Modeling Financial Uncertainty with Multivariate Temporal Entropy-based Curriculums

Ramit Sawhney*
Arnav Wadhwa*
Ayush Mangal2
Vivek Mittal3
Shivam Agarwal4
Rajiv Ratn Shah1
1IIIT Delhi
2Indian Institute of Technology Roorkee
3Indian Institute of Technology Mandi
4Manipal Institute of Technology

Abstract

In the financial realm, profit generation greatly relies on the complicated task of stock prediction. Lately, neural methods have shown success in exploiting stock affecting signals from textual data across news and tweets to forecast stock performance. However, the dynamic, stochastic, and variably influential nature of text and prices makes it difficult to train neural stock trading models, limiting predictive performance and profits. To transcend this limitation, we propose a novel multimodal curriculum learning approach: FinCLASS, which evaluates stock affecting signals via entropy-based heuristics and measures their linguistic and price-based complexities in a time-aware, hierarchical fashion. We show that training financial models can benefit by exposing neural networks to easier examples of stock affecting signals early during the training phase, before introducing samples having more complex linguistic and price-based temporal variations. Through experiments on benchmark English tweets and Chinese financial news spanning two major indexes and four global markets, we show how FinCLASS outperforms state-of-the-art across financial tasks of stock movement prediction, volatility regression, and profit generation. Through ablative and qualitative experiments, we set the case for FinCLASS as a generalizable framework for developing natural language-centric neural models for financial tasks.

1 INTRODUCTION

The ever-growing $60 trillion stock markets around the world present incredibly lucrative opportunities to make monetary profits. However, profit generation heavily relies on financial tasks such as forecasting stock movements and analyzing associated risk, which are complex due to the highly stochastic and dynamic nature of stock prices [Adam et al., 2016]. Prices are greatly influenced by factors such as investor sentiments, opinions about companies expressed across social media platforms (Twitter, Stocktwits), financial news, and much more. The abundance of stock affecting information present on the web helps investors analyze market trends and inspires the adoption of AI-based methods to study the interplay between online textual data and stock prices [Xu and Cohen, 2018, Du and Tanaka-Ishii, 2020].

However, analyzing and modeling the complex unstructured textual data involves numerous challenges. Stock affecting signals across textual sources like financial news and tweets exhibit sequential context dependencies and manifest a variably influential nature on the market [Hu et al., 2018]. Further, online text possesses inherent dynamic timing irregularities, which makes it challenging to model the influence of their temporal and linguistic variations on market trends.

Figure 1: Tweets for GameStop reflect greater linguistic complexity, temporal variations, and context-dependence. In contrast, tweets for Apple are relatively easier to comprehend. Consequently, predicting stock performance from tweets for Apple is easier when compared to GameStop.
The contributions of this study can be summarized as:

- We define a novel multi-modal entropy-based algorithm that measures the linguistic and price-based temporal variations across stock affecting data (text, prices) in a time-aware, hierarchical fashion. The proposed algorithm is flexible and generalizable across numerous tasks involving multi-modal multivariate data.

- We introduce FinCLASS, an end-to-end curriculum learning framework to enhance the training process and practical applicability of language-based neural methods for a variety of real-world financial tasks.

- We define a text-based stock trading model THA-Net, and show how FinCLASS helps THA-Net outperform state-of-the-art baselines on three financial tasks using stock affecting Chinese financial news and English tweets pertaining to stocks in four global markets.

2 RELATED WORK

2.1 FINANCE: STOCK PREDICTION

Conventional Methods  Stock prediction spans various methods, commonly formulated as either regression or classification tasks \cite{Jiang2020}. Conventional stock prediction methods rely on numeric features like historical prices \cite{Kohara1997,Lin2009}, technical indicators \cite{Shynkevich2017}, and macroeconomic indicators \cite{Hoseinzade2019}. These include discrete \cite{Bollerslev1986}, continuous \cite{Andersen2007}, and neural approaches \cite{Feng2019}. Despite their success, a fundamental shortcoming of these methods is that they are limited to numerical features and do not study crucial stock influencing factors across online textual media.

