Applied Soft Computing 112 (2021) 107788

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Sentiment-influenced trading system based on multimodal deep reinforcement learning

Yu-Fu Chen [a], Szu-Hao Huang [b,*]

[a] Institute of Information Management, National Chiao Tung University 1001 University Road, Hsinchu, 300, Taiwan

[b] Department of Information Management and Finance, National Yang Ming Chiao Tung University 1001 University Road, Hsinchu, 300, Taiwan

ARTICLE INFO

Article history:
Received 26 February 2021
Received in revised form 17 July 2021
Accepted 4 August 2021
Available online 11 August 2021

Keywords:
Finance
Multimodal learning
Reinforcement learning
Sentiment analysis
Stock trading

ABSTRACT

Owing to advancements in deep learning, studies involving the use of deep learning techniques to solve investment decision-making problems are increasing. However, although numerous aspects of the stock market may affect trends in financial data, previous studies have only considered price fluctuations. Therefore, investors may lose out on profits because of the complicated financial market condition. In this study, a multimodal reinforcement trading system is developed, which makes use of three techniques: reinforcement learning, sentiment analysis, and multimodal learning. The agent considers not only the price fluctuations but also news information when making a trading decision. Multimodal learning which can merge different modalities of data to enhance the performance of the model, and sentiment analysis for understanding the sentiment of news are introduced. In addition, an influence model is proposed to enable our agents to gain special insights on the impact that news has on the market. The influence model considers the relationship between sentiment of news and time. The experimental results show that multimodal agents outperform price-concerned agents by at least 13.26%. Our experimental results also indicate that the proposed influence model has the ability to shape the impact of news on the stock market. The model can aid the multimodal agents in evaluating the status of the market. The proposed multimodal reinforcement trading system is demonstrated to be robust in an experiment involving different sectors and evaluations by using various measures. In addition, because the data used are public, investors seeking to profit can easily implement the results of this paper. Therefore, it can be used in advanced research and financial applications.

1. Introduciton

Investment decision-making is as a sophisticated process that is difficult to model because financial time series data contain noise. Trading problems can be viewed as dynamic decision-making processes that are difficult to analyze. Information must be obtained from the market, and the right course of action can be chosen. A Markov decision process in reinforcement learning (RL) is similar in concept to investment decision-making in financial data. Therefore, many scholars have used RL for resolving financial problems. Moody and Saffell [1] proposed a method called direct reinforcement based on recurrent RL (RRL). Their proposed model was able to simplify representation and avoid dimensionality problems with increased efficiency. Deng et al. [2] used a deep recurrent neural network (DRNN) for financial signal representation to capture the characteristics of the market and proposed a task-aware back-propagation method to overcome the vanishing gradient problem.

However, developing a learning-based approach to model and analyze financial data is quite difficult because the principles and rules of social science are always hidden and dynamic. In addition, most studies on the use of RL in the field of finance only consider price data in the model and ignore other information available from news outlets or social media; such information is a useful representation of real-world events.

Sentiment analysis has been employed to model investor behavior more accurately. Kearney and Liu [3] summarized the seminal findings on how textual sentiments affect individual, firm-level, and market-level performance. Chan and Chong [4] extracted insights from unstructured data to provide hints on the trends in financial markets. They proposed a sentiment analysis engine that used linguistic analysis based on grammar. Some other studies [5–7] have also used sentiment analysis for predicting the stock market and for other financial applications [8]. Ren et al. [9] integrates sentiment analysis into a machine learning method based on a support vector machine. Furthermore, the authors consider the day-of-week effect and develop reliable and realistic sentiment indexes. Maia et al. [10] introduces a multi-class sentiment analysis classifier that categorizes financial and

* Corresponding author.
E-mail address: szuhaohuang@nctu.edu.tw (S.-H. Huang).

https://doi.org/10.1016/j.asoc.2021.107788
1568-4946/

market events, which are described in words, into three labels, namely positive, neutral, and negative, as other models in the field do. Numerous studies have been conducted on the extraction of insights from textual data and the use of the sentiments in the data to predict trends in the stock market; however, few studies have been conducted for realizing dynamic trading using RL.

To overcome these problems mentioned, multimodal learning is employed to integrate different modalities, such as newspapers and social media, to model human psychology more accurately. In a 2020 study, Lee et al. [11] incorporate the information into the stock prediction models with the help of multimodal deep learning and achieved significantly improved performance in comparison with the late fusion and single modality models. Li et al. [12] proposes an event-driven multimodal LSTM model to address the heterogeneity of the sampling times in different modes and achieves performance superior to that of other baseline algorithms. Sawhney et al. [13] also present a multitask solution that utilizes domain-specialized textual features and audio-attentive alignment for predictive financial risk and price modeling. Multimodal learning is used to construct models that can process and gather information from multiple sources [14]. Because multimodal learning involves cross-modal learning, it has been successfully employed for visual question answering [15,16], disease diagnosis [17,18], problems and other applications [19,20]. By merging discrete event information and continuous price data, the agents in the proposed system consider not only financial time series data but also government policies and news. In addition, news and social media play important roles during financial crises. Therefore, in this study, a model is proposed for determining the importance of news and events and used them as multimodal sources to train the agent. Also, a multimodal reinforcement stock trading system is developed by using deep learning techniques and solved investment decision-making problems. The trading agent is expected to make sensible trading decisions with the information provided.

Many studies based on RL and multimodal analysis have been developed to be combined with sentiment analysis in this field. For instance, Ma et al. [21] propose an algorithm derived from deep reinforcement learning that automatically learns to extract optimal feature representation to reduce the visual discrepancies between images of these two modalities. Chen et al. [22] propose a novel deep learning architecture for multimodal sentiment analysis that can perform modality fusion at the word level. Nonetheless, few studies discuss investment problems. For example, Yuan et al. [23] explore the influence of multimodal social media data on stock performance and investigate the underlying mechanism of two forms of social media data, namely text and pictures. Research on sentiment analysis is extensive. For example, Agarwal et al. [24] focused on sentiment classification based on the Valence Aware Dictionary and Sentiment Reasoner (VADER) tool and demonstrated the effect of sentiment on the change in stock market prices in 2020. In a 2021 study, Aguilar [25] constructs a new newspaper-based sentiment indicator that enables the real-time monitoring of economic activity in Spain. Moreover, few studies research the relationship between sentiment and time and the difference between titles and articles of news stories. To the best of our knowledge, combining RL and news sentiment analysis for resolving financial problems and solving investment decision-making problems.

