FinBPM: A Framework for Portfolio Management-based Financial Investor Behavior Perception Model

Zhilu Zhang[◇♣], Prochta Sen[♣*], Zimu Wang[♠], Ruoyu Sun[△],
Zhengyong Jiang[◇*], Jionglong Su[◇*]
[◇]School of AI and Advanced Computing, XJTLU Entrepreneur College (Taicang);
[♠]Department of Computing, School of Advanced Technology;
[△]Department of Financial and Actuarial Mathematics, School of Mathematics and Physics,
Xi'an Jiaotong-Liverpool University, Suzhou, 215123, China
[♣]Department of Computer Science, University of Liverpool, Liverpool, L69 3BX, UK
{zhilu,Procheta.Sen}@liverpool.ac.uk;
{Jionglong.Su,Zhengyong.Jiang02}@xjtlu.edu.cn

Abstract

The goal of portfolio management is to simultaneously maximize the accumulated return and also to control risk. In consecutive trading periods, portfolio manager needs to continuously adjust the portfolio weights based on the factors which can cause price fluctuation in the market. In the stock market, the factors affecting the stock price can be divided into two categories. The first is price fluctuations caused by irrational investment of the speculators. The second is endogenous value changes caused by operations of the company. In recent years, with the advancement of artificial intelligence technology, reinforcement learning (RL) algorithms have been increasingly employed by scholars to address financial problems, particularly in the area of portfolio management. However, the deep RL models proposed by these scholars in the past have focused more on analyzing the price changes caused by the investment behavior of speculators in response to technical indicators of actual stock prices. In this research, we introduce an RL-based framework called FinBPM, which takes both the factor pertaining to the impact on operations of the company and the factor of the irrational investment of the speculator into consideration. For our experimentation, we randomly selected 12 stocks from the Dow Jones Industrial Index to construct our portfolio. The experimental results reveal that, in comparison to conventional reinforcement learning methods, our approach with at least 13.26% increase over other methods compared. Additionally, it achieved the best Sharpe ratio of 2.77, effectively maximizing the return per unit of risk.

1 Introduction

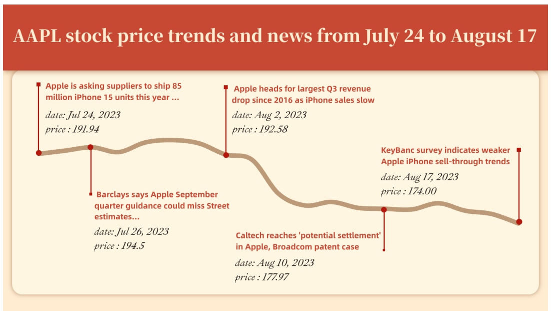

Figure 1: AAPL reported on August 2 that "Apple heads for largest Q3 revenue drop since 2016 as iPhone sales slow", and the stock price dropped significantly a few days later.

Many studies conclude that stock movements follow a random walk (Merton, 1980; Samuelson, 2015). With the development of artificial intelligence techniques, many scholars attempt to predict stock price movements or trends based on machine learning (Freitas et al., 2009; Niaki and Hoseinzade, 2013; Heaton et al., 2017; Goudar et al., 2022). However, it also turns out that accurately predicting future market prices remains difficult. Therefore, rather than directly forecasting prices, our work transforms the problem into predicting investment flows based on investor behavior. By modeling endogenous and exogenous behavioral factors influencing stock price, we aim to emulate macro-level investor dynamics that drive market prices. This agent-based, behavior-centric perspective circumvents the need to absolute valuation, instead focusing on more tractable signals correlated with crowd behavior.

Investor behavior is driven by two key factors: irrational price fluctuations from speculative market timing (Panchuk and Westerhoff, 2021), and fundamental company value (Syifaudin et al., 2020). Numerous models exploit irrational price fluctuations (Huang et al., 2016; Feng et al., 2019; Liu et al., 2020; Qin et al., 2022; Yang et al., 2022). While price and volume data can capture irrational fea-

[*] Jionglong Su, Procheta Sen and Zhengyong Jiang are corresponding authors.

Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics
Volume 1: Long Papers, pages 246–257
March 17-22, 2024 ©2024 Association for Computational Linguistics

tures (Goudar et al., 2022), these signals alone do not consider endogenous information from events like earnings surprises, mergers, or corporate actions (Chen and Huang, 2021). Such value-relevant events are often disclosed through financial news and social media, as well as substantially impact market movements (Oh and Sheng, 2011). Models such as S-Reward (Yang et al., 2018), SARL (Ye et al., 2020), and PROFIT (Sawhney et al., 2021) incorporate news text to estimate intrinsic value and inform decisions. However, they rely on correlating news with prices rather than directly extracting fundamental value, losing the core purpose of fundamental value discovery. Figure 1 gives the correlation between AAPL stock price trends and news. When Apple was reported to have slowed down iPhone sales on August 2, the stock price fell for several days. Additionally, using individual tweets or headlines provides limited information (Wang and Gan, 2023), as not all texts equally impact value (Hu et al., 2018).

These limitations motivate research info efficiently utilizing financial text with prices to model investor behavior. An integrative model of investor behavior requires incorporating market transaction data with textual sources reflecting intrinsic value. Our research addresses these limitations by: 1) Directly parsing semantic signals of intrinsic value from news content rather than just price correlations; 2) Leveraging full articles to extract richer insights versus restricted headlines or tweets; 3) Filtering text to focus on the most relevant endogenous valuation drivers. This allows more targeted modeling of how financial news influences investor demand through fundamental value, complementing price data that captures exogenous speculation.