Contemporary Methods  Newer models based on the Efficient Market Hypothesis \cite{Malkiel1989}, leverage natural language features extracted from investor sentiments \cite{Schumaker2009}, public earnings calls \cite{Qin2019}, online news \cite{Hu2018,Du2020} and social media \cite{Xu2018} for stock prediction. Recent work also demonstrates the benefits of modeling stock affecting signals from sources belonging to multiple modalities \cite{Sawhney2020,Zhou2020}. These works show how natural language sources can complement price-based methods in capturing the effect of events like market surprises, mergers and acquisitions over stock returns. However, a major limitation of these methods is that they do not account for the irregularities in release times of stock affecting text \cite{Sawhney2021a,Sawhney2021b}. For trading, the timing plays a critical role, as stock prices rapidly factor all available public market information. Firms may even exploit perception of information \cite{Forbes2009} by timing the release of negative news between positive ones to minimize losses \cite{Segal2016}.

2.2 CURRICULUM LEARNING

The idea of training neural models based on an easy-to-difficult fashion, was initially advanced as curriculum learning (CL) by \cite{Bengio2009}. Since then, CL found numerous applications in computer vision \cite{Pentina2015}, finance \cite{Koenecke2019}, NLP \cite{Xu2020}, reinforcement learning \cite{Narvekar2020}, and more. However, a prime limitation of existing NLP and finance-based CL methods is that they do not define the sample-difficulty by considering the complex linguistic and price-based temporal variations across a sequence of stock influencing modalities in a time-aware, hierarchical fashion. Building on the limitations of existing stock prediction and CL methods, we propose a new multi-modal time-aware hierarchical entropy-based CL framework that enhances the training and profitability of stock prediction models.
3 FINCLASS

**Algorithm 1: Multi-modal Multivariate Hierarchical Time-Aware Entropy-based Stock Complexity \( S_d \)**

**Input:** Text \( X_{\tau-T, \tau-T+1, \ldots, \tau-1} \) and Price Vectors 
\[
P_{\tau-T, \tau-T+1, \ldots, \tau-1} = [P_{\text{open}}, P_{\text{high}}, P_{\text{low}}, P_{\text{close}}] 
\]
during a Lookback Window \( T \), Financial BERT

**Output:** Stock Complexity \( S_d \)

**Step 1:** Extract Financial Sentiment Time-Series

**for each day \( i \) in \( \tau - T \) to \( \tau - 1 \) do**

**for each tweet \( X_{it} \) \( \in X_i \) do**

Get sentiment vector \( p_{it} \)

\[
[p_{\text{bearish}}, p_{\text{bullish}}, p_{\text{neutral}}] 
\]

\( \leftarrow \text{FinBERT}(X_{it}) \)

**end**

**end**

\( E_{\text{intraday}} \leftarrow PE(mpe_{\tau-T}, mpe_{\tau-T+1}, \ldots, mpe_{\tau-1}) \)

**Step 2:** Compute Intraday Multivariate Permutation Entropy \( (MPE) \) [Capturing intraday variations]

**Step 3:** Compute Interday Dynamic Time Warping (DTW) distance [Capturing interday variations]

**for each consecutive day-pair \( (i, j) \) \( \in T \) do**

\[
dtw_{i,j} \leftarrow DTW([p_{i0}, \ldots, p_{iK}], [p_{j0}, \ldots, p_{jK}]) 
\]

**end**

Compute the Permutation Entropy \( (PE) \) of the obtained DTW distances

\( E_{\text{interday}} \leftarrow PE(dtw_{\tau-T, \tau-T+1}, \ldots, dtw_{\tau-2, \tau-1}) \)

**Step 4:** Compute the Multivariate Permutation Entropy of stock price series across days [Capturing interday price variations]

\( E_{\text{price}} \leftarrow MPE(P_{\tau-T}, P_{\tau-T+1}, \ldots, P_{\tau-1}) \)

**Step 5:** Fuse the text and price entropies to obtain the stock complexity \( S_d \)

\( S_d \leftarrow E_{\text{price}} + E_{\text{intraday}} + E_{\text{interday}} \)

**return \( S_d \)**

Consider a dataset where each data sample comprises stock-relevant financial news items or tweets over a lookback of \( T \) days in range \( [\tau - T, \tau - 1] \) to study the performance of stock on day \( \tau \). We define FinCLASS as an approach to re-arrange the dataset according to a learning curriculum based on data-sample complexities to train neural models for financial tasks. To develop a neural model \( \Phi \) for stock prediction, let \( D \) be the set of training examples in a dataset. We compute a difficulty score \( \chi \) corresponding to each training sample \( d_s \in D \) and re-arrange the training dataset based on difficulty ranging from easy to hard. To determine the scores \( \chi \), we compute the stock complexity \( S_d \), and the model complexity \( M_d \) for each data sample as follows.

