



MANA-Net: Mitigating Aggregated Sentiment Homogenization with News Weighting for Enhanced Market Prediction

Mengyu Wang
The University of Edinburgh
Edinburgh, UK
mengyu.wang@ed.ac.uk

Tiejun Ma
The University of Edinburgh
Edinburgh, UK
tiejun.ma@ed.ac.uk

Abstract

It is widely acknowledged that extracting market sentiments from news data benefits market predictions. However, existing methods of using financial sentiments remain simplistic, relying on equal-weight and static aggregation to manage sentiments from multiple news items. This leads to a critical issue termed “Aggregated Sentiment Homogenization”, which has been explored through our analysis of a large financial news dataset from industry practice. This phenomenon occurs when aggregating numerous sentiments, causing representations to converge towards the mean values of sentiment distributions and thereby smoothing out unique and important information. Consequently, the aggregated sentiment representations lose much predictive value of news data. To address this problem, we introduce the Market Attention-weighted News Aggregation Network (MANA-Net), a novel method that leverages a dynamic market-news attention mechanism to aggregate news sentiments for market prediction. MANA-Net learns the relevance of news sentiments to price changes and assigns varying weights to individual news items. By integrating the news aggregation step into the networks for market prediction, MANA-Net allows for trainable sentiment representations that are optimized directly for prediction. We evaluate MANA-Net using the S&P 500 and NASDAQ 100 indices, along with financial news spanning from 2003 to 2018. Experimental results demonstrate that MANA-Net outperforms various recent market prediction methods, enhancing Profit & Loss by 1.1% and the daily Sharpe ratio by 0.252.

CCS Concepts

• Information systems → Sentiment analysis; • Applied computing → Economics; Forecasting.

Keywords

Market Prediction, Sentiment Analysis, News Aggregation

ACM Reference Format:

Mengyu Wang and Tiejun Ma. 2024. MANA-Net: Mitigating Aggregated Sentiment Homogenization with News Weighting for Enhanced Market Prediction. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM '24)*, October 21–25, 2024, Boise, ID, USA. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3627673.3679653>



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike International 4.0 License.

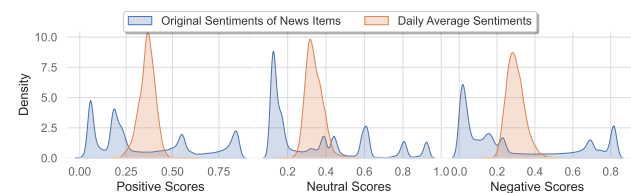
CIKM '24, October 21–25, 2024, Boise, ID, USA
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0436-9/24/10
<https://doi.org/10.1145/3627673.3679653>

1 Introduction

Many studies have documented the significant influence of news on financial markets [4, 18]. News sentiment analysis has emerged as a prominent research area, aiming to capture the emotional tone of news articles and their potential impact on market movements [39]. Unlike general sentiment analysis, financial news sentiment analysis deals with specialized vocabulary and requires consideration of market context [43]. Advancements in Natural Language Processing (NLP) have led to the development of tools for financial sentiment extraction, including dictionary-based methods [11], deep learning methods [2, 58], and recent large language models (LLMs) [38].

Despite the diverse methodologies for financial news sentiment analysis, the utilization of sentiments from multiple news items remains simplistic. Recent news-based stock prediction works, including various multi-modal models [34, 45, 56], and LLMs [38], primarily focus on prediction methods themselves. These methods commonly employ aggregation techniques like sentiment averaging, polarity counting, or custom sentiment indicators to derive variables representing overall market sentiments. However, these aggregation approaches are equal-weight and static, meaning they treat all news as equally influential and consistently generate fixed aggregated representations. This fails to capture the dynamic nature of market sentiments, where news articles exert varying degrees of influence depending on market conditions [6, 7, 48]. Additionally, market efficiency theory suggests that key news information is reflected in market prices [14, 64]. However, static aggregation does not account for the interactions between prices and news. Consequently, these equal-weight and static methods may not fully extract the predictive power available in news data, especially when dealing with large volumes of news.

Figure 1: The kernel density estimate plot illustrates the distributions of sentiment scores for both individual news items and their daily average values.



To illustrate this critical research gap, we analyzed an extensive financial news dataset spanning from 2003 to 2018, which comprises over 2.7 million news items, with up to 1953 news items per day. Each news item was scored for its potential positive, neutral, and negative impacts on market dynamics. Our investigation reveals a concerning phenomenon we term **Aggregated Sentiment Homogenization**, which results in weak news sentiment

representations. This phenomenon arises when aggregating market sentiments from a large volume of daily news, causing the aggregated sentiment scores to converge towards the average values of the individual sentiment distributions. As a result, the representations of aggregated news sentiments from different days become much more similar (homogenized) compared to the original sentiments of individual news items, thereby smoothing out unique and crucial information. Figure 1 visually presents an overview of this phenomenon, highlighting the notable disparity between the original sentiment scores and the daily average scores. The average score distribution exhibits significantly higher concentration, indicating the loss of information through homogenization. Detailed analysis of this challenge is presented in Section 4

Utilizing homogenized aggregated news sentiment representations for prediction suffers from two limitations. Firstly, given the diverse impact of news on the market, the aggregated scores mask influential news with extreme sentiments and loss crucial predictive information. Secondly, the similarity in average scores across different days homogenizes daily news sentiments, thereby failing to provide effective and distinguishing information for the prediction model. While we use sentiment averaging as an example here, aggregated sentiment homogenization is a broader issue prevalent in many existing aggregation methods. Because it stems from the equal treatment of various news items and static aggregation without considering market changes.

To address these challenges and extract more potential of financial news for prediction, this paper introduces the Market Attention-weighted News Aggregation Network (MANA-Net). MANA-Net innovates by implementing a dynamic market-news attention mechanism that assigns weights to news items based on their correlation with market prices. This approach allows for a nuanced aggregation of news sentiments, emphasizing items with greater predictive relevance. Consequently, MANA-Net can process extensive news of varying volumes and effectively convert it into a unified representation that captures the most predictive sentiment information.

Furthermore, MANA-Net integrates news aggregation and market prediction into a unified model. This holistic approach enables all model components to dynamically adjust during the training period through gradient back-propagation based on prediction loss. This process allows MANA-Net to optimize the aggregated news sentiment representations by considering price changes and prediction requirements. Therefore, MANA-Net offers an enhanced approach to utilizing market sentiments for market prediction.

In summary, this paper provides three key contributions:

- (1) We explore a critical issue inherent in the prevalent equal-weight and static news aggregation methods: aggregated sentiment homogenization. Through analysis of a large financial news dataset, we demonstrate these methods obscure valuable news insights and negatively impact market predictions.
- (2) We introduce MANA-Net, a market prediction model that employs a novel weighted news aggregation approach. MANA-Net mitigates the effects of sentiment homogenization and provides a method to weight news sentiments for market prediction.
- (3) MANA-Net demonstrates superior market prediction performance compared to existing approaches, including recent GPT-based methods. Our model achieves an improvement of 1.1% in Profit & Loss and a 0.252 increase in daily Sharpe ratio.

2 Related Works

2.1 Financial news sentiment analysis

The impact of news on the market has been studied for decades [53, 57]. Lexicon-based approaches, utilizing financial keywords dictionaries to extract sentiments, have provided foundational insights and investment guidance [11, 16, 39]. However, these methods struggle with complex sentence structures and semantics [43]. This has led to a shift towards machine learning for text extraction [1, 42, 51]. In addition, deep learning techniques have further enhanced sentiment analysis performance [21, 49, 62]. Techniques like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have significantly improved sentiment extraction accuracy [13, 30]. More recently, language models offer nuanced approaches to analyzing financial news sentiment [2]. Advanced Large Language Models (LLMs) also develop this topic significantly [38]. Despite these advances in extracting market sentiments from news, the challenge of aggregating sentiments from voluminous news data remains underexplored, particularly in addressing the different impact of individual news items on market predictions.