Contributions:

- We propose the FinBPM, a novel investor behavior-driven portfolio management framework using reinforcement learning. To the best of knowledge, we are the first to consider investor behavior in portfolio management. Our approach combines time series modeling of prices and volumes, with natural language processing of financial news, to jointly characterize both irrational and intrinsic drivers of investor behavior. This dual-view data fusion provides a more complete representation of the multifaceted factors governing financial markets.

- We perform extensive ablation experiments to determine optimal financial text processing for maximizing intrinsic value signals under the portfolio management task. The experimental results demonstrate that on our dataset, selecting the four most salient sentences from the full news text achieves the best portfolio management performance. Our results provide a basis for financial text-related research that selectively extracting and analyzing salient information from lengthy news reports may enhance portfolio management performance.

- We randomly select twelve stocks in Dow Jones Index to be used in our experiments. We also establish a financial news dataset encompassing twelve stocks with company classifications. This dataset facilitates into portfolio management approaches leveraging news content analysis and promotes academic study of finance techniques integrating textual data mining. Experimental results show that the cumulative return of FinBPM is at least 13.26% move than baseline strategies, while achieved the best Sharpe ratio of 2.77, and control the Maximum Drawdown in 10.10%.

2 Background

This paper presents a framework for portfolio management that incorporates investor behavior forecasting which combines financial text processing techniques and price volatility characteristics techniques. We provide an overview of the relevant technologies for each component, a new financial investor behavior perception module for enhancing investment portfolio management tasks.

Financial Text Processing Techniques. Financial text processing techniques involve the use of natural language processing (NLP) to analyze financial text data such as financial statements, news articles, and social media posts. TextRank (Mihalcea and Tarau, 2004) is a graph-based algorithm for text summarization and keyword extraction. It constructs a graph of relationships between sentences in the text and calculates the importance of sentences based on their similarities.

Pegasus (Zhang et al., 2020) is a pre-trained language generation model that uses the Transformer architecture and is trained on large-scale text data. The aim of Pegasus is to generate high-quality text summaries, compressing long texts into concise summaries. In the context of financial news processing, using TextRank can help extract the most important sentences from the full text, enabling faster access to key information. On the other hand, Pegasus can summarize the entire article, providing a concise summary to help users quickly grasp the core content of the article.

FinBERT (Yang et al., 2020) is a language model pre-trained on financial text data. It is based on BERT (Bidirectional Encoder Representations from Transformers) (Devlin et al., 2019), but pre-trained on a large corpus of financial documents like earnings reports, analyst reports, news articles, and financial forums. Achieves state-of-the-art results on many financial NLP benchmarks.

Price Volatility Characteristics Techniques. Historical stock prices, trading volume and other information can be used to predict stock prices (Soni et al., 2022). LSTM has been widely used for time series prediction due to its ability to handle long sequences using gating mechanisms (Obthong et al., 2020). However, it faces challenges with increasingly long sequences and poor performance in extreme cases. Informer uses a Transformer architecture that can process both long and short sequences while effectively learning and extracting features at different time scales (Zhou et al., 2021). It also incorporates attention mechanisms and adaptive lengths to further improve prediction performance. Analyzing social media data or the sentiment of news reports is also a way to predict stock prices (Yadav and Vishwakarma, 2020).

3 Problem Description

We shall describe the traditional model corresponding to the *Portfolio Optimization* problem. Let $S = \{s_1, s_2, \ldots, s_N\}$ denote a set of N stocks. In portfolio optimization we design a stock trading model to generate maximum cumulative return over all the stock trades across N stocks within a time period T. The cumulative return corresponding to any stock s_i at any time step τ depends on the state of the stock at that time step $\{s_{i_\tau}\}$ and the trading action applied to s_i, denoted as $a[s_i]_\tau$. We establish our State Space $\{s_{i_\tau}\}$ based on PROFIT (Sawhney et al., 2021). This research utilizing FinRL (Yang et al., 2021), an open-source framework, to systematically construct the action space $a[s_i]_\tau$ and rewards function r .

State Space: At each time-step τ, s_{i_τ} consists of two parts: Stock trading account status for s_i denoted as $o[s_{i_\tau}]$ and Market information status $m[s_{i_\tau}]$. $o[s_{i_\tau}]$ consists of account balance $b[s_{i_\tau}]$ and the holdings $n[s_i]_\tau$. $m[s_{i_\tau}]$ have financial news related to s_i released during a T-day lookback period.

Actions Space: At each time-step τ, trading actions can be of three types: buy, sell, and hold. $n[s_i]_\tau$ represents the volume of s_i at τ. Mathematically speaking,

$$n[s_i]_{\tau+1} = n[s_i]_\tau + a[s_i]_\tau. \tag{1}$$

If $a[s_i]_\tau$ is a buying action then $a[s_i]_\tau \in 1, 2, \ldots, h_{s_i}$ where h_{s_i} represents the maximum buying volume for a stock s_i. $a[s_i]_\tau = 0$ if it is a holding action. $a[s_i]_\tau \in -n[s_i]_\tau, \ldots, -2, -1$ if $a[s_i]_\tau$ is a selling action.

Rewards Function: Reward function r is defined as the change of the total value when the state changes from s_{i_τ} to $s_{i_{\tau+1}}$ due to a trading action,

$$r(s_i)_{\tau,\tau+1} = \left(b_{\tau+1} + p_{\tau+1}^T n_{\tau+1}\right) - \left(b_\tau + p_\tau^T n_\tau\right) - c_\tau, \tag{2}$$

where p_τ denotes the price at this time-step τ. We incorporate transaction fee rates for each transaction, denoted by c_τ.

4 Proposed Framework

Our proposed framework consists of two different modules. They are Financial Investor Behavior Perception Module and Investment Decision Module. Broadly speaking, the aim of perception module is to provide an idea about the current state of the stock. The perception module outputs an index that is exploited by the investment decision network to determine the trading action at any time t. Figure 2 gives the Framework of FinBPM. Each module of FinBPM is described as follows.