**Stock-complexity \( S_d \)**

For each data sample \( d_s \in D \), \( S_d \) reflects the linguistic complexity of the stock affecting textual signals and the variations in price signals present in \( d_s \). To capture the stock complexity of a sample, we propose a novel multi-modal multivariate hierarchical time-aware entropy-based algorithm, as presented in Algorithm[1] Entropy indicates the volatility and associated risk based on the complexity of financial time-series \cite{Pincus2004}. For a given lookback \( T \), we exploit the price and text information available corresponding to a stock \( s \). First, we compute the bearish, neutral, and bullish intent of each news item or tweet in the lookback using class probabilities obtained via fine-tuned Financial BERT for English tweets \cite{Araci2019} and Chinese financial text \cite{Rao2021}. We then form a time series corresponding to all three intents separately for each day in \( T \). To measure the three intents’ variations and trends over a day, we compute the Multivariate Permutation Entropy \( (MPE) \) \cite{Morabito2012} of the three time-series of intents, which reflects the linguistic complexity involved in studying financial news or tweets over the day. For multiple channel signals, each time series is usually considered separately while computing the entropy involved. Such a procedure may be acceptable for uncorrelated signals, but the multivariate financial sentiment time series we compute are highly correlated in nature. Thus we use MPE, which effectively captures the cross channel complexities as well. In MPE, the original multivariate time series is transformed into a time dependent matrix from which the relevant statistics and entropies are extracted. For days where the financial texts exhibit greater linguistic and intent variations, MPE would be higher compared to days where texts indicate relatively consistent sentiments towards a stock. To analyze how the temporal evolution of stock-affecting signals varies across days, we adopt Dynamic Time Warping (DTW) distance \cite{Mueller2007} which measures the similarities between time-series of the three intents across consecutive days in a time-aware fashion. For data-samples where financial sentiments across news and tweets show large inter-day variations, DTW distance would be higher, indicating greater difficulty in analyzing stock behavior across days. Next, for each day in the lookback, we form a price vector \( p_i = [p_i^0, p_i^1, p_i^1, p_i^1] \) comprising the stock’s opening, highest, lowest, and adjusted closing prices for trading day \( i \). Financial research indicates that irregular price trends lead to increased difficulties and complexities in predicting future stock performance \cite{Adam2016}. To this end, we calculate the MPE of all price vectors \( p_i \) to measure the irregularities in the time-series of historic prices across days. We then fuse the MPE of sentiment trends for each day, the DTW distance of sentiment trends for stock's opening, highest, lowest, and adjusted closing prices across a Lookback Window \( T \), and re-arrange the training dataset based on difficulty ranging from easy to hard. To determine the scores \( \chi \), we compute the stock complexity \( S_d \), and the model complexity \( M_d \) for each data sample as follows.
Figure 2: A high-level overview of FinCLASS sample difficulty score computation. Input: Data sample comprising stock-relevant financial news or tweets, and prices over a lookback period of $T$ days; Output: sample difficulty score $\chi$.

and the MPE of price features to obtain stock-complexity $S_d$, as shown in Figure 2. Given that entropy is an additive quantity, we add them to find $S_d$. We experiment with other fusion types such as dot product and weighted sum, but could find no significant improvement over addition.