The major contributions of this work are as follows. First, a multimodal reinforcement trading system is presented which includes a deep neural network and multimodal deep recurrent neural network. The multimodal integration of price information with news sentiment enables the proposed system to make profits. The result can help model human behavior when making investment decisions and proved to be superior to the RL trading system which only consider price in six different industries. Moreover, the influence model can help agents to understand the impact of news, which increases their opportunities to make profits. Because financial news data are public, investors seeking to gain excess profit can easily implement the results of this paper. Finally, this is the first work to quantify the influence of news in the field of finance and combine the results obtained with RL.

The remainder of this article is organized as follows. In Section 2: "Related works", sentiment analysis, and multimodal learning on RL are reviewed. Section 3: "Deep Multimodal Learning Methods", introduces the framework of this study, and a detailed explanation of the framework is provided. In Section 4: "Influence Model for Sentiment Analysis", the influence model is developed in the study to simulate the impact that real-world events have on the market. The experiment details and a discussion of the results are presented in Section 5: "Experimental Results". Finally, Section 6: "Conclusion", provides conclusions and directions for future research.

2. Related works

This study builds on two bodies of literature: RL to solve financial problems and sentiment-based multimodal learning. In the first section, the concept of RL and its application in finance are briefly introduced. Then, an overview of multimodal learning and introduce studies on sentiment analysis are presented.

2.1. Reinforcement learning in finance

RL is a branch of machine learning inspired by behaviorist psychology. The concept of RL is as follows. An agent exists in an environment, and its goal is to maximize long-term reward. During the training process, the agent observes the changes in states and chooses an action to execute. After choosing the action, a reward is received based on the action. In addition, the agent moves to the next state in the environment. This sequential decision-making process features a trade-off between short-term and long-term reward maximization. In recent years, many researchers have used RL in finance because the sequential decision-making process is similar to the investment behaviors of humans. Chen and Wang [26] designed a hybrid trading system based on genetic network programming (GNP). They combined GNP with RL to formulate effective trading strategies. Xiong and Liu [27] used deep RL to develop a stock trading strategy. The proposed approach outperformed the traditional min-variance portfolio allocation strategy in terms of the Sharpe ratio and cumulative returns. Li et al. [28] used stacked denoising autoencoders (SDAEs) and long short-term memory (LSTM) as parts of the function approximator to aid the trading agent. Conegundes et al. [29] beat the Brazilian stock market with a deep reinforcement learning (DRL) Day trading system. Thibaut et al. [30] apply a deep Q-learning method in DRL to develop algorithmic strategies.

2.2. Sentiment-based multimodal learning

Multimodal learning is a novel concept in the field of machine learning. In multimodal learning, models process and gather information of multiple modalities [14]. In this study, the joint representation and multimodal fusion techniques are focused in this study. Joint representation is achieved by concatenating multiple modalities with feature vectors or neural networks. Ngiam et al. [31] proposed using an autoencoder in a multimodal domain. They used a stacked autoencoder to deal with each modality and fused the modalities into a multimodal joint representation by using another autoencoder. In multimodal fusion, information from multiple modalities is integrated with the goal of predicting outcomes. The two main approaches to multimodal fusion are the model-based and model-free approaches. The model-free approach which does not directly rely on a specific machine learning method is employed in the present study. Fusion strategies include early fusion, late fusion, and hybrid fusion [32]. Among them, Hybrid fusion combines outputs from early fusion and late fusion, and it combines the advantages of the two approaches. This approach is used for multimedia classification, event detection [33], and audiovisual speaker identification [34]. As an aside, in the field of multimodal analysis, reinforcement learning is usually adopted. For instance, Qureshi et al. [35] developed a robot that could learn basic interaction skills successfully with multimodal deep reinforcement learning.

Sentiment analysis is the field of research in which the affective states and subjective information in a text unit are extracted, identified, quantified, and studied [8]. Sentiment analysis employs natural language processing [36], text analysis, computational linguistics, and biometrics. The primary aim is to determine the attitude toward a subject from a text and the overall polarity or emotional reaction to a document or event. It has been broadly used with textual data such as movies, product reviews, and survey responses. In finance, numerical data as well as analyst reports and earnings press releases have to be analyzed. Oliveira et al. [37] used StockTwits, a specialized stock market microblog, to create stock market lexicons based on statistical measures. These lexicons were useful for measuring investor sentiment and for producing Twitter investor sentiment indicators that have a moderate correlation with other well-known indicators. Tsai and Wang [38] used regression and ranking approaches to analyze risk from the financial reports of companies. The experimental results showed that financial sentiment words are highly correlated with risk prediction. Some researchers have assumed that trends in stock price movement are influenced by news events. In a 2020 study, Agarwal [24] focus on sentiment classification and demonstrated its effect on the change in stock market prices. In a 2021 study, Aguilar et al. [25] propose a newspaper-based sentiment indicator that enables the real-time monitoring of economic activity in Spain. The indicator not only outperforms the popular economic sentiment indicator of the European Commission but also performs excellently in nowcasting the Spanish gross domestic product (GDP).

Existing studies on RL, multimodal learning, and sentiment analysis have some limitations. First, in most studies on RL in finance, only price data have been considered. In other words, other sources of information such as news and expert opinions have been ignored. Pilvere-Javorska et al. [39] analyze the absolute value indicators of 510 listed companies in Nordic countries by using factor and cluster analysis, and they compare their results with a similar analysis applied to the Baltic states. Tiganescu et al. [40] explore the potential of various dynamic identification methods to provide information on building state changes. Seyedimany et al. [41] investigates the stock price reactions to special dividend announcements. Those mediums may be very useful in decision-making. Second, although the accuracy in predicting stock market trends when using sentiment analysis to extract insights from textual data is reasonable, it is not suitable for dynamic trading using RL. Dynamic trading is essential in trading because incorrect or improper trading decisions may lead to substantial transaction costs. Finally, multimodal learning has not been used in the field of finance thus far.

In light of the issues mentioned in the previous paragraph, RL and sentiment analysis are integrated with multimodal learning, which would enable agents to refer news sentiments and make dynamic trading decisions based on the market price conditions.