4.1 Financial Investor Behavior Perception

This module processes two types of heterogeneous data that influence investor behavior - numerical market data such as price and volume as well as textual news data. The rationale behind using numerical data is to learn the characteristic patterns of irrational price dynamics can be learnt. Numerical market data is comprised of time series data of historical prices and volumes corresponding to each stock. Similarly, the rationale behind using financial text is to model the effects of company intrinsic value factors expressed through financial text.

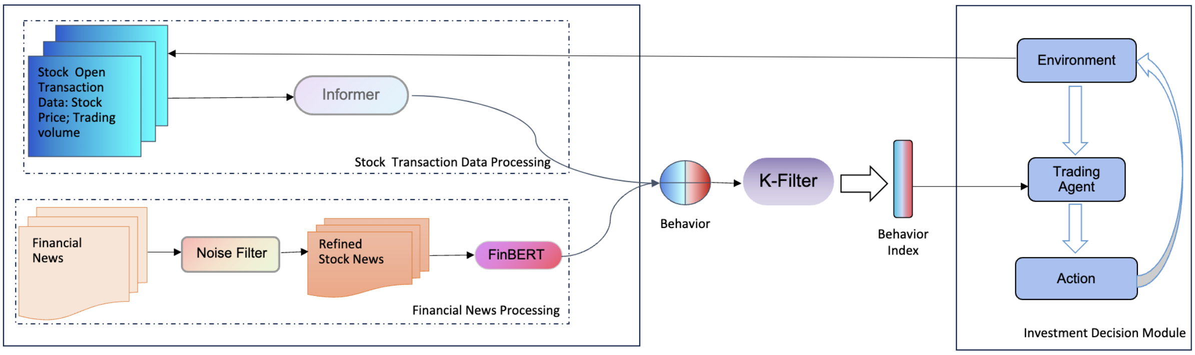
Stock Open Transaction Data: Stock Price; Trading volume
Informer
Stock Transaction Data Processing
Financial News
Noise Filter
Refined Stock News
FinBERT
Financial News Processing
Behavior
K-Filter
Behavior Index
Environment
Trading Agent
Action
Investment Decision Module

Figure 2: Framework of FinBPM. The financial investor behavior perception module (left) analyzes investor behavior by processing market transaction data and financial news. The investment decision module (right) integrates the current market environment (price and transaction volume) with the investor behavior index to generate investment decisions.

Separate modules tailored to each data type is used to handle distinct data modalities. An Informer (Zhou et al., 2021) model is trained on numerical market data which eventually estimates the impact of the irrational factors on the stock price. For text data, TextRank algorithm is first used to extract salient sentences from news articles, reducing noise. The filtered texts are then processed by FinBERT, a financial domain-specific BERT model fine-tuned on a large corpus of financial texts, to assess the influence degree on the associated company through sentiment score.

Combining the output of the above mentioned two modules, a final index S^{st} is used to estimate comprehensive investor behavior,

$$S^{st} = \begin{cases} S^s + \alpha * S^t, & \text{if } S^s * S^t > 0 \\ S^s, & \text{Otherwise} \end{cases} . \quad (3)$$

In Equation 3, the endogenous index $S^s \in [-1, 1]$ and similarly exogenous index $S^t \in [-1, 1]$. The range of S^{st} is also $[-1, 1]$. $\alpha \in (0, 1)$ represents a adjustable coupling coefficient. If S^s and S^t indicate the same directional trend, we combine the two indices. If S^s and S^t have opposing predicted trends, only S^s is retained as the final index, prioritizing the endogenous valuation signal. This selective coupling approach integrates the endogenous and exogenous factors when aligned, while filtering out exogenous noise when contradicting the endogenous financial text-based valuation index. The resulting S^{st} integrates the two aspects of investor behavior in a robust way.

4.2 Investment Decision Module

Intrinsic value derived from financial news is the primary driver of investor behavior in FinBPM. However, not all news can influence investor behavior. Consequently, we implement a k-filter layer to remove less impactful investor behavior. Only when the absolute value of investor behavior in the environment state is greater than $k \in [0, 1]$, the investor behavior index will then be applied to the trading process. After filtering the investor behavior, the state with behavior S^{st} is combined with environmental information such as stock prices and trading volume S^p to form an environment state $S(S^p, S^{st})$ with investor behavior.

We base the investment decision network on the original Proximal Policy Optimization (PPO) algorithm (Schulman et al., 2017) consisting of a fully connected Multilayer Perceptron (MLP) with two hidden layers of 64 units, and tanh nonlinearities. S then processed by MLP to output an action. For a single stock, the action space is defined as $\{-h, \ldots, -1, 0, 1, \ldots, h\}$, where $|h|$ is a predefined parameter that sets as the maximum volume of shares for each buying action as described in Section 3.

In the process of updating the decision network, the policy (decision) loss function is defined as follows:

$$\begin{aligned} L^{\text{CLIP}}(\theta) =& \hat{\mathbb{E}}_t \Big[\min \Big(r_t(\theta) \hat{A}(s_t, a_t), \\ & \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon) \hat{A}(s_t, a_t) \Big) \Big], \end{aligned} \quad (4)$$

In Equation 4, $\hat{\mathbb{E}}_t$ denotes expectation. $r_t(\theta)\hat{A}(s_t, a_t)$ is the normal policy gradient objective, and $\hat{A}(s_t, a_t)$ represents the advantage function, which reflects the degree of improvement of the current strategy relative to the old strategy. It can also be understood as the difference in cumulative return obtained under the current action. $r_t(\theta)$ represents the ratio of the probabilities of the old and new strategies. The function $\text{clip}\,(r_t(\theta), 1-\epsilon, 1+\epsilon)$ limits the ratio $r_t(\theta)$ to be within $[1-\epsilon, 1+\epsilon]$. ϵ is a predetermined constant, such as 0.1 or 0.2. The function of clip is to limit the update range of the policy, avoid drastic changes in the policy, perform stable policy updates, and prevent extreme policy changes, thereby achieving stable and efficient policy optimization (Schulman et al., 2017). The objective function of PPO takes the minimum of the clipped and normal objective. PPO discourages large policy beyond of the clipped interval. Therefore, PPO improves the stability of the policy networks training by restricting the policy update at each training step. We select PPO for stock trading because it is stable, fast, and simple to implement (Schulman et al., 2017; Zheng et al., 2023). We have provided a more detailed description of Equation 4 in Appendix A.

