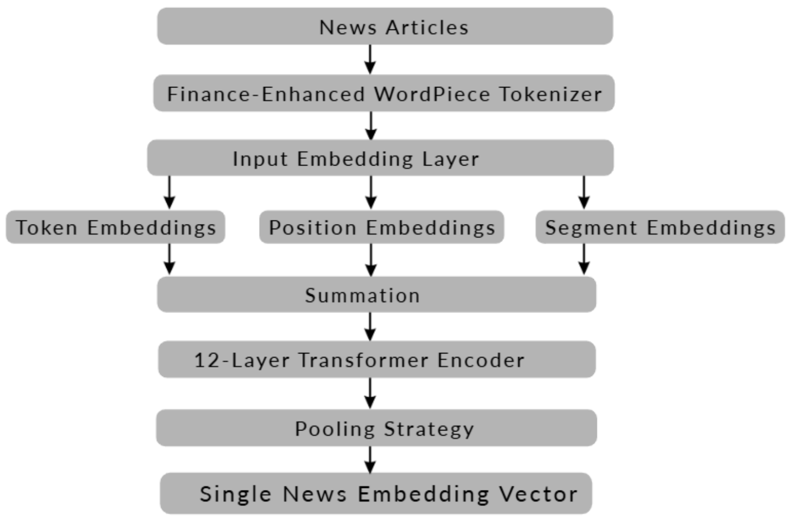

Figure 2: FinBERT

Notes:

- FinBERT outputs contextual embeddings. We use the `[CLS]` vector by default, but mean-pooling over token vectors is an alternative.
- FinBERT may be *frozen* during training to reduce memory and speed up experiments, or *fine-tuned* for potentially better performance.

5 Training Setup and Hyperparameters

Parameter	Original HAN	FinBERT-HAN
Loss Function	CrossEntropy	CrossEntropy
Optimizer	AdamW	AdamW
Learning Rate	1e-4	1e-5
Batch Size	16	16
Epochs	10	10
Dropout	0.3	0.3
Hidden Size	50	256

6 Results

6.1 Training Curves and Accuracy of Original HAN

No. of Stocks: 10
Context Days: 5
Accuracy: 44.44

Figure 3

No. of Stocks: 10
Context Days: 10
Accuracy: 45.6

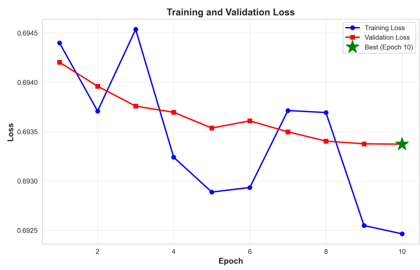

Figure 4

No. of Stocks: 20
Context Days: 5
Accuracy: 45.1

Figure 5

No. of Stocks: 20
Context Days: 10
Accuracy: 46.32

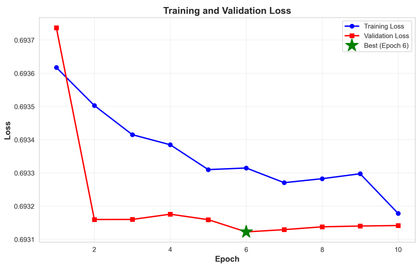

Figure 6

6.2 Training Curves and Accuracy of FinBERT HAN

No. of Stocks: 10
Context Days: 3
Accuracy: 51.7

Figure 7

No. of Stocks: 10
Context Days: 5
Accuracy: 52.83

Figure 8

No. of Stocks: 10
Context Days: 10
Accuracy: 53.2

Figure 9

No. of Stocks: 20
Context Days: 5
Accuracy: 55.2

Figure 10

No. of Stocks: 20
Context Days: 7
Accuracy: 56.0

Figure 11

No. of Stocks: 20
Context Days: 10
Accuracy: 57.1

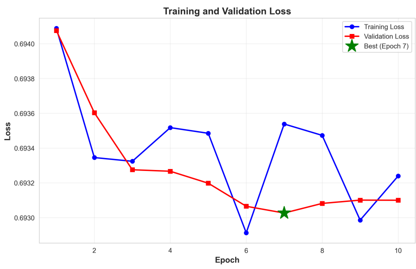

Figure 12

7 Analysis and Discussion

- The experimental results for the Original HAN are consistent with the findings reported in the original paper. Depending on the specific parameter settings, the model achieves an accuracy in the range of 44–46%.

- A slight improvement in accuracy is observed when increasing both the number of stocks and the context window size, indicating that the model benefits from richer temporal and cross-asset information.
- The results clearly demonstrate that the FinBERT-enhanced HAN significantly outperforms the Original HAN, achieving an improvement of almost 7-10%.
- The accuracy imporves when number of context days are increased thus proving that the news are related to each other and may significantly impact each other.

8 Limitations

- Dataset selection and preprocessing biases (sample filtering, thresholding).
- The model is trained only on 20 stocks which is much less used by the author, hence the loss curves are not very stable.
- Using the `[CLS]` token may discard fine-grained token-level information in some cases.
- Computational constraints may limit full end-to-end FinBERT fine-tuning experiments.

9 Conclusion

The modified FinBERT-HAN model outperforms the original HAN primarily because its financial-domain pretrained embeddings enable more precise interpretation of market-related news. By leveraging FinBERT's deep understanding of financial terminology, sentiment, and contextual nuances, the model extracts more informative and discriminative textual features compared to the generic word-level encoder used in the original architecture. Empirically, this enhanced representation capability translates into a significant improvement in prediction accuracy, reinforcing the value of domain-adapted language models for financial forecasting tasks. Overall, the results highlight that integrating specialized semantic encoders such as FinBERT within hierarchical attention frameworks can substantially advance the performance of news-driven stock movement prediction systems.

10 Future Work

- Integrating technical indicators into the HAN framework to complement news-based representations and enable a more comprehensive fusion of market signals.
- Extending the model beyond single-stock news sequences by incorporating relationships between news associated with different stocks, leveraging real-world industrial, sectoral, or supply-chain connections to capture cross-asset dependencies more effectively.

References

[1] Dogu Araci. *FinBERT: Financial Sentiment Analysis with Pre-trained Language Models*. 2019. arXiv: 1908.10063 [cs.CL]. URL: https://arxiv.org/abs/1908.10063.

[2] Ziniu Hu et al. *Listening to Chaotic Whispers: A Deep Learning Framework for News-oriented Stock Trend Prediction*. 2019. arXiv: 1712.02136 [cs.SI]. URL: https://arxiv.org/abs/1712.02136.

[3] Di Luo et al. "Causality-Guided Multi-Memory Interaction Network for Multivariate Stock Price Movement Prediction". In: *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2023, pp. 12164–12176.