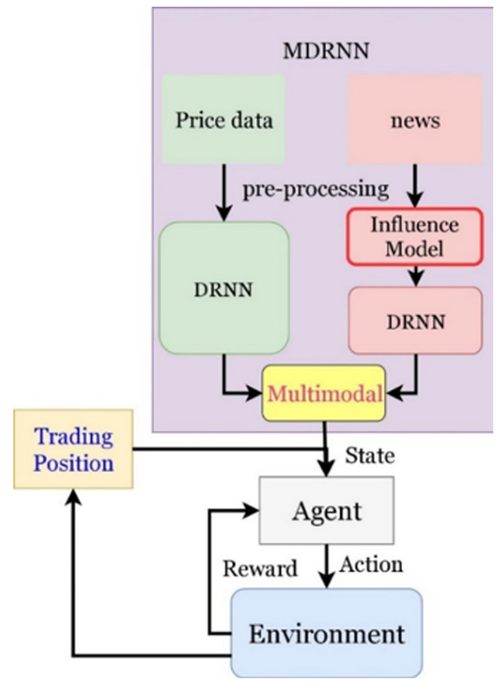


Fig. 1. Structure of the multimodal reinforcement trading system.

Furthermore, an influence model based on news propagation is proposed to measure the impact of news on the stock market in the period. A simulation of the influence of news on the stock market will enhance the correlation between sentiment and financial time series data.

3. Deep multimodal learning methods

In this section, the details of the proposed multimodal reinforcement trading system are described. First, the frameworks of the two basic models – RRL [1] and DRNN [2] – used in this study are reviewed. Although RRL is useful in solving the nonlinear trading problem, it cannot operate with feature learning. Thus, a DRNN model is proposed to solve the aforementioned problem. Furthermore, news information as multimodal features is introduced into the original DRNN model. Then, the multimodal DRNN (MDRNN) model is presented.

3.1. System overview

In this section, a brief overview of our multimodal reinforcement trading system is provided, including the deep recurrent network and the multimodal deep recurrent network. In contrast to the systems proposed in our previous studies, the proposed system combines data from multiple modalities, such as price data and news data, to construct a multimodal RL framework. Our system is divided into two parts. In the first part, data from different modalities are extracted, and, in the second part, the RL algorithm is executed. In the first part, modalities including price data and news data are processed by separate neural networks. As mentioned earlier, an influence model to measure the impact of news is proposed. In the second part, the extracted features are merged by using early fusion; the features are fused into a joint representation before classification. The agent makes trading decisions based on this representation and receive corresponding rewards. The structure of the multimodal reinforcement trading system is shown in Fig. 1.

3.2. Multimodal deep recurrent neural network

In this section, the review of the partial frameworks provided in [1] and [2] is taken because the proposed MDRNN is based on these frameworks. RRL uses the return of price time series data at different times t with the goal of maximizing the accumulated rewards U_T, as shown in Eq. (2) [2], where θ represents the parameters of the proposed model. In other words, the state of the RRL model indicates price sequences, and the action of the RRL model represents trading decisions, including the long, neutral, and short positions. A long position refers to the purchase of financial products, whereas a short position refers to the sale of financial products. A neutral position is one in which nothing is done with the assets. The reward in RRL is a profit or loss generated from the trading action. Eq. (1) [2] represents a single reward, where δ_t denotes the trading decision, z_t represents the return of price data, and c denotes the trading cost which takes not only transaction cost but also slippage cost into consideration.

$$R_t = \delta_{t-1} z_t - c|\delta_t - \delta_{t-1}| \tag{1}$$

$$\max_{\theta} U_T\{R_1, \dots, R_T|\theta\} \tag{2}$$

The major contribution of RRL is the application of a nonlinear function δ_t as the trading action, which is better than using a value function, which could result in efficiency problems. The DRNN, on the other hand, stated that RRL was lack of the process of feature learning and improved the problem by introducing deep learning techniques to do dynamic trading. The difference between RRL and a DRNN is that a DRNN extracts features from the original input data. With the assistance of deep learning, feature vectors can resist noise from the market because they are no longer directly established from the return.

In this study, the MDRNN framework based on a DRNN is presented. The MDRNN learns not only from the fluctuations in stock price but also from other information such as financial news. Some studies have investigated the relationship between news and price fluctuation. Negative news encourages individuals sell stock, whereas positive news prompts individuals to buy stock. The following assumptions are made in our study.

- Assumption 1: Actual events that impact companies are presented in financial news reports; these events often reflect the state of the companies, and they may impact investor confidence.

Financial news and stock price are assumed related, and the MDRNN is based on this relationship. However, because news and price are of different data types, they cannot be directly learned together. Therefore, the feature learning part is split into two channels and merged later. To make the news information learnable, Valence Aware Dictionary for sEntiment Reasoning (VADER) is used [42], which converts news texts into sentiment polarity scores. VADER combines lexical features and considers grammatical and syntactical conventions for expressing and emphasizing sentiment intensity. It can not only distinguish a single vocabulary but also handle complete sentence structures to provide precise sentiment evaluation. The output sentiment scores lie in the range $[-1, 1]$ (most negative to most positive). Through this method, textual data is converted into numeric data in the form of sentiment scores.

- Assumption 2: The sentiment scores of news items are additive and each has equal influence. Each news sentiment is assumed to have the same degree of influence on price and that they can be added.

Due to the fact that news is released randomly when an event occurs in the real world. Sometimes, news on companies may be reported only once a week. At other times, several news reports on a company may be published on the same day. To overcome the sample rate problem, the news is sorted by date and added up. Assumption 2 is inspired by Hawkes process [43] which supposes that past events can temporarily increase the probability of future events, assuming that such excitation is positive, additive over the past events and exponentially decaying with time. Because the traits of the Hawkes process are similar to sentiment-driven investor responses and may be useful to building up the structure [44], Hawkes process methods are adopted and sentiment scores are assumed to be additive and influential on future price trends.