5 Experiment Setup

In this section, we first provide an overview of the dataset used in our experiment. Subsequently, we introduce into the various baseline methods employed in our experimental setup, elaborating on the hardware and parameter configurations.

5.1 Dataset

The dataset consist of mainly two components: a) Stock information (i.e., daily closing stock prices, trading volumes, technical indicators (e.g., MACD, RSI)) and b) Financial news texts from 2018-07-01 to 2021-03-01. The price, volume, and indicator data are collected from Yahoo Finance[1], while the news texts comes from our financial news dataset. We divide the timeline into training (2018-07-01 to 2020-07-01), validation (2020-07-01 to 2020-09-01), and test (2020-09-01 to 2021-03-01) sets.

Our financial news dataset consists of nearly 20,000 articles from three major financial websites - Investing[2], Bloomberg[3], and Reuters[4] - covering 20 different companies. Unlike previous financial news datasets (Ding et al., 2014; Xu and Cohen, 2018), our corpus classifies each article by the associated company. Of the total articles, 7,813 contain the full text. Table 1 gives the distribution of the 7,813 full text articles across publication year, company, and text length in characters. As given in Table 2, each news item in our dataset includes the stock ticker, headline, publication date, and full text source. With multi-source labeled company news spanning recent years, our corpus provides a comprehensive up-to-date resource for analyzing the impact of financial texts on individual stocks.

5.2 Baseline

We compare our FinBPM strategy with both *traditional* and *reinforcement learning* strategies.

Traditional: Such methods use traditional models based on stock price fluctuations.

- **BK (Györfi et al., 2006):** The Nonparametric Kernel Based Log Optimal Strategy (BK) is a sophisticated approach in quantitative finance that leverages nonparametric kernel density estimation techniques to construct an optimal investment strategy. By utilizing log-optimal criteria, this strategy aims to maximize the expected logarithmic utility of wealth, while incorporating the underlying characteristics of the financial data through the flexible and adaptive nature of kernel methods.
- **CRP (Cover, 1991):** Constant Rebalanced Portfolios (CRP) is an investment strategy that aims to maintain a fixed allocation of assets over time by periodically rebalancing the portfolio. By adjusting the portfolio weights to their initial proportions at regular intervals, CRP can achieve risk reduction and potentially outperform other strategies in certain market conditions.
- **OLMAR (Li and Hoi, 2012):** Online Portfolio Selection with Moving Average Reversion (OLMAR) is a research area focused on developing trading strategies that dynamically allocate assets in an online manner while considering the moving average reversion phenomenon.
- **RMR (Huang et al., 2016):** Robust Median Reversion (RMR) is a quantitative investment strategy that employs statistical and mathematical techniques to identify and exploit the

[1] www.finance.yahoo.com
[2] www.investing.com
[3] www.bloomberg.com
[4] www.reuters.com

			Statistics						
Dataset				**Stock Price**			**Stock Volume**		
Type	Split	Time Range	#stocks	Min.	Avg.	Max.	Min.	Avg.	Max.
Stock Info	Train	2018-07-01 to 2020-07-01	12	244.2	30.45	108.12	106928300	612800	10334443
	Valid	2020-07-01 to 2020-09-01	12	277.63	34.99	127.63	182269900	1123800	9828417
	Test	2020-09-01 to 2021-03-01	12	281.25	30.85	135.27	124070700	585700	11117990
Dataset				**News Length**			**New Articles**		
Type	Split	Time Range		Min.	Avg.	Max.	Min.	Avg.	Max.
Financial News	Train	2018-07-01 to 2020-07-01	12	64	2574.32	31451	0	116.83	324
	Valid	2020-07-01 to 2020-09-01	12	116	2475.94	7937	0	12	26
	Test	2020-09-01 to 2021-03-01	12	120	2550.18	14028	0	35.16	69

Table 1: Introduction to the detailed information of the dataset. The stock trading information includes transaction prices and trading volumes. The news dataset includes the number of different sources of news and the length of the news.

Stock	Title	Date	Full Text
INTC	Intel slashes dividend to conserve cash ahead of U.S. capacity expansion	Feb 22, 2023	By Geoffrey Smith Investing.com – Intel (NASDAQ:INTC) said it will cut its dividend by two-thirds in an effort to conserve cash as it prepares for a massive expansion of chipmaking capacity in the U.S.The semiconductor giant said it will reset its quarterly dividend at 12.5c, down from 36.5c...

Table 2: Sample in the financial news dataset.

mean-reverting behavior of financial assets. RMR aims to mitigate the impact of outliers or extreme observations, thereby enhancing the resilience of strategy in volatile market conditions.

- **PAMR (Li et al., 2012):** Passive Aggressive Mean Reversion (PAMR) is a quantitative trading strategy that combines principles from machine learning and statistical arbitrage to identify and exploit the mean-reverting behavior of financial assets. By employing passive and aggressive actions based on observed market conditions, PAMR dynamically adjusts its trading position to optimize profitability while minimizing risk.

Reinforcement Learning: The following approaches optimize portfolio management through reinforcement learning.