2.2 News-based market prediction

While traditional market prediction models rely on price series data [19, 29, 59], incorporating news data has been proved to provide more valuable insights [39]. However, existing studies often employ static and equal-weight aggregation strategies for news sentiments, failing to consider the dynamic nature of markets and the varying impact of individual news items. Early works like sentiment categorization or treating news as a single corpus [26, 33, 35] suffered from limited accuracy due to neglecting the complexities of news language and context. Later advancements like knowledge graphs [9] and opinion extraction [41] aimed to extract more information from news data, but still lacked the ability to differentiate news importance. Recent models attempt to combine news and price data, but often resort to simple aggregation techniques like polarity counting [45, 55], sentiments averaging [56], or special sentiment indicators designing [27, 41]. This tendency toward simplistic aggregation techniques creates a significant challenge of data homogenization, a problem observed across various domains [5, 50] but underexplored within the context of news-based market prediction. As thousands of market-related news items become readily accessible on a daily basis [10], the inadequacy of current models in effectively harnessing this wealth of information becomes evident, presenting an opportunity for our proposed method, MANA-Net, to enhance market prediction.

2.3 Attention Mechanism

The introduction of the attention mechanism in machine translation [3] has revolutionized the ability of models to prioritize crucial information across different domains, including computer vision [66] and market selection [47]. In essence, the attention mechanism assigns weights to different elements, allowing the model to focus on the most crucial information. This is commonly achieved by comparing a “query” vector with a set of “key” vectors to compute attention scores. The output is then a weighted sum of corresponding “value” vectors. This approach, often referred

to as self-attention [52], allows elements within a sequence (e.g., words in a sentence) to attend to each other. The transformer architecture [52], which heavily relies on self-attention, has achieved success in market prediction by extracting insights from price sequences [37, 54, 63]. However, these models primarily focus on enhancing knowledge extraction from price data. MANA-Net breaks new ground by introducing a novel market-news attention mechanism. Unlike existing applications in market prediction, this mechanism focuses on dynamically estimating weights for each news item. These weights reflect the projected impact of each news item on market volatility. By assigning greater importance to news with a higher anticipated influence on the market, MANA-Net optimizes market prediction through more effective news sentiment utilization. This approach showcases a unique application of attention that extends beyond traditional sequence analysis and allows for the dynamic integration of financial news sentiments.

3 Preliminaries

3.1 Problem Formulation

The primary goal of this study is to predict the daily market price trends by analyzing daily market prices and news sentiments. For clarity, we introduce the following notations. The trading day is denoted by d , and the predicted daily trend, y_d , is represented as a binary outcome where 1 signifies a price increase and 0 a not increase. The vector \mathbf{p}_d encapsulates fundamental price indicators for the day, with \mathbf{p}_d^c specifically denoting the close price. The prediction target is formalized as follows:

$$y_d = \mathbb{I}(\mathbf{p}_{d+1}^c > \mathbf{p}_d^c) \quad (1)$$

The set of news sentiments for day d is represented by $S_d = \{s_{d,1}, s_{d,2}, \dots, s_{d,N_d}\}$, where $s_{d,i}$ indicates the sentiments of the i -th news item, and N_d is the total number of news items on day d . The function $f(\cdot, \cdot)$ represents the trend prediction model, which integrates market prices and news sentiments of the previous t days ($t \geq 1$) as inputs:

$$\hat{y}_d = f((\mathbf{p}_d, \mathbf{p}_{d-1}, \dots, \mathbf{p}_{d-t+1}), (S_d, S_{d-1}, \dots, S_{d-t+1})) \quad (2)$$

Our objective is to effectively learn and optimize the function $f(\cdot, \cdot)$.

3.2 Data

This research utilizes market index datasets: the Standard and Poor’s 500 (S&P 500) Index dataset and the NASDAQ 100 Index dataset, along with the Thomson Reuters News Analytics (TRNA) dataset comprising over 2.7 million news items. Our utilized news dataset is notably larger than those employed in previous studies, with a news volume at least 10 times greater than that of their datasets [27, 38, 45]. Moreover, in contrast to their online-collected news data, our dataset is sourced from industry practice, ensuring higher quality and more comprehensive market information.

Market Index Datasets: The S&P 500 Index tracks the market performance of 500 leading publicly traded companies in the United States and serves as a critical benchmark for the financial market [17, 31]. Its comprehensive company coverage makes it a pivotal index for evaluating market trends and forecasting. The NASDAQ 100 Index consists of 100 of the largest, most actively traded non-financial companies listed on the Nasdaq Stock Market. It represents

the top companies outside of the financial sector and is widely used as a benchmark dataset for stock prediction [12, 20]. We collected daily S&P 500 and NASDAQ 100 prices from 2003 to 2018, including six basic features: open price, close price, adjusted close price, high price, low price, and trading volume. Descriptive statistics for the two index datasets are detailed in Table 1.

Table 1: Statistics of Market Index Datasets

Variable	S&P 500 Index			NASDAQ 100 Index		
	Mean	Std Dev	Median	Mean	Std Dev	Median
Open	1559.8	531.3	1367.3	2908.9	1686.8	2197.0
Close	1560.1	531.1	1367.6	2910.2	1687.6	2199.1
High	1568.3	532.0	1374.8	2931.9	1696.9	2216.9
Low	1550.8	530.2	1360.8	2885.0	1675.4	2167.3
Adj Close	1560.1	531.1	1367.6	2910.2	1687.6	2199.1
Volume	3.41×10^9	1.38×10^9	3.40×10^9	2.83×10^5	1.31×10^5	2.68×10^5

News Dataset: Our utilized TRNA datasets spans from 2003 to 2018. It tracks over 25,000 equities and nearly 40 commodities and energy topics, making it a comprehensive news source for studying financial markets [44]. All news items are annotated with market sentiment scores by the news provider with an industry’s advanced news-analytics. Each news item is assigned three sentiment scores, which indicate the likelihood that the news item will have a positive, neutral, or negative impact on the mentioned instrument. We denote $\mathbf{s}_{d,i} = (pos_{d,i}, neu_{d,i}, neg_{d,i})$, where $pos_{d,i}$, $neu_{d,i}$, and $neg_{d,i}$ represent the positive, neutral, and negative scores of the i -th news on day d . The sum of these three scores for the same news equals one. Statistics of TRNA dataset is illustrated in the next section.

4 Sentiment Aggregation Challenges

Our investigation into the TRNA dataset has revealed a fundamental challenge inherent to the prevailing equal-weight and static approaches to news sentiment aggregation: **Aggregated Sentiment Homogenization**. In essence, assigning equal weight to all news items regardless of their content or potential market impact homogenizes the overall sentiment representations. Furthermore, the static nature of these methods prevents the aggregated representations from considering price data and adapting to different market conditions. Consequently, diverse sentiment knowledge embedded within individual news items are erased, making the aggregated sentiments less informative for detailed market analysis.

There are various widely used news sentiment aggregation representations such as Count Features, Sentiment Factors and Average Features [27, 38, 56], which will be introduced in Section 6.2. Existing aggregated representations are primarily based on two core indicators, the average sentiment scores and the news polarity counts. Both of them can lead to the homogenization issue, affecting all related aggregation methods. To demonstrate the prevalence of this challenge beyond our specific news sentiments from TRNA dataset, we collected additional sentiment scores using FinBERT [2], a widely recognized area-specific language model proficient in financial sentiment analysis. We generated FinBERT scores for our news. Table 2 presents statistics of two sentiment sets for analysis.

Statistics of individual news items reveal that sentiment scores are widely scattered, rendering overall statistics less representative. Considering the 0-1 scoring range, the high standard deviations (around 0.25 and 0.15 for both datasets) indicate significant

Table 2: News sentiment statistics from both TRNA dataset and FinBERT. The 'IQR' refers to the interquartile range. We analyze these statistics at two levels: individual news items and daily aggregated sentiment, including average-based and count-based values. For average aggregation, we calculate the daily average sentiment scores to represent the overall daily sentiments. Count aggregation calculates the proportion of news items belonging to each sentiment category to represent daily sentiments.