Sentiment scores are assumed to be additive and have an influence on future price trends. Based on Assumption 2, sentiment scores are firstly collected according to the date on which the news containing company information was published or reported. Then, determining whether a company and piece of news are related using two approaches. The first approach is to check whether the company's name appears in the title of the news article; which may be rough but topical and strongly related to the company. The other approach is to check whether the company's name appears at least once in the article; this may be a more comprehensive approach, but also noisier because the related company may not play a main role in the news. Then, these sentiment scores are converted into a sentiment feature vector S_t using the descriptive statistic of the day, which summarizes the central tendency, dispersion, and shape of the distribution. The sentiment feature vector is given in Eq. (3):

$$S_t = [count, mean, std, min, Q_1, median, Q_3, max] \tag{3}$$

where *count* is the number of related news items in a given day; *mean* and *std* are respectively the arithmetic mean and standard deviation of the related news' sentiment scores; *min*, $Q1$, *median*, $Q3$, and *max* are the minimum, first quartile, median, third quartile, and maximum sentiment scores, respectively, for a given day. Next, learning sentiment features and combining them with the features obtained from the DRNN of the original price. The sentiment scores for each day are regarded as the sentiment influence sequence that represents the company sentiment trend. A long short-term memory (LSTM) is used to model the sentiment influence series and adopted early fusion to merge the sentiment and price parts as a multimodal vector M_t, as shown in Eq. (4), where f_t denotes recent return values.

$$M_t = [LSTM(f_t), LSTM(S_t)] \tag{4}$$

The correlation and interactions are exploited between price and news sentiment features using the early fusion method instead of the late fusion method because late fusion outputs two decisions simultaneously. Therefore, early fusion is chosen to merge information from different modalities and allow agents to make trading decisions based on the multimodal features. This also allows our agents to adjust the proportion between news and price information dynamically for the trading action. The trading action is given by Eq. (5), where w, u, and b are the parameters of the proposed MDRNN.

$$\delta_t = tanh\,[\langle w, M_t\rangle] + b + u\delta_{t-1} \tag{5}$$

The MDRNN uses the joint representation Mt to replace the feature vectors that represent the market condition and finance news opinions in a DRNN. This not only makes the MDRNN receive the market state and news sentiment simultaneously but also makes the agent behave like an experienced investor.

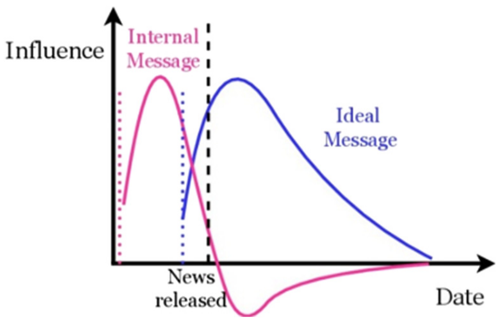


Fig. 2. Influence curve of ideal message and internal message.

4. Influence model for sentiment analysis

In this section, a method for modeling the multimodal influence is proposed based on the propagation phenomena of real-world event information. The proposed method is hoped to catch the influence of events which differs as time goes by. Then, the multimodal influence is applied to the MDRNN to verify whether it can help agents gain better insights of the market.

4.1. News influence model

First, a news influence model is introduced to overcome the issues associated with the proposed system. Then, using multimodal information to generate better outcomes because the propagation and influence of information in modern human society are complex. As a result, an influence model is constructed based on some assumptions related to journalism and communication. These observations facilitate the quantification of the propagation and influence of news.

- Assumption 1: The time when news is released, but not the occurrence of it, anchors the events. In addition, information regarding an event takes time to reach people, and the influence on price increases when more people react to events and decreases over time.
- Assumption 2: The influence of events may start before the news release; therefore, the influence must be modeled backward from the anchor point.

Some events are only by specific people before the news is made public. Investors with such insider information may be able to make substantial profits. This phenomenon also follows the previous assumptions but the inconsistent order of communication in the influence of events may have an impact on stock price before the news release.

Then, summarizing the assumptions above and discussing the difference between an ideal message which is publicly disclosed. And an internal message that does not disclosed in public and is related to insider trading [45] as defined in Fig. 2.

The blue curve is the ideal message influence curve. The dotted blue line indicates when an event occurs; the influence of information starts to increase from this point. The dotted black line indicates when the news is released; the influence of information continues to increase at this point. Over time, the influence starts to decrease until there is no influence of the events. An internal message has a different process flow compared with an ideal message and a different degree of influence. The pink curve in Fig. 2. indicates the influence of news when it is released; news can even have a negative influence after it is released. The impact of internal messages on the market is influenced by insiders and overreact after the news released. Furthermore, the distribution of influence is believed as an indicator of corporate information disclosure.

Therefore, our influence model focuses on the date on which news is released. The influence impacts the same range forward and backward from the anchor date. To calculate the total influence of days following Assumption 3, which states that each news sentiment has equal influence and is additive. The influence of news is given by Eq. (6).

$$\gamma_t = b + w_{t-r}S_{t-r} + \cdots + w_{t+r}S_{t+r} \tag{6}$$

where t is the date that the news is released, r is the range of days that a piece of news or an event has an influence on market price, w is the weight of the influence of the news on market price, S is the sum of that day's sentiment scores, and b is the bias of the influence. Based on our definition of influence, influence is considered as a type of drive impacting the market price.

Next, the learning target is shown. In technical analysis, support and resistance prevent the price of stocks from being pushed in a certain direction, analogous to a force in physics. Extending the concepts of support and resistance, the percentage change in price is chose as the learning target because it has a similar effect as influence; this is analogous to acceleration in physics. The percentage change in price in the time series $P = \langle p_1, p_2, \ldots, p_t \rangle$ is given by Eq. (7).

$$\rho = \frac{p_t - p_{t-1}}{p_{t-1} * 100} \tag{7}$$

To verify the presence of a correlation between influence and the percentage change in price, the following statistical parameters are used to measure the performance of the model: mean square error (MSE), mean absolute percentage error (MAPE), and R squared (R^2).

The mean square error (MSE) is used to estimate the difference between influence and the percentage change in price. The formula of the MSE is given in Eq. (8), where n is the length of the influence series. Thus, the influence model is a linear regression model.

$$MSE = \frac{1}{n}\sum_{t=1}^{\infty}(\rho_t - \gamma_t)^2 \tag{8}$$

The mean absolute error (MAE) can be used to account for outliers and scale, perhaps making it more reasonable to use compared with the MSE. Therefore, the accuracy (expressed as a percentage) of the model, that is, the MAPE, indicates the quality of the prediction result. The formula for MAPE is provided in Eq. (9). As is the case in Eq. (8), n is the length of the influence series.

$$MAPE = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{\rho_t - \gamma_t}{\rho_t}\right| \tag{9}$$

In addition, R^2, which is used to explain the strength of the linear relationship between dependent and independent variables, is considered. The formula for R^2 is expressed in Eq. (10).