**Model-complexity $M_d$**

For a sample $d_s \in D$, $M_d$ reflects the difficulty faced by a neural model in accurately mapping $d_s$ to the ground truth given a financial task. We compute the model-complexity for each sample $d_s \in D$ based on the confidence score (movement classification) and regression (volatility) scores obtained from the model $\Phi$. To compute $M_d$, we first scatter the training set $D$ into $B$ uniform meta-datasets, each having $1/B$ of the total samples in $D$. We then train $2 \times B$ identical copies of the model $\Phi$, two over each meta-dataset, for price movement classification and volatility regression, respectively. For each example $d_k$ belonging to meta-dataset $b$, we compute the classification (softmax probabilities for price movement) and regression (predicted volatilities) scores for corresponding tasks from all copies of the model except the ones which were trained on meta-dataset $b$. We then identify the class $Y$ predicted by a majority of the classification copy-models for sample $d_k$, and compute the average $p_{Y,avg}$ of the confidence scores generated for class $Y$ by each classification copy-model. Next, we compute the average $\hat{p}_{avg}$ of magnitude of difference between the true $v_\tau$ and the predicted $\hat{v}_\tau$ volatilities for sample $d_k$ as obtained by each regression copy-model. We then compute the model-complexity as $M_d = p_{avg} + 1/p_{avg}$. Lastly, for each training sample $d_s \in D$, we define and compute the **overall difficulty score** $\chi$ as:

$$\chi = \alpha M_d + (1 - \alpha) S_d$$

where $\alpha$ is a learnable parameter for the two complexities.

**Curriculum definition**

We re-arrange the training dataset based on difficulty scores $\chi$ and expose the model to more difficult samples with each consecutive epoch. Formally, we first sort the training samples $d_s$ by their overall difficulty scores $\chi$ and divide the scores into $B$ equal sized ranges. Next, we arrange each sample $d_s \in D$ into the $B$ groups, based on their difficulty scores. After this arrangement, we train the stock prediction model $\Phi$, for ten epochs each on samples from every group. For the first epoch group, we train the model on the data-groups having in-decreasingly difficulty samples. After $10 \times B$ epochs, when the model has seen all the ordered samples, we expose it to the original data distribution in entire training set $D$, and train it until convergence. **Note:** Different curriculums can be arranged based on difficulties computed across combinations of price, text, and model complexities.

### 4 NEURAL STOCK TRADING

We derive inspiration from [Hu et al., 2018], [Sawhney et al., 2021a], and define a model that takes as input stock relevant texts in lookback $T \in [\tau - T, \tau - 1]$, and predicts the price movement or volatility for day $\tau$. While any trading model can be used, we propose **THA-Net:** Time-aware Hierarchical Attention Network for these financial tasks.

**Intra-Day Encoder**

First, THA-Net encodes the texts (news or tweets) $t$ for stock $s$ released in a day via an embedding layer: BERT [Devlin et al., 2019] as $m = \text{BERT}(t) \in \mathbb{R}^d, d=768$, by averaging the final outputs from BERT.
For each stock $s_i$ on a day $i$, a variable number (K) of tweets (t) are posted at irregular times (k). Studying a sequence of tweets over the day provides a more unified context to understand a stock, as compared to a single tweet alone [Barber and Odean 2007]. LSTMs are a natural way to capture such sequential context dependencies over time. However, LSTMs assume inputs to be equally spaced in time whereas the time interval between release of consecutive news or tweets can vary widely, from a few seconds to many hours, which can have a drastic impact on their influence on the market [O’Hara, 2015].

To this end, we use a time-aware LSTM (TLSTM) [Baytas et al. 2017] to model stock-relevant texts over a day. We feed the time between texts to the TLSTM to capture the temporal irregularities in their release times. We encode the news, and tweets for a stock $s_i$ on a day $i$ using the TLSTM as:

$$h_t = \text{TLSTM}(m_t, \Delta k, h_{t-1}); t \in [1, K]$$

where $h_t$ represents the hidden state for text $t$. Owing to the variably influential nature of news and tweets, we use an attention mechanism [Luong et al. 2015] to emphasize texts that have a higher impact on the stock, as shown in Equation 3. This attention (intra-day attention) learns to aggregate the hidden states of the TLSTM into an intra-day vector $x_i$:

$$x_i = \sum_{t} \gamma_i h_t; \gamma_i = \frac{\exp(W_1(W_2 h_t + W_3 h_K))}{\sum_{t=1}^{K} \exp(W_1(W_2 h_t + W_3 h_K))}$$

where $\gamma_i$ denotes the attention weights, and $W_1$, $W_2$, and $W_3$ are learnable network parameters.