Category	Sentiments from TRNA Dataset						Sentiments from FinBERT					
	Mean	Std Dev	Median	IQR	Skewness	Kurtosis	Mean	Std Dev	Median	IQR	Skewness	Kurtosis
Individual News Items												
Positive	0.360	0.268	0.239	0.394	0.570	-1.066	0.334	0.135	0.301	0.153	3.881	1.639
Neutral	0.343	0.248	0.216	0.405	0.908	-0.386	0.366	0.194	0.411	0.186	-1.553	0.213
Negative	0.297	0.295	0.165	0.575	0.682	-1.153	0.301	0.182	0.257	0.155	2.212	0.540
Daily Average Values												
Positive	0.364	0.043	0.366	0.054	-0.293	0.493	0.334	0.023	0.330	0.027	2.256	18.602
Neutral	0.338	0.043	0.333	0.059	0.610	0.733	0.368	0.032	0.371	0.041	-0.698	1.324
Negative	0.298	0.047	0.293	0.062	0.603	0.541	0.299	0.028	0.297	0.034	0.423	1.921
Daily Count Ratios												
Positive	0.397	0.073	0.400	0.092	-0.216	0.245	0.119	0.048	0.112	0.060	1.258	5.009
Neutral	0.273	0.076	0.264	0.101	0.628	0.674	0.659	0.072	0.664	0.098	-0.417	1.204
Negative	0.330	0.071	0.324	0.094	0.526	0.433	0.222	0.063	0.219	0.084	0.331	2.024

dispersion. Furthermore, the substantial gaps between mean and median values, along with large interquartile ranges (IQR), support a relatively even distribution across a broad range. Skewness and kurtosis provide further insights into the distribution shapes. TRNA sentiment scores exhibit skewness and significant platykurticity. FinBERT scores, while less platykurtic, still demonstrate a strong skew. These characteristics – high dispersion, wide ranges, and skewed distributions – pose challenges for existing aggregation methods that rely heavily on average scores or counts.

Our analysis of aggregated sentiment scores from both TRNA and FinBERT highlights a concerning trend: homogenization. After aggregation, the standard deviation values for daily sentiment on both datasets drop below 0.1, with most values even less than 0.05. This represents a decrease of more than 80% compared to the standard deviation of individual news item scores. Similarly, the difference between mean and median values, along with the interquartile range (IQR), all shrink significantly. These observations suggest that the aggregated sentiment scores become highly concentrated. Kurtosis values further support this finding. Compared to the original sentiment scores, the kurtosis values of the aggregated sentiment scores increase significantly. This indicates that the distributions become peaked and more concentrated around the average value. These observations suggest that the aggregated

sentiment scores become highly concentrated, losing the variability present in the original data and providing less distinguished representations for the following prediction models.

To complement our statistical analysis, we have included the distribution figure in Section 1. In addition, we provide a boxplot (Figure 2) to offer a visual representation of our findings. We randomly selected 30 days from the dataset and drew a boxplot of the positive news scores for these days. The figure clearly illustrates the homogenization effect of aggregation. As observed, all daily average scores (represented by green triangles) fall within a narrow value range (highlighted by the blue rectangle). In contrast, the boxplot for individual news item scores encompasses a much wider range, visually demonstrating the greater variability in sentiment before aggregation.

In summary, our investigation exposes a fundamental challenge in current sentiment aggregation methods: Aggregated Sentiment Homogenization. This issue obscures the valuable diversity of sentiments within individual news articles, leading to the generation of uninformative aggregated sentiment representations.

5 Methodology

MANA-Net tackles the challenge of aggregated sentiment homogenization by incorporating a trainable news aggregation process within the prediction model itself. It takes both daily market prices and sentiments of individual news item as inputs. A core component of MANA-Net is a market-news attention mechanism. This mechanism dynamically evaluates the relevance between each news item and the daily market prices, assigning news weights that determine the impact of each news item on the prediction outputs. These importance scores are then used to weight individual sentiment scores during aggregation. By integrating news aggregation and market prediction into a unified trainable model, MANA-Net ensures that the aggregation process adapts to the prediction objective, mitigating sentiment homogenization and generating a more informative representation for superior market prediction performance. Figure 3 presents an overview of MANA-Net.

Figure 2: The boxplot of positive sentiment scores for individual news items across 30 randomly sampled days. Green triangles represent the daily average scores, while red lines indicate the median values. The blue rectangle highlights the range of averaged sentiments.

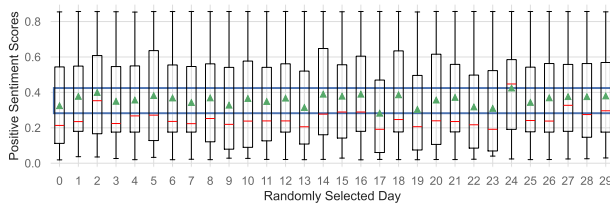
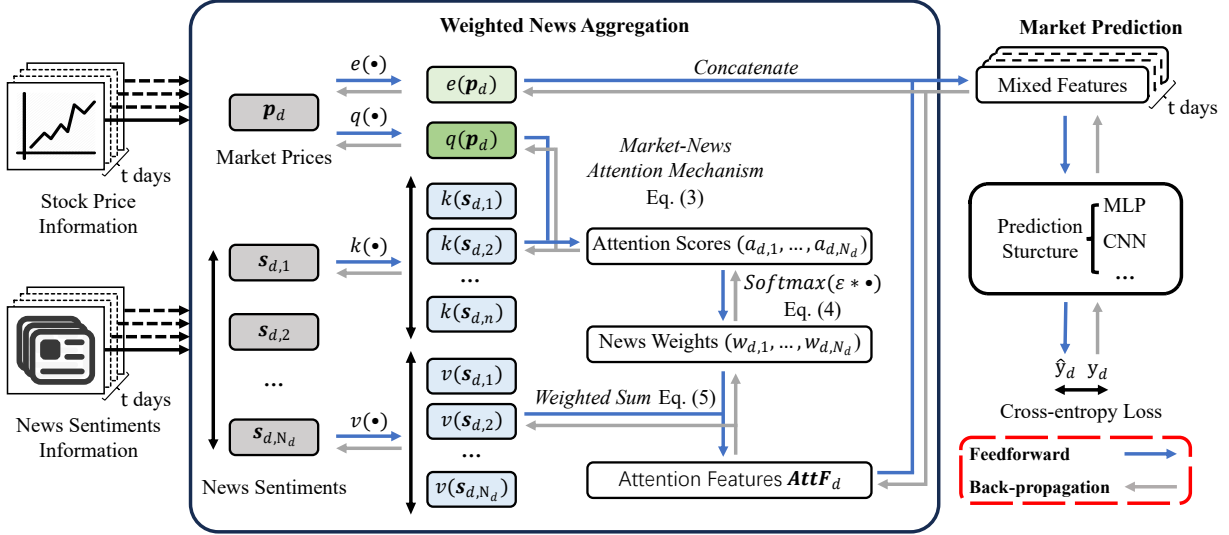


Figure 3: An overview of MANA-Net. The market prices and news sentiments data are as defined in Section 3.1. Blue arrows show the feed-forward process and grey arrows show the back-propagation process.



5.1 Weighted News Aggregation

MANA-Net’s weighted news aggregation component is specifically designed to address the issue of aggregated sentiment homogenization. It meets three key requirements for effective news sentiment aggregation: (i) quantifying the impact of news on the market based on price changes, (ii) generating more effective aggregated representations incorporating the market impact of news, and (iii) accommodating fluctuating news volumes.