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(\rho_t - \gamma_t)^2}{\sum_{t=1}^{n}(\rho_t - \bar{\gamma}_t)^2} \tag{10}$$

Two types of sentiments — article sentiments and synthesized sentiments are adopted for the model influence. Kardinata et al. [46] perform an ontology-based sentiment analysis on the basis of news titles to determine whether GE received a positive or negative response in Indonesia in 2019 by using different instruments. Cui et al. [47] demonstrate the effectiveness of the proposed method in detecting fake news. Synthesized sentiments contain news titles and articles related to companies, whereas article sentiments only contain news related to companies. In addition, considering the aforementioned studies, news titles are argued greatly influence public sentiment. The results of the experiment regarding the influence scores in Section 5.2 (Figs. 6 and 7) also indicate that the news title has a larger influence than the news article does. Thus, because news titles are believed to be more influential than articles, the sentiment scores are combined with a titles-to-articles ratio of 0.6 : 0.4. Then, building the influence with six companies and different influence days to choose a suitable influence-day range. An influence-day range of ± 6 was chosen because of the assumption that the influence of news does not last for more than 2 weeks. The MSEs of the six companies are shown in Fig. 3

As shown in Fig. 3, the MSE of influence decreases when the influence-day range increases. The proper number of influence days must be determined from the variation in MSEs. Companies should not be selected with the least variation in MSEs, but with the least variation in MSEs between different days. First, focusing on the article influence (Fig. 3(a)), which is shown to find out the proper setting of influence day. In GE and Google (GOOG), a considerable drop in the MSE is observed from ± 3 days to ± 4 days; however, not much variation in the MSE is observed subsequently. The MSEs of MSFT, AMZN and CVX are almost constant after ± 4 days, respectively, whereas the MSE of MCO drops from ± 4 days to ± 5 days. As a result, the influence-day range is set as ± 5 days because the MSEs for most companies become steady after ± 4 days. Next, observing synthesized approaches for determining influence (Fig. 3(b)). The variation in synthesized MSEs is different, but the conclusion based on these observations is the same. Then, the influence day range is set to ± 5 days. This setting appears appropriate for the influence that MSEs become steady; this means that influence is somewhat similar to the percentage change in prices.

To ensure that the influence day range setting is robust, the MAPE, which is sensitive to outliers and scale, is adopted. R^2 is also adopted to measure the different types of influence that change over time. The results of MSE, MAPE, and R^2 are presented in Tables 1–3. The first part of Table 1 displays the MSE results, which are consistent with the results presented in Fig. 3. The second and third parts of Tables 2 and 3 present the MAPE and R^2 results. According to Tables 1–3, regardless of which measurement (MSE, MAPE or R^2) is selected, the same conclusion can be used. In most of the companies, the values of the different measurements change slightly and stabilize after ± 4 or 5 days. Thus, the influence day range setting is considered robust.

4.2. Integration of the news influence model into the MDRNN

To verify the effectiveness of the influence model, it was merged with RL in the MDRNN framework and compared the performance with that of the original MDRNN to determine whether the proposed model provides agents with more insights on the market. Then, concatenating the vectors from each channel to form a multimodal influence vector Mi_t, as shown in Eq. (11)

$$Mi_t = [LSTM(f_t), LSTM(\rho_t)] \tag{11}$$

In contrast to the original multimodal vector M_t, the multimodal influence vector Mi_t includes the propagation of news influence and the impact of news instead of simply the sentiment scores on the day when news is released. Based on this vector, the RL agents judge the situation and make appropriate trading decisions. The trading action is now given by Eq. (12)

$$\delta_t = tanh\,[\langle w, Mi_t\rangle] + b + u\delta_{t-1} \tag{12}$$

This action function is similar to the action function of the previous MDRNN; however, the influence model has more specific inputs from the status of news. The agent receives feature vectors that contain price fluctuations and news influence, and appropriate decisions are made based on these vectors. The complete algorithm of the MDRNN stock trading framework with the influence model is given in Alg. 1. The steps involved in the operation of the multimodal reinforcement trading system are as follows. The proposed framework has two inputs. One is the historical stock price return, which reveals price trends, and the other is the influence that news has on the market. The output of the framework is the final cumulative return that agents make through the trading action, which can be replaced with other appropriate risk-adjusted returns. In the trading time from 1 to n, the MDRNN learns the price and influence of news separately from individual DRNNs; the features of price and news influence are fused into a joint multimodal influence vector Mi_t. The agent observes the multimodal feature vector and makes trading decisions. Subsequently, the environment gives a corresponding return to the agent, and the agent updates its cumulative reward. Finally, the agent receives its final reward and updates its trading rules based on the reward.

Algorithm 1 Framework of the MDRNN stock trading system with the influence model.

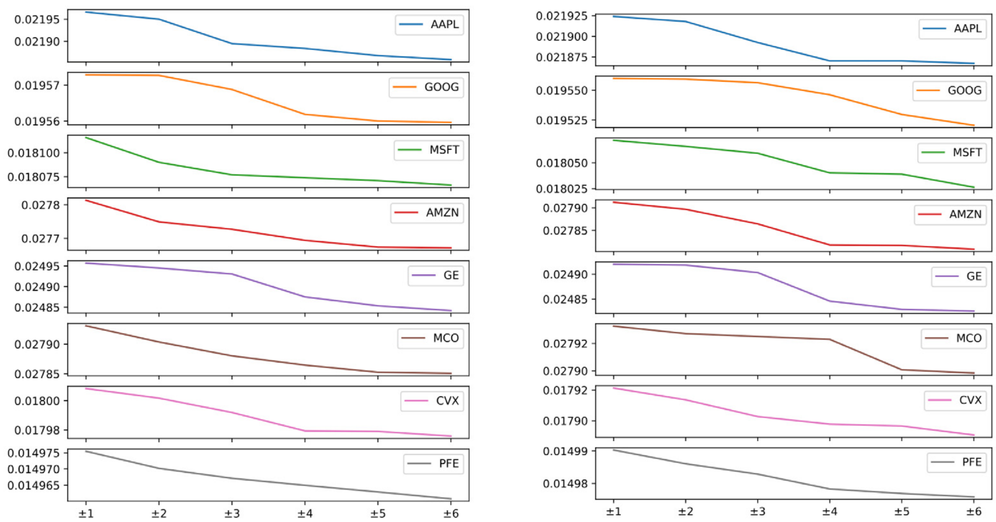
Input: $\{z_1, z_2, ..., z_n\}$:Historical stock price return;
$\{\gamma_1, \gamma 2, ..., \gamma n\}$:News influence toward the market;
Output: U_n:Final cumulative return;
1: Initialize $R_0 = 0$, $\delta_0 = 0$
2: **for** $t = 1, 2, ..., n$ **do**
3: DRNNs respectively learn on Z_t and γ_t
4: Early fusion into multimodal influence feature vector Mi_t
5: Agent based on feature vector make action δ_t
6: Realized the period return R_t and update cumulative return U_t
7: **end for**
8: Trading agents update its trading rules