- **PPO (Schulman et al., 2017):** A policy gradient method that uses multiple epochs of minibatch updates along with clipping of the objective function to improve sample efficiency and stabilize training.

- **A2C (Mnih et al., 2016):** A synchronous version of the asynchronous A3C algorithm. It uses an actor-critic approach with multiple workers interacting with environments in parallel. The gradients from each worker are synchronized periodically.

- **DDPG (Lowe et al., 2017):** An actor-critic method for continuous action spaces that uses a replay buffer and target networks. The actor maps states to actions directly while the critic evaluates the policy.

- **SAC (Haarnoja et al., 2018):** An off-policy actor-critic algorithm that incorporates entropy regularization to encourage exploration. It learns a stochastic policy along with state-value and policy-value functions.

- **SARL (Ye et al., 2020):** Incorporates heterogeneous data sources into RL training for portfolio management. The key idea is to augment the state with additional predictive signals (e.g. predicted price movements from news articles)

5.3 Evaluation Metrics

We chose **Sharpe ratio (SR)**, the **Cumulative Return (CR)**, and the **Max Drawdown (MDD)** which are commonly used in the financial field as our evaluation metrics for evaluating model performance.

The SR is a measure of the risk and return of an investment portfolio (Sharpe, 1964). $SR = \frac{R_p - R_f}{\sigma_p}$. We use R_p to denote the return of the investment portfolio, the risk-free rate R_f to represent. σ_p is the standard deviation of the portfolio's annualized return.

The CR is the change in the investment over time and is computed using the initial (b_0) and the final (b_f) account balance as: $CR = \frac{b_f - b_0}{b_0} * 100\%$.

MDD is used to evaluate risk control. The calculation method of MDD can be expressed as: $MDD = Max(\frac{r_t - r_p}{r_p} * 100\%)$. Larger values (in magnitude) of MDD indicate higher volatility. MDD represents the unit rate of change from the highest net asset value r_p to the lowest net asset value r_t after the highest net asset value during a period of continuous decline.

In addition, we use β to measure the risk and return of strategies. The calculation method of β can be expressed as: $\beta = \frac{Cov(r_a, r_m)}{\mu_m}$. $Cov(r_a, r_m)$ represents the covariance between cumulative returns r_a and market returns r_m. μ_m represents the variance of market returns. The β reflects the degree of systematic risk of the model strategy relative to the overall market. A higher β value indicates higher volatility and systematic risk. β equal to 1 means its risk level is comparable to the market; β greater than 1 means its risk is higher than the market average; β less than 1 means its risk is lower than the market average.

5.4 Trading Setting

We evaluate portfolio management strategy using a 12-stock portfolio with an initial capital of \$100,000 and no initial stock holdings. Starting from the first trading day, the buy, hold and sell actions on each stock are determined dynamically based on market conditions. To prevent overconcentration in a single stock, a maximum daily buy volume of 10 is set for each stock ($h_{\max} = 10$). Due to the inherent randomness in reinforcement learning, each RL strategy was evaluated across five experimental runs, and the average performance was reported. During trading, we record CR, SR, and MDD. This setup allows realistic simulation of managing a diversified portfolio by taking modulated positions in individual assets based on the policy of our model. Tracking key portfolio risk and return metrics provides a comprehensive performance assessment.

All our experiments are performed using the PPO reinforcement learning framework with the same parameter settings. The specific parameter settings are: n_steps (cumulative returns from the current moment forward for n_steps moments to update the value function or strategy function): 2048, ent_coef (control the trade-off between strategy entropy and reward): 0.01, learning rate: 0.00025, batch size: 128. The GPU computing resource used in the experiment is Tesla T4.

6 Results and Discussion

In Table 3, among the traditional methods, RMR achieves the highest cumulative return of 38.86% and Sharpe ratio of 1.93. It also has the highest β of 4.62. The investment strategy returns substantially exceed the benchmark over the experiment period. For RL methods, SAC performs best with a 19.87% cumulative return and 1.6097 Sharpe ratio. DDPG also performs with a 16.77% return. Our proposed FinBPM approach achieves superior performance compared to other methods. FinBPM (Mean) obtains the highest cumulative returns, with a 35.19% cumulative return and Sharpe ratio of 2.30. FinBPM (Best) achieves the maximum 52.12% return with a Sharpe ratio of 2.77, exceeding all other models in cumulative returns and Sharpe ratio metrics. FinBPM also exhibits strong risk control with very low maximum drawdown. The high β coefficients of 4.15 and 6.33 for FinBPM further demonstrate its strong capabilities.

Figure 3 gives the cumulative return for FinBPM versus baseline methods. The red line (FinBPM) shows exhibiting stable growth overall, with only a minor drawdown around the 60th trading day. In contrast, other methods display larger fluctuations in the second half of the trading period, while FinBPM maintains a relatively steady upward trajectory. This highlights advantage of FinBPM in risk control and consistent returns, particularly in the latter trading days, where even RMR sees volatile shocks. The stable growth trend of FinBPM throughout the timeline, with minimal drawdown, demonstrates its robustness in portfolio management. Overall, the experimental results validate the advantages of FinBPM in dual investor behavior modeling and reinforcement learning optimization for enhanced portfolio management.