Firstly, to quantify the impact of news, we design a market-news attention mechanism. Traditional self-attention mechanisms focus on a sequence’s internal information to weight its content [52]. In financial areas, such mechanisms are used for price series analysis by attending to prices at different points [12]. MANA-Net innovates by applying a split-attention approach to distinct input information. Specifically, we prepare functions for query, key, and value vectors, denoted as $q(\cdot)$, $k(\cdot)$, and $v(\cdot)$, which encodes inputs into the corresponding features. Unlike traditional methods, these functions are applied to different data sources. As shown in Figure 3, $q(\cdot)$ operates on market price data p_d , while $k(\cdot)$ is used on news sentiment data ($s_{d,1}, s_{d,2}, \dots, s_{d,N_d}$). We then employ scaled dot-product attention [52] to calculate attention scores $a_{d,i}$ as:

$$a_{d,i} = \frac{1}{\sqrt{d_k}} q(p_d)^T \cdot k(s_{d,i}) \quad (3)$$

where d_k denotes the key vector dimension, using for scaling the products of $q(p_d)$ and $k(s_{d,i})$ to prevent exploding gradients during training. This design leverages the core principle of attention mechanisms: measuring relevance between query and key vectors to weight key-corresponding data. In MANA-Net, this separation allows news information to be evaluated based on market prices. The resulting attention scores fulfill the first goal by quantifying news impact based on price changes.

Secondly, we design a difference enlargement normalization to translate attention scores into news weights. Large daily news

volumes can lead to less differentiated values when directly applying the softmax function to attention scores. To address this and accelerate training convergence, MANA-Net introduces a difference enlargement factor ϵ ($\epsilon \geq 1$) within the softmax function. We normalize attention scores to news aggregation weights $W_d = (w_{d,1}, w_{d,2}, \dots, w_{d,N_d})$ as:

$$W_d = \text{Softmax}(\epsilon \cdot (a_{d,1}, a_{d,2}, \dots, a_{d,N_d})) \quad (4)$$

This factor emphasizes the relative importance of news items by enhancing the distinction between news weights while preserving their rankings. The factor ϵ is tuned as a hyperparameter during training. The resulting weights W_d determine the contribution of each news item to the aggregated representation. They are used in a weighted sum over the value vectors $v(s_{d,i})$, generating a unified vector called **Attention Features (AttF)**.

$$\text{AttF}_d = \sum_i^{N_d} w_{d,i} v(s_{d,i}) \quad (5)$$

This AttF_d vector captures the sentiment information from all news on day d , while weighting their contributions based on market impact. To this end, we achieve our second goal of generating more effective aggregated representations for market prediction.

Thirdly, upon reviewing the process of MANA-Net’s weighted news aggregation, it becomes evident that there are no constraints on news volume. Equations (3), (4), and (5) operate with any value of N_d . The market-news attention mechanism and the weighted news aggregation operations are adaptable to varying news volumes. There is no need for time-consuming and potentially biased process of news selection or filtering, common in approaches that struggle with fluctuating news volumes [23, 33]. Consequently, the third goal of accommodating fluctuating news volume is achieved.

In summary, MANA-Net’s market-news attention mechanism assesses the importance of news to the market, and its weighted news aggregation emphasizes highly influential news. This process effectively addresses the challenges of sentiment homogenization

by ensuring that critical news items with extreme positive or negative sentiment scores are not simply smoothed over. Capturing these extremes allows the model to gain a nuanced understanding of market sentiment, crucial for effective market prediction.

5.2 Market Prediction and Training

MANA-Net further addresses the issue of sentiment homogenization by an integrated market prediction model that incorporates our news aggregation component. Existing works typically treat aggregated news sentiments as a static input [36, 38, 41]. Therefore, they are limited by unchangeable news representations, which has been homogenized due to the equal-weight aggregation. Unlike these methods, MANA-Net creates a trainable news aggregation process directly connected to the prediction model. This holistic approach treats the aggregation and prediction components as a unified system, taking both prices and original news sentiments as inputs to output market predictions.

Specifically, once the aggregated news sentiment vector, AttF_d , is obtained, it is concatenated with $e(\mathbf{p}_d)$, a distinct price feature vector designed for prediction. The function $e(\cdot)$ here is an encoding function different from $q(\cdot)$, allowing for potentially different encoding strategies for different data utilization. The concatenated vector $\mathbf{m}_d = [e(\mathbf{p}_d), \text{AttF}_d]$ serves as the input for the prediction model. If the prediction leverages multi-day data (previous t days, where $t > 1$), we create a sequence of features $(\mathbf{m}_d, \mathbf{m}_{d-1}, \dots, \mathbf{m}_{d-t+1})$ for each day within the chosen range. The prediction model then utilizes a cross-entropy loss function, commonly used for classification tasks, to quantify prediction error for training.

This seamless integration of news aggregation with prediction facilitates the optimization of both components in alignment with prediction objectives. Unlike static aggregation methods where backpropagation terminates at the aggregated representations, MANA-Net allows gradients to flow through the entire process. This enables the model to learn from prediction errors and adjust the weights throughout the news aggregation pipeline. The backpropagation process can be illustrated by the following equation :

$$\frac{\partial y}{\partial s_{d,i}} = \frac{\partial y}{\partial \text{AttF}_d} \cdot \frac{\partial \text{AttF}_d}{\partial s_{d,i}} \quad (6)$$

This equation delineates how the error signal (represented by the partial derivative of the loss function with respect to the prediction y) is propagated backwards. The dynamic adjustment allows the model to modify the news aggregation process based on prediction errors, channeling gradients from the aggregated representations AttF_d back to individual news sentiments $s_{d,i}$. Such a mechanism ensures that the news feature vector continuously evolves to encapsulate the most relevant information for market prediction. Figure 3 visually depicts the training process workflow, highlighting the forward pass (blue arrows) and backward pass (grey arrows) of data and gradients. This integrated training approach ensures that all model components, including the news aggregation mechanism, are optimized for better market prediction.

MANA-Net is adaptable and can be used with various neural network architectures for prediction. We evaluate its performance using five popular structures: shallow network (SN), Multi-Layer Perceptron (MLP) [8], Convolutional Neural Network (CNN) [32], Long Short-Term Memory (LSTM) model, and Transformer (Trans)

structure [52]. These architectures, widely acknowledged in market prediction literature [25, 27], are evaluated through time series cross-validation to identify the optimal structure for MANA-Net. For SN, MLP, and CNN models, the mixed features are concatenated into a single vector before being fed into the prediction model. LSTM and Transformer structures, designed for sequential data, take the vectors $(\mathbf{m}_{d-t+1}, \mathbf{m}_{d-t+2}, \dots, \mathbf{m}_d)$ as input in sequential order. A detailed analysis of the prediction structures will be presented in Section 7.1.

6 Experiment Settings

We conducted experiments using both 2080Ti and A100 GPUs. Training a model requires 28 hours on a 2080Ti GPU or 6 hours on an A100 GPU. Trained models run for over a month in our tests, requiring only 9.45 seconds for a single prediction. Given the common usage of A100 and superior GPUs in the industry, our computing requirements are practical for real-world applications.

6.1 Evaluation Metrics

We evaluate our model based on three metrics: **Accuracy**, **Profit & Loss (PnL)** and **Sharpe Ratio (SR)**. They are widely used in market prediction literature [60, 61]. Among these three metrics, SR is considered the most important one as it incorporates both returns and risk, which are crucial factors in market investment.

PnL quantifies the cumulative profit or loss experienced by a portfolio during a designated time frame. We calculate the PnL for each trading day and sum up the daily PnL values across the entire test sets. The PnL for all forecasts spanning D days can be expressed as:

$$\text{PnL} = \sum_{d=1}^D \text{flag}_d \cdot \frac{\mathbf{p}_{d+1}^c - \mathbf{p}_d^c}{\mathbf{p}_d^c}, \quad (7)$$

where $\text{flag}_d = 1$ if $\hat{y}_d = y_d$ and $\text{flag}_d = -1$ otherwise.

Sharpe Ratio measures the investment performance in relation to a risk-free asset. To compute the daily Sharpe Ratio within the test sets, we employ the following formula:

$$\text{SR} = \left(\frac{1}{n} \sum_{d=1}^D \text{flag}_d \frac{\mathbf{p}_{d+1}^c - \mathbf{p}_d^c}{\mathbf{p}_d^c} - R_f \right) / \sigma(R), \quad (8)$$

where R_f represents the return rate of an investment with zero risks, meaning that it is the return that investors could expect for taking no risk, such as a Treasury bond investment. And $\sigma(R)$ denotes the standard deviation of the excess return of the asset. In our settings, we use a value of 0.02 for R_f , which corresponds to the average US Treasury rate during our data collection period. Furthermore, we also explored different risk-free rate values between 0.00 and 0.05. MANA-Net exhibited robustness to this parameter, consistently achieving superior performance compared to baseline methods.