5. Experimental results

In this section, an introduction of the used data and experiments conducted by the proposed method was given. First, a detailed description of the data used and the experiment settings was provided. Subsequently, our experimental results are presented to verify the effectiveness of our proposed multimodal reinforcement trading system.

5.1. Data description and experimental setting

Different modalities in the present study were adopted, namely price data and news data. The price data were obtained from Wharton Research Data Services (WRDS)'s NYSE Trade and Quote database, which contains intraday transaction data for all securities listed on the NYSE, AMEX, and NMS. The S&P 500 index was adopted because it comprises the leading 500 companies in the US stock markets. Furthermore, these companies have influence and popularity in their domain based on the Global Industry Classification Standard (GICS). After adjusting the data frequency from milliseconds to minutes, the data was incorporated for transactions that took place within regular trading hours into the model. The news data were obtained from the financial news datasets provided by Bloomberg and Reuters from 2006 to 2013. There were 450,341 news items from Bloomberg and 109,110 news items from Reuters. Also, 37 247 news items of titles of financial news are collected from Bloomberg from 2014 to 2020. The time information was maintained at the date level, and rearranged the news items according to trading days because real-world events that occur on weekends often influence price on the next trading day. Then, RL was applied. The following conditions were set:



(a) Article influence (b) Synthesized influence

Fig. 3. MSEs of different influence-day ranges for various companies.

Table 1
6-day MSE of price movement and sentiment measures (in 10^{-2}).

Types	Days	AAPL	GOOG	MSFT	AMZN	GE	MCO	CVX	PFE
Synthesized	±1	2.1924	1.9560	1.8072	2.7912	2.4920	2.7933	1.7921	1.4990
	±2	2.1918	1.9559	1.8066	2.7897	2.4918	2.7927	1.7914	1.4986
	±3	2.1893	1.9556	1.8059	2.7864	2.4903	2.7925	1.7903	1.4983
	±4	2.1870	1.9546	1.8040	2.7818	2.4846	2.7923	1.7898	1.4978
	±5	2.1870	1.9530	1.8039	2.7817	2.4829	2.7901	1.7897	1.4977
	±6	2.1867	1.9521	1.8026	2.7808	2.4826	2.7898	1.7891	1.4976
Article	±1	2.1966	1.9573	1.8116	2.7814	2.4957	2.7922	1.8008	1.4975
	±2	2.1950	1.9573	1.8090	2.7749	2.4945	2.7921	1.8002	1.4970
	±3	2.1895	1.9569	1.8077	2.7727	2.4931	2.7920	1.7992	1.4967
	±4	2.1884	1.9562	1.8074	2.7694	2.4875	2.7911	1.7979	1.4965
	±5	2.1868	1.9560	1.8071	2.7674	2.4854	2.7900	1.7979	1.4963
	±6	2.1859	1.9560	1.8066	2.7672	2.4842	2.7898	1.7976	1.4961

Table 2
6-day MAPE of price movement and sentiment measures.

Types	Days	AAPL	GOOG	MSFT	AMZN	GE	MCO	CVX	PFE
Synthesized	±1	40.4886	20.0601	38.6771	53.7004	33.2259	37.4133	73.3421	9.1161
	±2	44.7519	19.9143	39.4987	56.5720	35.0961	40.1594	74.2356	11.8619
	±3	49.9264	21.2169	39.2917	64.8060	43.9097	40.4738	73.9323	12.5982
	±4	51.9956	26.3206	46.5178	74.0800	71.1627	42.3216	76.9528	16.6301
	±5	51.9643	31.5870	49.3934	74.4321	76.6147	58.6910	78.9453	15.6675
	±6	51.1582	37.6457	53.2460	77.7539	80.0671	57.5272	79.0464	16.1729
Article	±1	43.3655	19.1967	27.1028	52.0980	27.2070	32.9483	46.4822	11.4582
	±2	53.1612	19.1116	39.4609	71.2576	32.4247	53.7344	49.3903	16.5412
	±3	64.6031	20.8174	42.6027	68.1125	39.9931	60.9448	50.1521	17.4378
	±4	69.8506	24.2064	40.7588	79.2379	66.8171	64.7589	54.8130	19.3716
	±5	77.2806	26.4260	41.0206	83.0965	80.2456	71.4454	54.5245	20.5184
	±6	77.4743	26.3998	39.9679	84.3514	83.9607	69.1149	53.7909	20.9713

- Agents only trade between 9:30 and 16:00.
- Agents make their trading decision at minute intervals.
- Agents can long or short one lot.
- The strike price is the closing price per minute.
- Agents have sufficient funds to perform a trading action.
- The capital is settled at the end of the trading day.
- The transaction cost is fixed at 0.005.

All neural networks were trained for 150 episodes at a learning rate of 0.01 by using the Adam optimizer. In the price-related network, the input size was four, comprising the opening, high, low, and closing price of stocks. The hidden unit size of the LSTM was set to 20. After making a trading decision, the agent used the closing price of that minute for the strike to receive the reward for the decision.