**Inter-Day Encoder** We combine the representations learned from texts in each day across multiple days in a lookback in a hierarchical fashion using the sequence of intra-day vectors $x_i$. We feed the vectors $x_i$ to an LSTM:

$$h_i = \text{LSTM}(x_i, h_{i-1}); \tau - T \leq i \leq \tau - 1$$

where $h_i$ is the hidden state representation for day $i$. Further, tweets and news published across different days have shown to have a varying impact on stock prices, due to financial phenomena such as calendar anomalies, the week-day effect, etc. Thus, to selectively weigh critical days, we employ an inter-day attention mechanism to aggregate all days into an overall representation $z_i$. The inter-day and the intra-day attention together form a hierarchical attention, allowing the model to emphasize crucial textual signals within and across days in the lookback in an attentive fashion. We now define the following financial tasks to evaluate the benefits of using FinCLASS to train THA-Net.

**Stock Movement Classification** We define the price movement of stock $s_i$ from day $\tau - 1$ to $\tau$ as:

$$Y_{\tau} = \begin{cases} 0, & p_{\tau}^c < p_{\tau-1}^c \\ 1, & p_{\tau}^c \geq p_{\tau-1}^c \end{cases}$$

where $p_{\tau}^c$ is the closing price of the stock on day $\tau$, and 0 and 1 denote price downfall and rise. We feed $z_i$ to a feed-forward layer followed by softmax that outputs the predicted price movement $\hat{Y}_\tau$ for the stock $s_i$ on day $\tau$.

**Stock Volatility Regression** Next, we define the single day log volatility using the daily log of absolute returns as:

$$v_{\tau} = \ln \left( \frac{p_{\tau} - p_{\tau-1}}{p_{\tau-1}} \right)$$

We input the representation $z_i$ to a feed-forward layer followed by a linear activation that outputs the predicted volatility $\hat{v}_{\tau}$ for stock $s_i$ on day $\tau$.

**Stock Trading Strategy** To assess the profitability of THA-Net, we propose a rule-based trading strategy defined using the predicted movement and volatility for stock $s_i$ on day $\tau$. We execute trades only if THA-Net classifies the price movement with a confidence higher than 70% and the predicted volatility lies within the 1st standard deviation from the mean of true volatilities across the training dataset. Note that true volatility is computed by parsing the ground truth closing price values of the data in the training set through Equation 6.

5 EXPERIMENTS AND SETUP

5.1 PUBLIC DATASETS

**US S&P 500** [Xu and Cohen 2018] Comprises 109,915 English tweets from social media platform Twitter spanning January 2014 to December 2015, related to 88 high-trade-volume-stocks from the NASDAQ stock exchange forming the S&P 500 index. [Xu and Cohen 2018] extract stock specific tweets using regex queries made of stock tickers (e.g., $AAPL$ for Apple, where $ is a cashtag on Twitter).

**China & Hong Kong** [Huang et al. 2018] Comprises 90,361 financial news headlines in Chinese spanning January to December 2015, aggregated by Wind related to 85 top China A-shares stocks in Shanghai, Shenzhen and Hong Kong Exchanges. [Huang et al. 2018] extract corporate news from major Chinese financial websites. We extract historic stock prices from Yahoo Finance for both the datasets.

**Preprocessing** We align trading days by dropping data samples that do not possess any news or tweets in the 5-day (T) window. We split the US S&P 500 dataset temporally based on date ranges from January 01, 2014 to July 31, 2015 for training, August 01, 2015 to September 30,
We contrast the performance of THA-Net and FinCLASS against the following baselines on the public datasets.

We adopt grid search to find optimal hyperparameters based on the validation MCC and MSE (\(5.4\)) for all classification and regression models, respectively. We explore the hidden states for both TLSTM and LSTM \(d \in [64, 128, 256]\), and find the best performance at \(d = 128\) for both encoders. We divide the training dataset in 30 groups, and present the performance variation across different values of \(B\) in \(6.5\).

We use a learning rate of \(1e-4\) and a decay rate of \(1e-5\) to train the models using the Adam [Kingma and Ba, 2014] optimizer for 500 epochs. We further elucidate on THA-Net and FinCLASS training setup in the supplementary material.