6.2 Baseline Methods

To assess MANA-Net's news aggregation capabilities, we compare it with five popular news aggregation methods:

(1) **Count Features (CF)**: This common method calculates ratios of news categories' counts to represent overall sentiment (e.g., positive news articles / total news articles) [38, 56].

- (2) **Sentiment Factor (SenF)**: Proposed by Jing et al. [27], SenF is a sentiment factor for LSTM model to predict the market trends. SenF is calculated as: $SenF_d = (num_d^{pos} - num_d^{neg}) / num_d^{pos+neg}$, where num_d^* is the volume of news belonging to class * on day d .
- (3) **Sum Features (SumF)**: This method simply sums the daily news sentiment scores to create a single sentiment vectors [45].
- (4) **Average Features (AF)**: Similar to SumF, AF aggregates sentiment by averaging daily individual news sentiment scores [56].
- (5) **Frequency Aggregation Features (FAF)**: Huynh et al. [24] proposed FAF, which assigns weights based on news appearance frequency, considering the potential impact of news repetition.

When comparing these methods to MANA-Net, we implemented these approaches by substituting our news aggregation model and utilizing their news representations instead of our **AttF**. The inputs and prediction structures remain the same to ensure a fair comparison. Additionally, since these methods' original prediction structures are included in our prediction structure candidates, they also serve as market prediction baselines compared to MANA-Net.

In addition to news aggregation, we evaluate MANA-Net's overall market prediction performance against three advanced methods:

- (6) **Ensemble Deep Learning Model** [36]: This model integrates sentiment analysis with a two-level ensemble architecture, known for its effectiveness in capturing complex time series patterns and combining textual and numerical data.
- (7) **Frozen Pretrained Transformer (FPT)** [65]: FPT leverages a pre-trained Transformer model, demonstrating state-of-the-art performance in various time series tasks.
- (8) **GPT-based Methods** [38]: This recent work explores the potential of LLMs like GPT-3.5 and GPT-4 for market prediction using news headlines. It highlights the growing interest in leveraging these powerful language models for financial forecasting.

These market prediction methods covers the advanced time-series models, ensemble models, and large language models. Therefore, they represent the state-of-the-art methods of market prediction.

6.3 Training and Validation

We assess the performance of all methods using data from the preceding t days. Our experimental design involves varying values for t , including 1, 3, 5, 10, and 20. To ensure a rigorous and comprehensive assessment, we adopt a time series cross-validation approach to determine the optimal hyper-parameters. The dataset is partitioned into 10 sliding windows, each encompassing 500 consecutive days of data, arranged chronologically. The initial window starts from the first available day, with subsequent windows commencing 391 days later than their predecessor. Within each window, we allocate the data into training, validation, and testing sets following an 8:1:1 ratio, respectively. The presented results are averages from the test sets across all 10 windows, ensuring stability and reliability across different temporal segments of the dataset.

7 Results and Discussions

In our experiments, using sentiments derived from TRNA datasets generally yields better market prediction results compared to sentiments obtained from other methods like FinBERT. Therefore, our discussions are primarily based on results with TRNA sentiments.

7.1 Prediction model Selection

To determine the most effective prediction structure for MANA-Net, we maintain the weighted news aggregation model constant while experimenting with candidate prediction structures. The results are summarized in Table 3. Notably, the MLP emerges as the top-performing structure across most tested settings in terms of SR and PnL, establishing itself as the most suitable choice. Despite their renown for handling time-series data, both the LSTM model and the Transformer structure do not outperform the other models.

Table 3: MANA-Net's prediction results with different prediction structures. Refer to Section 5.2 for model abbreviations.

Time Length	Metrics	Model Architectures				
		SN	MLP	CNN	LSTM	Trans
1-day	Acc	0.535	0.537	0.518	0.529	0.534
	PnL	0.022	0.031	0.017	0.007	0.020
	SR	0.499	1.110	0.576	0.382	0.599
3-day	Acc	0.535	0.541	0.543	0.529	0.507
	PnL	0.004	0.037	0.053	0.011	0.032
	SR	0.080	0.943	1.565	0.557	0.692
5-day	Acc	0.525	0.542	0.507	0.530	0.511
	PnL	0.024	0.041	0.005	0.008	0.031
	SR	0.180	1.331	-0.028	0.416	0.831
10-day	Acc	0.563	0.552	0.516	0.530	0.503
	PnL	0.045	0.054	0.004	0.008	0.029
	SR	1.317	1.624	-0.192	0.416	0.563
20-day	Acc	0.550	0.528	0.526	0.503	0.509
	PnL	0.023	0.045	0.029	0.029	0.033
	SR	0.834	1.533	0.608	0.597	0.747

This outcome can be attributed to the complexity of the model. The LSTM model has been criticized for its intricate structure, characterized by multiple gates and parameters lacking immediate interpretability or clear purpose [28]. This complexity poses challenges to the training process. In the context of MANA-Net, which already incorporates a sophisticated news aggregation structure, a highly complex prediction network could potentially hinder overall performance. Our validation results support this notion, as the most effective LSTM configuration involved using a single layer and one-directional model, the simplest form. Similarly, the inherent complexity of Transformer does not align well with MANA-Net. The best validation results obtained from a transformer with 2 attention layers, whereas general transformers typically employ 6/12 layers. Thus, we require a prediction structure with an optimal level of complexity that complements our weighted news aggregation.

Additionally, while LSTM and Transformer are renowned for their ability to capture sequential information, other structures can also leverage temporal dependencies. Several studies have successfully applied MLP models to time series prediction tasks, showcasing their capacity to use time-dependent message from input data [15, 40, 46]. Moreover, when provided with high-quality features, MLP models are frequently employed as subsequent layers due to their simplicity [22, 46]. Given the concatenated features, the MLP model can inherently learn the significance of time order if it is critical for the prediction. Therefore, the MLP model, with

Table 4: Market prediction results of prevalent aggregation methods and advanced prediction methods. Abbreviations for these methods are detailed in Section 6.2. The “Price Only” lines serve as the control group, relying solely on prices for predictions. Results for the GPT-4 on the NASDAQ 100 dataset are exclusive due to its less competitive results on S&P 500 dataset and its limited availability compared to GPT-3.5. Underlined values highlight the highest PnL and SR achieved across all settings.