6.1 Financial News Processing:

Unlike prior works using just news headlines or tweets (Yang et al., 2018; Ye et al., 2020; Du and Tanaka-Ishii, 2020), we leverage full text articles for richer financial news modeling. To optimally extract and utilize the embedded information, we conduct ablation experiments on different news text processing methods, evaluating the cumulative returns. Due to randomness in reinforcement learning, each method is tested in five runs to ob-

Method	Strategy	Cumulative Return	Sharp Ratio	Max Drawdown	β
Traditional	BK (Györfi et al., 2006)	-0.63%	0.0443	-0.1633	-0.46
	CRP (Cover, 1991)	21.35%	2.2289	<u>-0.0867</u>	2.36
	OLMAR (Li and Hoi, 2012)	25.54%	1.3509	-0.1046	2.90
	RMR (Huang et al., 2016)	<u>38.86%</u>	1.9399	-0.1086	<u>4.62</u>
	PAMR (Li et al., 2012)	-11.19%	-0.5315	-0.1495	-1.83
RL	PPO (Schulman et al., 2017)	-2.095%	-0.2392	**-0.0673**	-0.657
	A2C (Mnih et al., 2016)	-5.49%	-0.9920	-0.0811	-1.095
	DDPG (Lowe et al., 2017)	16.77%	1.5962	-0.1040	1.776
	SAC (Haarnoja et al., 2018)	19.87%	1.6097	-0.1152	2.176
	SARL (Ye et al., 2020)	22.06%	1.466	-0.1009	2.459
	FinBPM (Mean)	35.19%	<u>2.3000</u>	-0.0918	4.15
	FinBPM (Best)	**52.12%**	**2.7759**	-0.1010	**6.33**

Table 3: Performance comparison between FinBPM and baseline methods (mean of 5 runs). The best performing method for each metric is highlighted in bold font, while the second-best-performing method are underlined.

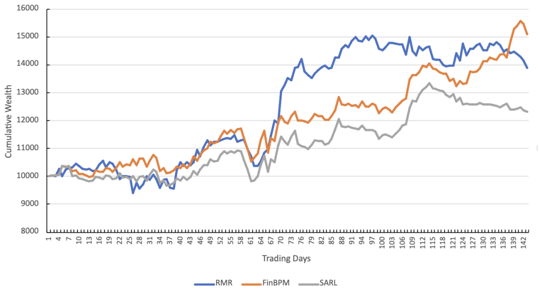

Figure 3: FinBPM cumulative return performance comparison with baselines.

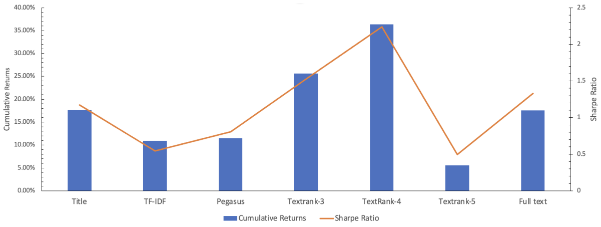

Figure 4: Comparison of results of different processing methods for financial news (TextRank-4 means that 4 sentences are filtered from the full text using the TextRank method).

tain the average. As shown in Figure 4, TextRank sentence selection of four sentences achieved the highest average return of 36.372% and Sharpe ratio. Headline-only and full-text approaches have similar gains, indicating headlines sufficiently capture key content while full-texts introduce more noise. Figure 4 gives that Pegasus underperformed, likely distorting the original text. For our dataset, TextRank filtering of full articles improved gain by retaining four salient sentences. Overall, selectively extracting important sentences with TextRank from full news text maximizes gains compared to headlines-only and indiscriminate full-text use. The results demonstrate properly filtering full articles can better leverage their information richness over headlines alone, while avoiding noise from irrelevant content.

7 Effect of News on Trading

In order to better observe the impact of news information on stock trading, we first traded a single stock. That is, there is no linkage between stocks, and only buy, hold, and sell operations are made for a certain stock. We selected eight target stocks such as AAPL and JPM for the experiment, and compared the impact of using financial news signals and not using them on stock buying and selling decisions. For each stock, we conducted five experiments with and without news data, and Table 4 shows the cumulative return results for these eight stocks in the test set. By observing the results in Table 4, it can be seen that in most cases, the cumu-

	Without News		With News(FinBERT)		With News(BERT)	
Stock	Max	Mean	Max	Mean	MAX	Mean
JPM	53.38%	32.12%	33.29%	27.77%	32.98%	22.75%
CAT	0%	0%	16.1%	6.25%	4.32%	3.19%
MMM	0%	0%	21.78%	17.80%	16.95%	13.43%
AAPL	8.18%	7.31%	8.94%	8.10%	8.02%	7.93%
DIS	0%	0%	48.34%	9.67%	42.39%	13.52%
GS	25.24%	12.60%	43.42%	32.49%	40.21%	23.73%
INTC	23.42%	17.05%	25.35%	24.58%	22.85%	17.32%
MSFT	10.74%	9.80%	13.77%	10.53%	11.37%	10.68%

Table 4: Single Trading For Certain Stock. 0% result represent at all test peroid the decision agent did not find any chance to buy this stock.

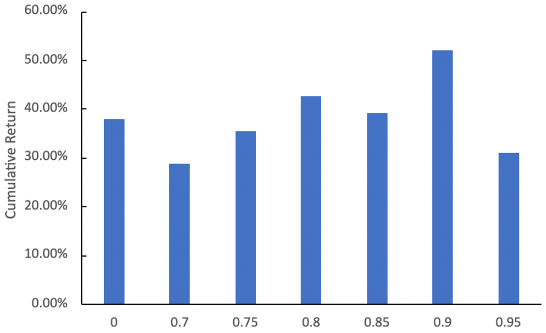

Figure 5: Performance of cumulative return under different k values.

lative return will increase after the introduction of news information. Among them, CAT and MMM default to no buy action when not using news indicators. After adding news information, their return rates are greatly improved. At the same time, compared with other stocks, the results of MSFT and AAPL did not improve much. By comparing the sentiment index distribution chart of individual stocks, it can be found that AAPL and MSFT have significant differences in sentiment index distribution compared with other stocks. The reason is that AAPL and MSFT are popular stocks with more news and attention, so combining news also brings more noise, which weakens the influence weight of sentiment index in the trading process.