5.2 THA-NET AND FINCLASS TRAINING SETUP

We conduct all experiments on an NVIDIA Tesla T4 GPU. We adopt grid search to find optimal hyperparameters based on the validation MCC and MSE (\(5.4\)) for all classification and regression models, respectively. We explore the hidden states for both TLSTM and LSTM \(d \in [64, 128, 256]\), and find the best performance at \(d = 128\) for both encoders. We divide the training dataset in 30 groups, and present the performance variation across different values of \(B\) in \(6.5\).

We use a learning rate of \(1e-4\) and a decay rate of \(1e-5\) to train the models using the Adam [Kingma and Ba, 2014] optimizer for 500 epochs. We further elucidate on THA-Net and FinCLASS training setup in the supplementary material.

5.3 BASELINE APPROACHES

We contrast the performance of THA-Net and FinCLASS against the following baselines on the public datasets [5.1]:

- **WLSTM**: LSTMs with stacked autoencoders that encode noise-free data obtained through wavelet transform of prices [Bao et al., 2017].
- **RandForest**: Random Forest classifiers trained over text embeddings obtained using word2vec [Mikolov et al., 2013].
- **TSLDA**: Topic Sentiment Latent Dirichlet Allocation – a generative model that uses sentiments and topics in text [Nguyen and Shirai, 2015].
- **CH-RNN**: An RNN-based model with cross-modal attention on price movement trends and texts across days [Wu et al., 2018].
- **SN-HFA**: StockNet - HedgeFundAnalyst - a variational autoencoder with attention on text and prices in the lookback period [Xu and Cohen, 2018].
- **SN-DA**: StockNet - DiscriminativeAnalyst - a StockNet model that directly optimizes the log likelihood objective [Xu and Cohen, 2018].
- **Chaotic**: A hierarchical attention network that uses Gated Recurrent Units with attention across words, texts and days [Hu et al., 2018].

- **Adv-LSTM**: An Adversarial LSTM-based model which leverages adversarial training to improve the training process [Feng et al., 2019].
- **StockEmb**: Stock embeddings acquired using prices, and dual vector (word-level and context-level vectors) representation of texts [Du and Tanaka-Ishii, 2020].
- **FAST**: A BERT-based hierarchical time-aware encoder for financial text using hierarchical attention while modeling stocks together [Sawhney et al., 2021a].

5.4 EVALUATION METRICS

**Classification** We use Accuracy and Matthew’s Correlation Coefficient (MCC) for evaluating the stock movement prediction performance. MCC avoids potential bias due to data skew as it does not depend on the choice of the positive class and also accounts for the true negatives. For a given confusion matrix \(\begin{pmatrix} tp & fn \\ fp & tn \end{pmatrix}\):

\[
MCC = \frac{tp \times tn - fp \times fn}{\sqrt{(tp + fp)(tp + fn)(tn + fp)(tn + fn)}},
\]

**Regression** To evaluate the volatility regression performance, we adopt the Mean Squared Error (MSE) to compute the error between the actual and the predicted volatilities.

**Profit** To evaluate the practical applicability of our approach to real-world trading, we assess profitability using two metrics: Sharpe Ratio (SR) [Sharpe, 1994] and Maximum Drawdown (MDD). Note that we trade one unit of each stock independently. The Sharpe ratio is a measure of the return of a portfolio compared to its risk. We calculate the Sharpe ratio by computing the ratio of the expected return \(\mu\) of a portfolio to its standard deviation as: \(SR = \frac{\mu}{\sigma}\). The Maximum Drawdown measures the maximum loss from a peak \(r_m\) to a trough \(r_d\) of a portfolio (in terms of returns), and is defined as: \(MDD = \left(\frac{r_m - r_d}{r_m}\right) \times 100\). Larger values of MDD indicate potentially risky trades.