Method	1-day			3-day			5-day			10-day			20-day		
	Acc	PnL	SR	Acc	PnL	SR	Acc	PnL	SR	Acc	PnL	SR	Acc	PnL	SR
Results of S&P 500 Market															
Price Only	0.535	0.026	0.870	0.512	0.030	0.886	0.525	0.024	0.592	0.531	0.029	0.820	0.546	0.026	0.686
CF	0.498	0.024	0.619	0.533	0.008	0.309	0.536	0.015	-0.051	0.530	0.021	0.621	0.526	0.024	0.915
SenF	0.539	0.021	0.767	0.496	0.010	-0.164	0.554	0.030	1.083	0.525	0.038	1.013	0.554	0.045	1.085
SumF	0.512	-0.022	-0.637	0.514	0.019	0.276	0.503	-0.019	-0.409	0.501	0.024	0.814	0.513	0.036	0.853
AF	0.535	0.017	0.515	0.541	0.037	0.943	0.513	-0.004	-0.098	0.509	0.026	0.732	0.515	0.023	0.712
FAF	0.518	0.018	0.389	0.534	0.031	0.845	0.515	-0.016	-0.716	0.511	0.012	0.339	0.503	0.029	0.597
MANA-Net	0.537	0.031	1.110	0.518	0.029	0.978	0.542	0.041	1.331	0.552	0.056	1.624	0.528	0.045	1.533
Ensemble	0.522	0.018	0.886	0.514	0.031	0.672	0.510	0.018	0.824	0.534	0.031	1.343	0.545	0.028	1.028
FPT	0.508	0.002	0.323	0.524	0.033	0.917	0.520	0.042	1.169	0.528	0.025	0.969	0.547	0.035	1.273
GPT-3.5	0.522	0.022	0.914	0.530	0.030	1.133	0.528	0.022	0.946	0.522	0.014	0.765	0.524	0.014	0.769
GPT-4	0.504	0.032	0.696	0.520	0.008	0.194	0.516	0.008	-0.104	0.484	0.003	-0.244	0.494	0.005	-0.109
Results of NASDAQ 100 Market															
Price Only	0.512	-0.021	-0.834	0.498	-0.057	-1.537	0.517	-0.010	-0.476	0.557	0.024	0.950	0.517	-0.008	0.016
CF	0.548	0.010	0.711	0.482	-0.045	-1.227	0.527	-0.003	0.136	0.533	-0.005	0.231	0.507	-0.023	-0.475
SenF	0.528	0.024	0.150	0.510	-0.020	-0.257	0.569	0.023	0.937	0.505	-0.035	-0.594	0.517	-0.016	-0.381
SumF	0.536	0.010	0.570	0.514	-0.019	-0.217	0.545	0.006	0.588	0.527	0.009	0.457	0.540	0.014	0.883
AF	0.504	-0.035	-0.984	0.536	-0.013	0.059	0.521	0.008	0.306	0.513	-0.026	-0.516	0.491	-0.024	-0.884
FAF	0.552	-0.002	0.401	0.536	-0.034	-0.425	0.539	-0.016	-0.101	0.525	-0.016	-0.120	0.527	-0.006	0.059
MANA-Net	0.528	0.004	0.592	0.550	0.023	0.891	0.549	0.020	0.901	0.561	0.031	1.017	0.538	0.041	1.347
Ensemble	0.540	-0.006	0.107	0.534	0.007	0.628	0.546	0.010	0.627	0.532	0.020	0.985	0.542	0.024	1.095
FPT	0.536	0.009	0.523	0.537	0.003	0.457	0.549	0.009	0.766	0.540	0.017	0.837	0.506	0.003	0.322
GPT-3.5	0.538	0.010	0.795	0.522	0.014	0.882	0.526	0.015	0.692	0.530	0.017	0.896	0.522	0.009	0.639

its appropriate level of complexity, appears to be the most suitable prediction structure for our MANA-Net framework.

7.2 Prediction Results and Analysis

We assess MANA-Net against various news aggregation methods and contemporary market prediction models, illustrating its effectiveness in harnessing news for market predictions. Comparative results, shown in Table 4, use the ‘Price Only’ method as a benchmark to demonstrate outcomes without integrating news data. Abbreviations for baseline methods are consistent with those introduced in Section 6.2. Results for the GPT-4 on the NASDAQ 100 dataset are exclusive due to its less competitive results on S&P 500 dataset and its limited availability compared to GPT-3.5.

7.2.1 Comparison with Aggregation Methods. Our evaluation of MANA-Net against existing news aggregation methods (presented in the upper side of MANA-Net in Table 4) reveals three key findings that highlight MANA-Net’s more effective news utilization. (i) None of the baseline methods consistently outperform the “Price Only” strategy, indicating that inappropriate news aggregation methods may not only fail to improve but could even negatively affect the prediction. (ii) No single baseline method demonstrates superior performance across all scenarios, highlighting a general lack of robustness. (iii) MANA-Net demonstrates consistent improvement, outperforming the control group in all settings and surpassing baselines. It achieves the best performance in most scenarios and

achieves the highest PnL and SR across all settings on both datasets. This consistent improvement highlights MANA-Net’s effectiveness in utilizing news for prediction. Notably, at the optimal settings for each methods, MANA-Net outperforms baselines by at least 1.1% in PnL and 0.539 in SR for the S&P 500 market, and 1.6% in PnL and 0.397 in SR for the NASDAQ 100 market.

The comparison with existing news aggregation methods underscores the effectiveness of MANA-Net’s weighted news aggregation. The key difference lies in how news aggregated representations are obtained. Most existing methods, except for “FA”, treat news articles equally and use static representations for prediction, making them susceptible to sentiment homogenization issue. The “FAF” method partially addresses this issue by considering news frequency as a proxy for news importance, but it still relies on static representations. MANA-Net’s superior performance over these methods highlights its more effective news aggregation in overcoming sentiment homogenization (see Section 7.3 for detailed analysis).

7.2.2 Comparison with Advanced Market Prediction Methods. MANA-Net’s strength in leveraging news knowledge positions it as a frontrunner among contemporary market prediction models. It surpasses Ensemble and FPT methods, which also exhibit commendable and stable performance, particularly when incorporating multi-day data. MANA-Net achieves a PnL improvement of 1.4% and an SR increase of 0.281 on the S&P 500 market and a 1.7% PnL gain and a 0.252 SR boost on the NASDAQ 100 market compared

Table 5: News Cases with Weights Exceeding 0.98. This table lists specific news items that received weights greater than 0.98 after normalizing daily weights to a 0-1 scale, along with their sentiment scores. The sentiment scores are presented in the format of (positive scores, neutral score, negative score).

Date	News Title	News Sentiments	Daily Avg. Sentiments
03/10/2018	Reuters Insider - Trading at Noon: A look at what is powering Dow Industrials to another record	(0.851, 0.118, 0.031)	(0.373, 0.302, 0.325)
05/10/2018	Macron to campaign for tougher anti-monopoly rules in EU elections	(0.057, 0.125, 0.819)	(0.393, 0.325, 0.282)
12/10/2018	Saudis face business backlash over missing journalist	(0.056, 0.126, 0.818)	(0.347, 0.320, 0.332)
16/10/2018	Pompeo meets Saudi king on Khashoggi case, Turks study "toxic materials"	(0.058, 0.134, 0.808)	(0.375, 0.309, 0.317)

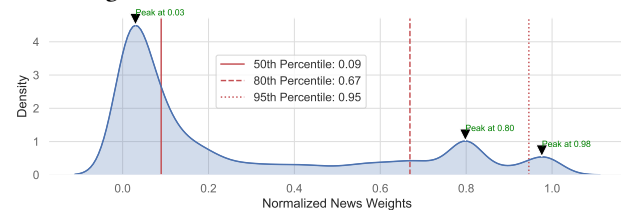
to these models. The advanced GPT-based methods underperform in our settings, primarily due to the significantly larger size of our news dataset compared to those used in their original development. GPT-based methods, which rely on polarity counts for news aggregation, suffer from the sentiment homogenization issues. For instance, GPT-3.5 categorizes a substantial portion (55.38%) of news articles in the S&P 500 dataset as neutral, with 29.25% classified as positive and 15.36% as negative, leading to highly homogenized aggregated news sentiments that are predominantly neutral. Similarly, GPT-4 classifies 66.78% of news as neutral, further exacerbating the homogenization issue. The NASDAQ 100 dataset displays a similar trend as well. These observations underscore the necessity for a more nuanced approach to news aggregation. By effectively parsing and weighting news sentiments, MANA-Net proves to be a superior solution for news-driven market prediction.

7.3 Analysis of News Weights

To further validate the effectiveness of our weighted news aggregation, we analyze the distribution of generated news weights and use case studies for a comprehensive understanding. Since we transform news sentiments into trainable representations, which cannot be directly compared to the original sentiment scores, we focus on individual news utilization to demonstrate how we mitigate sentiment homogenization issues.

Figure 4 depicts the distribution of news weights across the test set. Since the range of news weights varies with daily news volume, we normalize these weights to a 0-1 scale, maintaining their rankings and ensuring comparability across different days. A key observation is that news weights are not uniform, reflecting varying degrees of market influence. Notably, the percentile lines indicate that over 50% of news items are assigned very low weights, suggesting that most news has only a minimal impact on market movements. Furthermore, more than 80% of news items receive weights below 0.8, illustrating that the majority do not significantly stand out amidst the voluminous news data. Conversely, fewer than 5% of news items have weights exceeding 0.98, highlighting a select few pieces of news that are particularly influential and

Figure 4: The kernel density estimate plot of MANA-Net’s news weights on the test set.



warrant specific attention. This distribution confirms that MANA-Net assigns varying weights to news items, emphasizing certain news by allocating higher weights, thereby distinguishing their relative importance in influencing market movements.