Effect of α in Equation 3 : We also investigated the variation of cumulative return for three extreme values of α (i.e. The parameter controlling the contribution of news content in estimating overall index as described in Equation 3). The value of cumulative return with $\alpha = 0.1$ is 35.19%. and with $\alpha = 1$, it is 1.72% in our approach.

7.1 K-Filter Analysis:

Not all news can influence investor behavior (Chen and Huang, 2021), and language models still cannot fully capture semantics accurately (Yadav and Vishwakarma, 2020). Therefore, we implement a k-filter layer to refine the behavior index by retaining only the strongest signals most likely to impact investors. For our dataset and overall portfolio framework, setting $k = 0.9$ yields the optimal performance.

8 Conclusions and Future Work

In this research, we propose FinBPM, a portfolio management framework based on predicting investor behavior using reinforcement learning. It leverages linguistic models to comprehensively analyze intrinsic value information of companies from news text, and captures irrational volatility characteristics from price and trading volume data. Extensive ablation experiments are conducted to determine the optimal processing of financial text for maximizing intrinsic value signals under portfolio management. The results demonstrate the strengths of FinBPM in integrating textual and time series signals for investor behavior modeling and portfolio management optimization. Experiments show that FinBPM gains 13.26% returns over state-of-the-art models, while controlling maximum drawdown to 10.10%. In the future, we will support high-frequency trading strategy and explorer more financial market trading.

Limitations

The major limitations of FinBPM are twofold: (1) FinBPM operates on daily data, which can result in imprecise trading execution. Our future work will develop high-frequency trading module to overcome this limitation. (2) FinBPM currently only supports US stock market transactions and analysis of English financial news. In the future, more languages and more diverse financial market transactions (such as futures market, etc.) will be supported.

Ethical Considerations

We will discuss the ethical considerations and broader impact of this work here: (1) Fair competition. A trading system should not hide information. We evaluate FinBPM only on public data in highly regulated stock markets. We follow broad ethical guidelines to design and evaluate FinBPM, and encourage readers to follow both regulatory and ethical considerations pertaining to the stock market. (2) Intellectual property. We adhere to the original licenses for all datasets and models used. Regarding the issue of data copyright, we do not provide the original data and we only provide processing scripts for the original data. (3) Environmental Impact. The experiments are conducted on the GPUs. This results in a amount of carbon emissions. (4) Intended Use. FinBPM can be utilized to provide portifilo management advice for users. (5) Misuse risks. FinBPM should not be utilized for processing and analyzing sensitive or uncopyrighted data.

References

Yu-Fu Chen and Szu-Hao Huang. 2021. Sentiment-influenced trading system based on multimodal deep reinforcement learning. *Applied Soft Computing*, 112:107788.

Thomas M. Cover. 1991. Universal portfolios. *Mathematical Finance*, 1(1):1–29.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. 2014. Using structured events to predict stock price movement: An empirical investigation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1415–1425, Doha, Qatar. Association for Computational Linguistics.

Xin Du and Kumiko Tanaka-Ishii. 2020. Stock embeddings acquired from news articles and price history, and an application to portfolio optimization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3353–3363, Online. Association for Computational Linguistics.

Fuli Feng, Xiangnan He, Xiang Wang, Cheng Luo, Yiqun Liu, and Tat-Seng Chua. 2019. Temporal relational ranking for stock prediction. *ACM Trans. Inf. Syst.*, 37(2).

Fabio D. Freitas, Alberto F. De Souza, and Ailson R. de Almeida. 2009. Prediction-based portfolio optimization model using neural networks. *Neurocomputing*, 72(10):2155–2170. Lattice Computing and Natural Computing (JCIS 2007) / Neural Networks in Intelligent Systems Designn (ISDA 2007).

Ananya Divakar Goudar, Hema K S, Inchara T R, and Meghana Kalmat. 2022. Predicting stock market trends using machine learning and deep learning algorithm. *International Research Journal of Computer Science*.

László Györfi, Gábor Lugosi, and Frederic Udina. 2006. Nonparametric kernel-based sequential investment strategies. *Mathematical Finance*, 16(2):337–357.

Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. 2018. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 1861–1870. PMLR.

James B. Heaton, Nick G. Polson, and Jan Hendrik Witte. 2017. Deep learning for finance: deep portfolios. *Applied Stochastic Models in Business and Industry*, 33(1):3–12.

Ziniu Hu, Weiqing Liu, Jiang Bian, Xuanzhe Liu, and Tie-Yan Liu. 2018. Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, WSDM '18, page 261–269, New York, NY, USA. Association for Computing Machinery.

Dingjiang Huang, Junlong Zhou, Bin Li, Steven C. H. Hoi, and Shuigeng Zhou. 2016. Robust median reversion strategy for online portfolio selection. *IEEE Transactions on Knowledge and Data Engineering*, 28(9):2480–2493.

Bin Li and Steven C. H. Hoi. 2012. On-line portfolio selection with moving average reversion. In *Proceedings of the 29th International Coference on International Conference on Machine Learning*, ICML'12, page 563–570, Madison, WI, USA. Omnipress.

Bin Li, Peilin Zhao, Steven C. H. Hoi, and Vivekanand Gopalkrishnan. 2012. Pamr: Passive aggressive mean reversion strategy for portfolio selection. *Machine Learning*, 87(2):221–258.

Yang Liu, Qi Liu, Hongke Zhao, Zhen Pan, and Chuanren Liu. 2020. Adaptive quantitative trading: An imitative deep reinforcement learning approach. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(02):2128–2135.

Ryan Lowe, YI WU, Aviv Tamar, Jean Harb, OpenAI Pieter Abbeel, and Igor Mordatch. 2017. Multi-agent actor-critic for mixed cooperative-competitive environments. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Robert C. Merton. 1980. On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, 8(4):323–361.