6 RESULTS AND DISCUSSION

6.1 PERFORMANCE COMPARISON

Table 2 shows the superior stock movement classification and volatility regression performance of THA-Net and FinCLASS over baseline and proposed methods. In general, methods that study stock affecting information across text, classify stock movements and predict stock volatility more accurately than methods that only exploit historical prices. These improvements re-validate the effectiveness of leveraging textual sources to capture stock affecting signals such as market surprises, announcements, and public sentiment. We attribute the higher performance of THA-Net + FinCLASS.
We now study how THA-Net’s stock classification and volatility regression performance benefits via curriculums defined using model-based and multi-modal entropy-based complexity heuristics. Table 1 presents the influence of factoring price-based, text-based, and both multi-modal (price and text-based) stock complexities ($S_d$), and the model complexity ($M_d$) to compute the difficulty scores $\chi$ for arranging a curriculum to train THA-Net. The difficulty score as an intrinsic property of a training example is sometimes best decided by the neural model itself [Xu et al., 2020], as is indicated by the performance improvement observed for the task of volatility regression in the case of US S&P 500. However, as stock prices and textual signals are highly stochastic in nature, model training benefits from a potent blend of hand-crafted heuristics to better analyze sample difficulty. THA-Net enjoys more significant performance gains across all tasks and datasets when trained over price and text-based curriculums compared to model-based curriculum.

However, we observe that the price curriculum typically leads to lower relative performance gains over the base model compared to the text curriculum. This difference arises as textual stock affecting signals reflect more complex linguistic and temporal variations, especially around events such as the release of quarterly earnings calls, mergers and acquisitions, breaking news, etc. Such variations are likely to be more apparent in the case of the China & Hong Kong data, which comprises chaotic stock-relevant news from the period of 2015-16 China Stock Market Turbulence [Liu et al., 2016]. Consequently, a text curriculum is more likely to introduce the difficult textual samples to THA-Net at later stages of training. We observe optimum performance when using difficulty scores obtained via a blend of price, text, and model complexities to define the curriculum. We attribute these improvements to the entropy-based multi-
We now study the performance improvements obtained via FinCLASS to real-world stock trading by analyzing the risk-adjusted returns (Sharpe ratio) and the maximum risk (Maximum Drawdown) associated with the trades executed using THA-Net across stocks in China A-shares and the S&P 500 indexes. We first train THA-Net for stock trading without curriculum learning, and observe poor performance in terms of profits and a higher risk over both indexes as shown in Table 3. This observation indicates that THA-Net takes riskier trading decisions and often experiences losses of large magnitude. However, when we train THA-Net using FinCLASS, we observe significant improvements in risk-adjusted returns (US S&P 500: 168.5%, China A-Shares: 447.3%) and a strong reduction in maximum losses (US S&P 500: 20.7%, China A-Shares: 2.8%). Such improvements indicate the efficacy of FinCLASS in enhancing the real-world applicability of neural stock prediction methods. We further elucidate on the benefits of FinCLASS via a qualitative study on the stocks in the China & Hong Kong dataset.

### 6.3 PROFIT ANALYSIS

We examine the practical applicability of FinCLASS to real-world stock trading by analyzing the risk-adjusted returns (Sharpe ratio) and the maximum risk (Maximum Drawdown) associated with the trades executed using THA-Net across stocks in China A-shares and the S&P 500 indexes. We first train THA-Net for stock trading without curriculum learning, and observe poor performance in terms of profits and a higher risk over both indexes as shown in Table 3. This observation indicates that THA-Net takes riskier trading decisions and often experiences losses of large magnitude. However, when we train THA-Net using FinCLASS, we observe significant improvements in risk-adjusted returns (US S&P 500: 168.5%, China A-Shares: 447.3%) and a strong reduction in maximum losses (US S&P 500: 20.7%, China A-Shares: 2.8%). Such improvements indicate the efficacy of FinCLASS in enhancing the real-world applicability of neural stock prediction methods. We further elucidate on the benefits of FinCLASS via a qualitative study on the stocks in the China & Hong Kong dataset.