To illustrate the functionality of MANA-Net’s news weighting system, we analyze several news cases. Analyzing every news item is infeasible due to the sheer volume of daily news and the intricate link between news and market fluctuations. While many news pieces, like “Johnson & Johnson Completes Divestiture of LifeScan to Platinum Equity”, provide useful information, their specific market impact is difficult for humans to ascertain. Therefore, for a clear and focused analysis, we select major news items that apparently exert significant market influence. Table 5 lists these selected news cases, including broad market news such as the performance of the Dow Industrials and statements by Macron affecting EU and global markets. We also include influential incidents with clear market consequences, the 2018 murder of Saudi journalist Khashoggi, which significantly affected U.S.-Saudi relations and market sentiment.

These selected news items are recognized for their significant influence on the markets. In MANA-Net, each of them received a weight above 0.98, surpassing over 95% of all news items analyzed. The high weights underscore that MANA-Net effectively elevates the importance of these influential news items in its predictions. Furthermore, the sentiment scores of these news items are exceptionally polarized, with values exceeding 0.8, indicating strong positive or negative impacts. However, the last column shows that the corresponding daily average sentiments are highly homogenized, obscuring the distinct sentiments of these influential news items. This discrepancy highlights the shortcomings of equal-weight aggregation methods and demonstrates how MANA-Net successfully leverages critical news information to enhance market prediction.

8 Conclusion

In this paper, we address a significant challenge existing in financial news sentiment aggregation: Aggregated Sentiment Homogenization. This issue was explored through the analysis of an extensive news dataset. We developed MANA-Net, which is specifically designed to counteract this challenge. MANA-Net employs a market-news attention mechanism that dynamically aggregates news data, assigning varying weights to news items based on their assessed market impact. This approach not only facilitates a more sophisticated utilization of news sentiments but also ensures that the aggregation process remains adaptable and trainable, thus producing more effective news representations for market prediction. Experimental results underscore MANA-Net’s superior performance relative to various contemporary methods, highlighting its successful exploitation of news information for market analysis.

References

- [1] Chid Apte. 2010. The role of machine learning in business optimization. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*. Citeseer, 1–2.
- [2] Dogu Araci. 2019. Finbert: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063* (2019).
- [3] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* (2014).
- [4] Brad M Barber and Terrance Odean. 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The review of financial studies* 21, 2 (2008), 785–818.
- [5] Defang Chen, Jian-Ping Mei, Can Wang, Yan Feng, and Chun Chen. 2020. Online knowledge distillation with diverse peers. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 34. 3430–3437.
- [6] Yin-Wong Cheung, Rasmus Fatum, and Yohei Yamamoto. 2019. The exchange rate effects of macro news after the global financial crisis. *Journal of International Money and Finance* 95 (2019), 424–443.
- [7] Benjamin Clapham, Michael Siering, and Peter Gomber. 2021. Popular news are relevant news! How investor attention affects algorithmic decision-making and decision support in financial markets. *Information Systems Frontiers* 23, 2 (2021), 477–494.
- [8] George Cybenko. 1989. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems* 2, 4 (1989), 303–314.
- [9] Shumin Deng, Ningyu Zhang, Wen Zhang, Jiaoyan Chen, Jeff Z Pan, and Huajun Chen. 2019. Knowledge-driven stock trend prediction and explanation via temporal convolutional network. In *Companion Proceedings of The 2019 World Wide Web Conference*. 678–685.
- [10] Justina Deveikyte, Helyette Geman, Carlo Piccari, and Alessandro Proveti. 2022. A sentiment analysis approach to the prediction of market volatility. *Frontiers in Artificial Intelligence* 5 (2022), 836809.
- [11] Michael Dewally. 2003. Internet investment advice: Investing with a rock of salt. *Financial Analysts Journal* 59, 4 (2003), 65–77.
- [12] Qianggang Ding, Sifan Wu, Hao Sun, Jiaodong Guo, and Jian Guo. 2020. Hierarchical Multi-Scale Gaussian Transformer for Stock Movement Prediction.. In *IJCAI*. 4640–4646.
- [13] Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. 2015. Deep learning for event-driven stock prediction. In *Twenty-fourth international joint conference on artificial intelligence*.
- [14] Eugene F Fama. 1970. Efficient capital markets: A review of theory and empirical work. *The journal of Finance* 25, 2 (1970), 383–417.
- [15] Amirreza Farnoosh, Bahar Azari, and Sarah Ostadabbas. 2021. Deep switching auto-regressive factorization: Application to time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 7394–7403.
- [16] Ronen Feldman, Benjamin Rosenfeld, Roy Bar-Haim, and Moshe Fresko. 2011. The stock sonar—sentiment analysis of stocks based on a hybrid approach. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 25. 1642–1647.
- [17] Thomas Fischer and Christopher Krauss. 2018. Deep learning with long short-term memory networks for financial market predictions. *European journal of operational research* 270, 2 (2018), 654–669.
- [18] Rohitha Goonatilake and Susantha Herath. 2007. The volatility of the stock market and news. *International Research Journal of Finance and Economics* 3, 11 (2007), 53–65.
- [19] Yuechun Gu, Da Yan, Sibo Yan, and Zhe Jiang. 2020. Price forecast with high-frequency finance data: An autoregressive recurrent neural network model with technical indicators. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 2485–2492.
- [20] Min Hou, Chang Xu, Yang Liu, Weiqing Liu, Jiang Bian, Le Wu, Zhi Li, Enhong Chen, and Tie-Yan Liu. 2021. Stock trend prediction with multi-granularity data: A contrastive learning approach with adaptive fusion. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 700–709.
- [21] Wenpeng Hu, Mengyu Wang, Bing Liu, Feng Ji, Haiqing Chen, Dongyan Zhao, Jinwen Ma, and Rui Yan. 2019. Transformation of dense and sparse text representations. *arXiv preprint arXiv:1911.02914* (2019).
- [22] Wenpeng Hu, Mengyu Wang, Qi Qin, Jinwen Ma, and Bing Liu. 2020. Hrn: A holistic approach to one class learning. *Advances in neural information processing systems* 33 (2020), 19111–19124.
- [23] Chenn-Jung Huang, Jia-Jian Liao, Dian-Xiu Yang, Tun-Yu Chang, and Yun-Cheng Luo. 2010. Realization of a news dissemination agent based on weighted association rules and text mining techniques. *Expert Systems with Applications* 37, 9 (2010), 6409–6413.
- [24] Thanh D Huynh and Daniel R Smith. 2017. Stock price reaction to news: The joint effect of tone and attention on momentum. *Journal of Behavioral Finance* 18, 3 (2017), 304–328.
- [25] Weiwei Jiang. 2021. Applications of deep learning in stock market prediction: recent progress. *Expert Systems with Applications* 184 (2021), 115537.
- [26] Fang Jin, Nathan Self, Parang Saraf, Patrick Butler, Wei Wang, and Naren Ramakrishnan. 2013. Forex-foreteller: Currency trend modeling using news articles. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. 1470–1473.
- [27] Nan Jing, Zhao Wu, and Hefei Wang. 2021. A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction. *Expert Systems with Applications* 178 (2021), 115019.
- [28] Rafal Jozefowicz, Wojciech Zaremba, and Ilya Sutskever. 2015. An empirical exploration of recurrent network architectures. In *International conference on machine learning*. PMLR, 2342–2350.
- [29] Kelvin JL Koa, Yunshan Ma, Ritchie Ng, and Tat-Seng Chua. 2023. Diffusion Variational Autoencoder for Tackling Stochasticity in Multi-Step Regression Stock Price Prediction. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*. 1087–1096.
- [30] Mathias Kraus and Stefan Feuerriegel. 2017. Decision support from financial disclosures with deep neural networks and transfer learning. *Decision Support Systems* 104 (2017), 38–48.
- [31] Christopher Krauss, Xuan Anh Do, and Nicolas Huck. 2017. Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research* 259, 2 (2017), 689–702.
- [32] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. 1989. Backpropagation applied to handwritten zip code recognition. *Neural computation* 1, 4 (1989), 541–551.
- [33] Feng Li. 2010. The information content of forward-looking statements in corporate filings—A naïve Bayesian machine learning approach. *Journal of Accounting Research* 48, 5 (2010), 1049–1102.
- [34] Qing Li, Jinghua Tan, Jun Wang, and Hsinchun Chen. 2020. A multimodal event-driven LSTM model for stock prediction using online news. *IEEE Transactions on Knowledge and Data Engineering* 33, 10 (2020), 3323–3337.
- [35] Xiaodong Li, Haoran Xie, Li Chen, Jianping Wang, and Xiaotie Deng. 2014. News impact on stock price return via sentiment analysis. *Knowledge-Based Systems* 69 (2014), 14–23.
- [36] Yang Li and Yi Pan. 2022. A novel ensemble deep learning model for stock prediction based on stock prices and news. *International Journal of Data Science and Analytics* (2022), 1–11.
- [37] Jintao Liu, Hongfei Lin, Xikai Liu, Bo Xu, Yuqi Ren, Yufeng Diao, and Liang Yang. 2019. Transformer-based capsule network for stock movement prediction. In *Proceedings of the first workshop on financial technology and natural language processing*. 66–73.
- [38] Alejandro Lopez-Lira and Yuehua Tang. 2023. Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models. *Return Predictability and Large Language Models (April 6, 2023)* (2023).
- [39] Tim Loughran and Bill McDonald. 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of finance* 66, 1 (2011), 35–65.
- [40] André Luis Santiago Maia and Francisco de AT de Carvalho. 2011. Holt’s exponential smoothing and neural network models for forecasting interval-valued time series. *International Journal of Forecasting* 27, 3 (2011), 740–759.
- [41] Pooja Mehta, Shamil Pandya, and Ketan Kotecha. 2021. Harvesting social media sentiment analysis to enhance stock market prediction using deep learning. *PeerJ Computer Science* 7 (2021), e476.
- [42] Piotr Mirowski, M Ranzato, and Yann LeCun. 2010. Dynamic auto-encoders for semantic indexing. In *Proceedings of the NIPS 2010 Workshop on Deep Learning*, Vol. 2.
- [43] Kostadin Mishev, Ana Gjorgjevikj, Irena Vodenska, Lubomir T Chitkushev, and Dimitar Trajanov. 2020. Evaluation of sentiment analysis in finance: from lexicons to transformers. *IEEE access* 8 (2020), 131662–131682.
- [44] Leela Mitra and Gautam Mitra. 2011. Applications of news analytics in finance: A review. *The handbook of news analytics in finance* 596, 1 (2011).
- [45] Saloni Mohan, Sahitya Mullanpudi, Sudheer Sammeta, Parag Vijayvergia, and David C Anastasiu. 2019. Stock price prediction using news sentiment analysis. In *2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService)*. IEEE, 205–208.
- [46] Larry Olanrewaju Orimoloye, Ming-Chien Sung, Tiejun Ma, and Johnnie EV Johnson. 2020. Comparing the effectiveness of deep feedforward neural networks and shallow architectures for predicting stock price indices. *Expert Systems with Applications* 139 (2020), 112828.
- [47] Ramit Sawhney, Shivam Agarwal, Arnav Wadhwa, Tyler Derr, and Rajiv Ratn Shah. 2021. Stock selection via spatiotemporal hypergraph attention network: A learning to rank approach. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 497–504.
- [48] Ramit Sawhney, Arnav Wadhwa, Shivam Agarwal, and Rajiv Shah. 2021. FAST: Financial news and tweet based time aware network for stock trading. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*. 2164–2175.
- [49] Aliaksei Severyn and Alessandro Moschitti. 2015. Twitter sentiment analysis with deep convolutional neural networks. In *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval*. 959–962.