Rada Mihalcea and Paul Tarau. 2004. TextRank: Bringing order into text. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pages 404–411, Barcelona, Spain. Association for Computational Linguistics.

Volodymyr Mnih, Adria Puigdomenech Badia, Mehdi Mirza, Alex Graves, Timothy Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. 2016. Asynchronous methods for deep reinforcement learning. In *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 1928–1937, New York, New York, USA. PMLR.

Seyed Taghi Niaki and Saeid Hoseinzade. 2013. Forecasting s&p 500 index using artificial neural networks and design of experiments. *Journal of Industrial Engineering International*, 9(1).

Mehtabhorn Obthong, Nongnuch Tantisantiwong, Watthanasak Jeamwatthanachai, and Gary Wills. 2020. A survey on machine learning for stock price prediction: Algorithms and techniques. In *Proceedings of the 2nd International Conference on Finance, Economics, Management and IT Business - Volume 1: FEMIB,*, pages 63–71. INSTICC, SciTePress.

Chong Oh and Olivia R. Liu Sheng. 2011. Investigating predictive power of stock micro blog sentiment in forecasting future stock price directional movement. In *International Conference on Interaction Sciences*.

Anastasiia Panchuk and Frank Westerhoff. 2021. Speculative behavior and chaotic asset price dynamics: On the emergence of a bandcount accretion bifurcation structure. *Discrete and Continuous Dynamical Systems - B*, 26(11):5941–5964.

Yixin Qin, Fengchen Gu, and Jionglong Su. 2022. A novel deep reinforcement learning strategy for portfolio management. In *2022 7th International Conference on Big Data Analytics (ICBDA)*, pages 366–372.

Paul A. Samuelson. 2015. *Proof that Properly Anticipated Prices Fluctuate Randomly*, chapter 2.

Ramit Sawhney, Arnav Wadhwa, Shivam Agarwal, and Rajiv Ratn Shah. 2021. Quantitative day trading from natural language using reinforcement learning. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4018–4030, Online. Association for Computational Linguistics.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal policy optimization algorithms.

William F. Sharpe. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk*. *The Journal of Finance*, 19(3):425–442.

Payal Soni, Yogya Tewari, and Deepa Krishnan. 2022. Machine learning approaches in stock price prediction: A systematic review. *Journal of Physics: Conference Series*, 2161(1):012065.

Ahmad Syifaudin, Yusuf Yusuf, Roni Mulyatno, and Benny Dhevyanto. 2020. Fundamental financial information as a signal of company value. In *Proceedings of the 1st International Conference on Accounting, Management and Entrepreneurship (ICAMER 2019)*, pages 22–24. Atlantis Press.

Zimu Wang and Hong-Seng Gan. 2023. Multi-level adversarial training for stock sentiment prediction. In *2023 IEEE 3rd International Conference on Computer Communication and Artificial Intelligence (CCAI)*, pages 127–134.

Yumo Xu and Shay B. Cohen. 2018. Stock movement prediction from tweets and historical prices. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1970–1979, Melbourne, Australia. Association for Computational Linguistics.

Ashima Yadav and Dinesh Kumar Vishwakarma. 2020. Sentiment analysis using deep learning architectures: A review. *Artif. Intell. Rev.*, 53(6):4335–4385.

Hongyang Yang, Xiao-Yang Liu, Shan Zhong, and Anwar Walid. 2021. Deep reinforcement learning for automated stock trading: An ensemble strategy. In *Proceedings of the First ACM International Conference on AI in Finance*, ICAIF '20, New York, NY, USA. Association for Computing Machinery.

Steve Y. Yang, Yangyang Yu, and Saud Almahdi. 2018. An investor sentiment reward-based trading system using gaussian inverse reinforcement learning algorithm. *Expert Systems with Applications*, 114:388–401.

Xuting Yang, Ruoyu Sun, Xiaotian Ren, Angelos Stefanidis, Fengchen Gu, and Jionglong Su. 2022. Ghost expectation point with deep reinforcement learning in financial portfolio management. In *2022 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*, pages 136–142.

Yi Yang, Mark Christopher Siy UY, and Allen Huang. 2020. Finbert: A pretrained language model for financial communications.

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. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(01):1112–1119.

Jingqing Zhang, Yao Zhao, Mohammad Saleh, and Peter Liu. 2020. PEGASUS: Pre-training with extracted gap-sentences for abstractive summarization. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 11328–11339. PMLR.

Rui Zheng, Shihan Dou, Songyang Gao, Yuan Hua, Wei Shen, Binghai Wang, Yan Liu, Senjie Jin, Qin Liu, Yuhao Zhou, Limao Xiong, Lu Chen, Zhiheng Xi, Nuo Xu, Wenbin Lai, Minghao Zhu, Cheng Chang, Zhangyue Yin, Rongxiang Weng, Wensen Cheng, Haoran Huang, Tianxiang Sun, Hang Yan, Tao Gui, Qi Zhang, Xipeng Qiu, and Xuanjing Huang. 2023. Secrets of rlhf in large language models part i: Ppo.

Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. 2021. Informer: Beyond efficient transformer for long sequence time-series forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(12):11106–11115.

A Decision Network Loss Function

In Equation 4, the mathematical formula for $r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{\text{old}}}(a_t|s_t)}$, π_θ represents the probability of the current policy network, at state s_t, outputting action a_t. $\pi_{\theta_{\text{old}}}$ represents the probability of the policy network, which has not undergone network parameter updates, outputting action "at" in state s_t. Similarly the mathematical definition of $\hat{A}\left(s_t, a_t\right)$ is given as follows $\hat{A}\left(s_t, a_t\right) = Q_\pi(s,a) - V_\pi(s)$. $Q_\pi(s,a)$ represents the value of the action performed on a in state s, $V_\pi(s)$ represents the expectation of the value of all actions in state s.