### 6.4 ANALYZING STOCK COMPLEXITY

We now study the performance improvements obtained via FinCLASS over non-curriculum THA-Net against samples of varying difficulty levels $\chi$. In Table 4, we divide the dataset into three buckets of low, medium, and high sample difficulty according to the stock complexity $S_d$. We observe significant improvements over all three difficulty levels, demonstrating that FinCLASS improves performance across both financial tasks over data having varying levels of complexity. Interestingly, we see a corresponding increase in relative improvement for volatility regression as the stock complexity increases from low to medium to high, demonstrating that FinCLASS incurs greater performance gains on increasingly difficult samples, which are otherwise harder to learn for THA-Net. In Figure 3, we divide our training data into 10 fine-grained groups according to the stock complexity $S_d$, and show a heat-map of relative improvement via FinCLASS over non-curriculum THA-Net across 10 stocks. We observe improvements across a diverse range of stocks where FinCLASS improves training performance across a wide range of sample complexities. The heat-map shows some interesting trends; for instance, there is a significant performance increase for high entropy samples compared to low entropy ones. Such trends may arise as non-curriculum approaches may better learn easy samples only, but FinCLASS also enables THA-Net to better learn samples having complex linguistic and price-based temporal variations along with easy ones to bolster stock prediction.

### 6.5 PARAMETER SENSITIVITY

In Figure 4, we study the influence of the parameter $B$ on the performance of FinCLASS. The value of $B$ directly impacts the number of meta-datasets formed, granularity of the model difficulty score, and the number of learning stages in the curriculum. We gradually vary $B$ from smaller to larger values in the range $B \in [2, 20]$, as shown in Figure 4. Initially, we observe poor performance across both tasks as for lower values of $B$, the copy-models are likely to inherit a bias towards only one type of difficulty [Xu et al., 2020]. This bias occurs as the sample distribution in meta-datasets would remain identical to the original dataset for lower values of $B$. On the other hand, larger values of $B$ lead to smaller meta-datasets, which may individually not possess enough information (variations in stock-affecting signals) to train robust copy-models. Further, we observe that middle-ranged values ($B = 10$) give the best performance across both tasks and datasets. Note that FinCLASS is robust to the value of $B$, owing to the competitive stock prediction
Figure 4: The performance of THA-Net + FinCLASS for stock movement prediction and volatility regression tasks across both the datasets, as the number of meta-datasets \( B \) is varied in the range of 2 to 20.

performance against non-curriculum THA-Net for a breath of variations in the value of \( B \).

6.6 QUALITATIVE ANALYSIS

We now conduct an extended study to elucidate the benefits of FinCLASS for stock prediction on the US S&P 500, as shown in Figure 5. Tweets about Apple in training samples A and B possess a sarcastic tone, making it hard to analyze their plausible influence on the stock. Without a curriculum, such linguistically challenging samples would appear before the easier sample (C) due to the chronology of the training dataset, likely making it harder for THA-Net to learn stock affecting signals across text to predict movements accurately. We observe that FinCLASS correctly assigns a higher difficulty score to samples A and B while defining the learning curriculum and presents them after the less complicated sample C to train THA-Net, allowing it to learn stock affecting trends via easier samples first. This paradigm later makes it easy for THA-Net to learn stock-affecting information across difficult samples, thus maximizing performance. Note that predictions improve across both training and testing datasets when using the curriculum generated via FinCLASS to train THA-Net. Lastly, we show that for a moderately complex test-data sample, movement trend is wrongly classified when training THA-Net without the curriculum, but when trained using FinCLASS, the trend is classified accurately. We attribute THA-Net+FinCLASS’s overall improved performance to the generated curriculum that ameliorates the efficiency of the learning process.

7 CONCLUSION

We present FinCLASS, a curriculum learning framework to enhance the training of price-based and language-based neural models for financial tasks. FinCLASS defines an optimum learning curriculum using model-based and hand-crafted entropy-based multi-modular sample difficulty heuristics that reflect the linguistic and price-based temporal complexities of training samples in datasets. FinCLASS is generalizable to a variety of language-based tasks involving a sequence of textual and multi-modal data, embeddings, or feature representations. Experiments on benchmark English tweets and Chinese financial news headlines related to stocks in the S&P 500 and the China A-Shares indexes demonstrate the efficacy of FinCLASS to enhance the performance of neural quantitative trading methods. FinCLASS helps our neural trading model THA-Net outperform state-of-the-art baselines and increases its practical applicability across multiple real-world financial settings.

References


Fuli Feng, Huimin Chen, Xiangnan He, Ji Ding, Maosong Sun, and Tat-Seng Chua. Enhancing stock movement prediction with adversarial training, 2019.


Minh-Thang Luong, Hieu Pham, and Christopher D. Manning. Effective approaches to attention-based neural machine translation, 2015.