- [50] Karthik Shivaram, Ping Liu, Matthew Shapiro, Mustafa Bilgic, and Aron Culotta. 2022. Reducing Cross-Topic Political Homogenization in Content-Based News Recommendation. In *Proceedings of the 16th ACM Conference on Recommender Systems*. 220–228.
- [51] Sahar Sohangir, Nicholas Petty, and Dingding Wang. 2018. Financial sentiment lexicon analysis. In *2018 IEEE 12th International Conference on Semantic Computing (ICSC)*. IEEE, 286–289.
- [52] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).
- [53] Pietro Veronesi. 1999. Stock market overreactions to bad news in good times: a rational expectations equilibrium model. *The Review of Financial Studies* 12, 5 (1999), 975–1007.
- [54] Chaojie Wang, Yuanyuan Chen, Shuqi Zhang, and Qiuhui Zhang. 2022. Stock market index prediction using deep Transformer model. *Expert Systems with Applications* 208 (2022), 118128.
- [55] Gang Wang, Tianyi Wang, Bolun Wang, Divya Sambasivan, Zengbin Zhang, Haitao Zheng, and Ben Y Zhao. 2014. Crowds on wall street: Extracting value from social investing platforms. *arXiv preprint arXiv:1406.1137* (2014).
- [56] Qili Wang, Wei Xu, and Han Zheng. 2018. Combining the wisdom of crowds and technical analysis for financial market prediction using deep random subspace ensembles. *Neurocomputing* 299 (2018), 51–61.
- [57] Walter Wasserfallen. 1989. Macroeconomics news and the stock market: Evidence from Europe. *Journal of Banking & Finance* 13, 4-5 (1989), 613–626.
- [58] Qiong Wu, Christopher G Brinton, Zheng Zhang, Andrea Pizzoferrato, Zhenming Liu, and Mihai Cucuringu. 2021. Equity2vec: End-to-end deep learning framework for cross-sectional asset pricing. In *Proceedings of the Second ACM International Conference on AI in Finance*. 1–9.
- [59] Sheng Xiang, Dawei Cheng, Chencheng Shang, Ying Zhang, and Yuqi Liang. 2022. Temporal and heterogeneous graph neural network for financial time series prediction. In *Proceedings of the 31st ACM international conference on information & knowledge management*. 3584–3593.
- [60] Yunan Ye, Hengzhi Pei, Boxin Wang, Pin-Yu Chen, Yada Zhu, Ju Xiao, and Bo Li. 2020. Reinforcement-learning based portfolio management with augmented asset movement prediction states. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 1112–1119.
- [61] Xu Yuemei, Wang Zihou, and Wu Zixin. 2021. Predicting Stock Trends with CNN-BiLSTM Based Multi-Feature Integration Model. *Data Analysis and Knowledge Discovery* 5, 7 (2021), 126–138.
- [62] Lei Zhang, Shuai Wang, and Bing Liu. 2018. Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 8, 4 (2018), e1253.
- [63] Qiuyue Zhang, Chao Qin, Yunfeng Zhang, Fangxun Bao, Caiming Zhang, and Peide Liu. 2022. Transformer-based attention network for stock movement prediction. *Expert Systems with Applications* 202 (2022).
- [64] Yongjie Zhang, Weixin Song, Dehua Shen, and Wei Zhang. 2016. Market reaction to internet news: Information diffusion and price pressure. *Economic Modelling* 56 (2016), 43–49.
- [65] Tian Zhou, Peisong Niu, Liang Sun, Rong Jin, et al. 2023. One fits all: Power general time series analysis by pretrained lm. *Advances in neural information processing systems* 36 (2023), 43322–43355.
- [66] Xizhou Zhu, Weijie Su, Lewei Lu, Bin Li, Xiaogang Wang, and Jifeng Dai. 2020. Deformable detr: Deformable transformers for end-to-end object detection. *arXiv preprint arXiv:2010.04159* (2020).