5.2. Influence model with multiple companies

However, certain influences of the multimodal framework could not be observed in the previous section. An attribution was made to improper modeling of the influence of multiple modalities on investment decisions. In the original MDRNN, sentiment scores at the date when news was released; however, there might be a doubt that details on when and how news was actually communicated to the public. To overcome this problem, a method was proposed to evaluate the influence of news on market price.

Table 3
6-day R^2 of price movement and sentiment measures (in 10^{-2}).

Types	Days	AAPL	GOOG	MSFT	AMZN	GE	MCO	CVX	PFE
Synthesized	±1	1.2216	0.1726	0.8150	0.8181	0.4896	0.2123	2.4789	0.0257
	±2	1.2769	0.1791	0.8794	0.9288	0.5045	0.2516	2.5624	0.0823
	±3	1.5054	0.2096	0.9525	1.1578	0.6245	0.2665	2.6811	0.1248
	±4	1.7051	0.3125	1.1605	1.4878	1.0832	0.2815	2.7347	0.1857
	±5	1.7052	0.4811	1.1742	1.4952	1.2157	0.4405	2.7481	0.2043
	±6	1.7320	0.5736	1.3130	1.5542	1.2433	0.4576	2.8117	0.2184
Article	±1	0.8416	0.0408	0.3312	1.5180	0.1945	0.2253	1.5298	0.2235
	±2	0.9867	0.0423	0.6144	1.9730	0.2901	0.4198	1.6015	0.2937
	±3	1.4849	0.0824	0.7572	2.1290	0.4064	0.5864	1.7086	0.3345
	±4	1.5829	0.1530	0.7907	2.3618	0.8489	0.6974	1.8459	0.3633
	±5	1.7291	0.1719	0.8239	2.5041	1.0214	0.7830	1.8498	0.3908
	±6	1.8107	0.1757	0.8749	2.5196	1.1134	0.7962	1.8847	0.4190

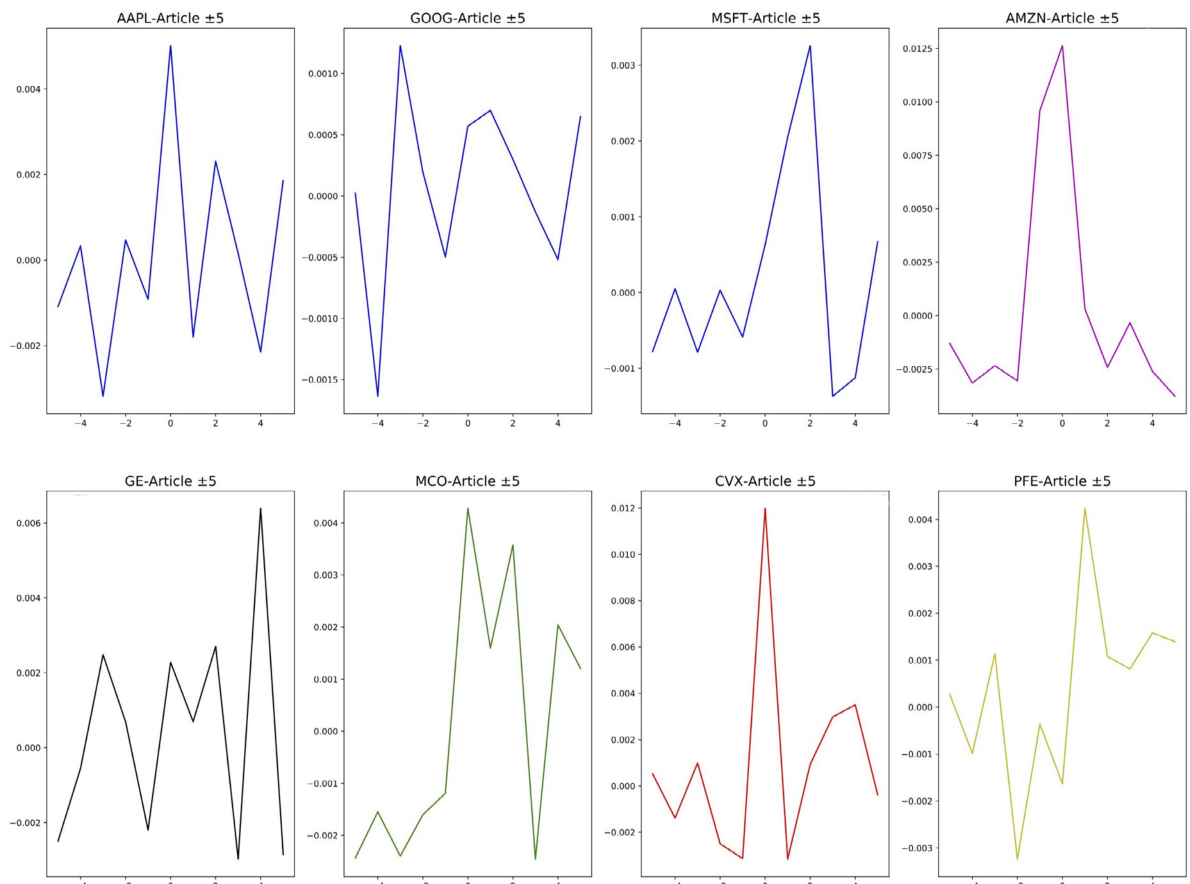


Fig. 4. Article influence distribution for different sectors in the S&P 500.

As discussed in Section 4.1, the influence-day range was set to ±5 days based on the variation in MSEs. To determine the distribution of influence, two types of influence distribution were observed in different sectors of the S&P 500 in Figs. 4 and 5. Each corporation has different influence distributions. Considering the analysis laid out in Section 4.1, an argument can be made that the influence distribution can be regarded as a property of the corporation, such as information disclosure. Insider trading [48] can be discovered by examining its influence distribution if its price always reflects the impact of news [49,50] long before the news is released. Leaders of stocks in six different detectors based on the GICS were selected, and each sector is illustrated by a different color. Blue represents the information technology sector, black represents the industrial sector, purple represents the consumer-related sector, dark green represents the financial sector, red represents the energy sector, and yellow represents the healthcare sector. The horizontal axis represents the range of day and the vertical axis shows the influence of different types of news.

Compared with the article influence distributions, the synthesized influence distributions were more concentrated on the day news was released and had a narrower range of values. The synthesized influence distributions for the technology sector and industrial sector appeared to fluctuate the most, whereas the energy sector and consumer-related sector exhibited the least variation in synthesized influence distributions. The article influence and synthesized influence distributions for the AAPL and GE were similar. Influence distributions provided us with an alternative means to investigate the relationship between the stock price and company news.

Before discussing the performance of the influence model in multimodal reinforcement trading, the article and title influence

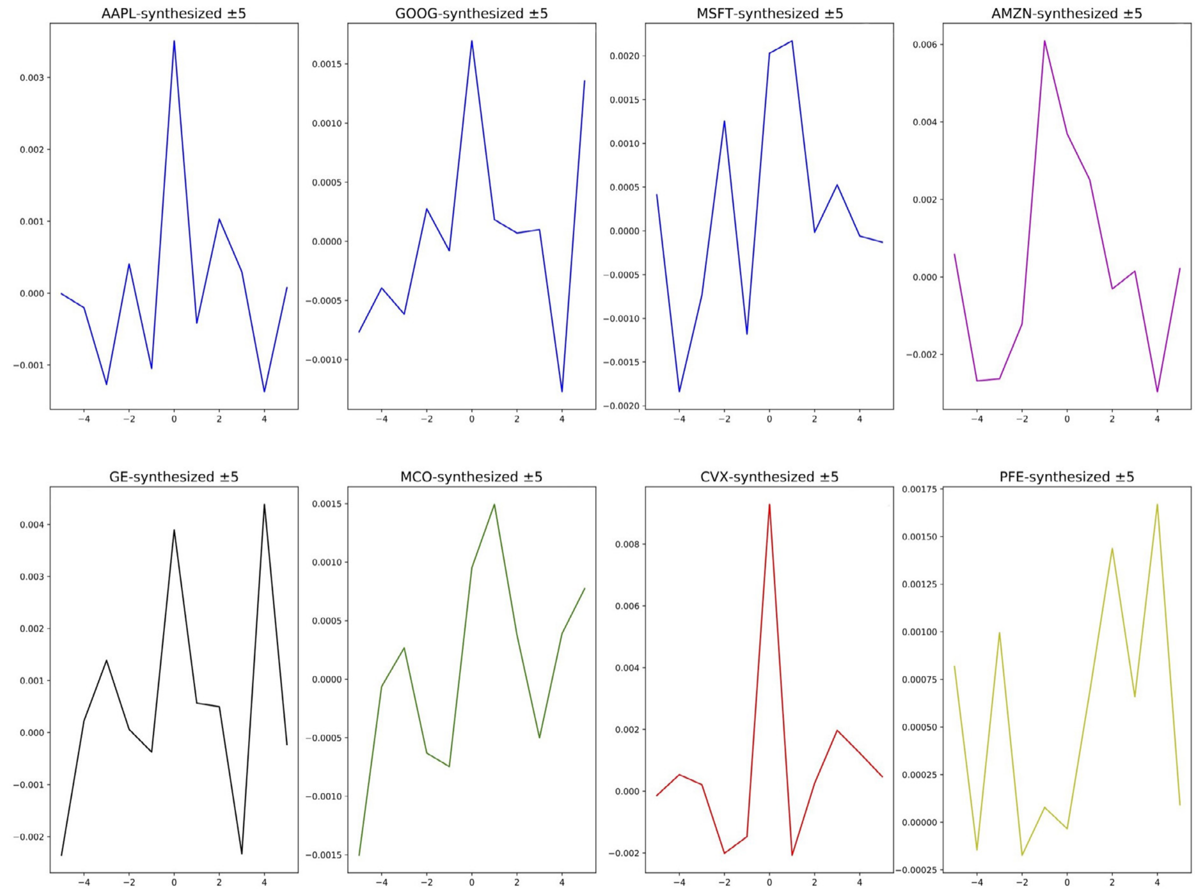


Fig. 5. Synthesized influence distribution for different sectors in the S&P 500.

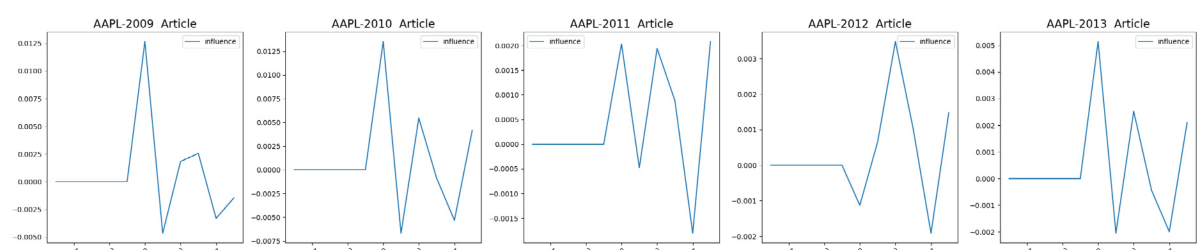


Fig. 6. Article influence distribution of AAPL used in reinforcement trading from 2009 to 2013.

distributions of AAPL 2009 to 2013 are present in Figs. 6 and 7. The influence ratio is presented on the y-axis in the figures. As shown in Fig. 6, all the influence distributions for AAPL had similar shapes. The influence was positive the day news was released; negative influence the following day may fix the overreaction of the previous day. In 2012, the influence of AAPL was quite different because the influence on the day news was released was negative and the influence looks like delay for one day to happen. This may be a focal point that the agents' performance with this article influence in 2012 on AAPL. As shown in Fig. 7, the influence distributions for titles are similar to those for articles, particularly in 2009; however, they still differ. Most of the synthesized influence distributions peak on the day news is released; however, the distribution sharply decreases after 2 days, increases on the next day, and falls again on day 4. Five days after the news release, the influence distribution increases again. This type of roller coaster shape is common in the influence distributions for news, although the influence does not decrease at the same time every year.

5.3. Performance of the influence model in multimodal reinforcement trading

In this section, the performance of the influence model in multimodal reinforcement trading was discussed. First, data covering 2 years was selected to establish the influence models in rolling and adjusted for the influence outputs. For example, price and news data from 2007 and 2008 was selected to build the influence model for reinforcement trading in 2009. The forward influence of news was erased to avoid look-ahead bias. The experiment was conducted to examine the effectiveness of introducing news sentiments in trading decisions. The training period was between 2007 and 2013. The rolling window approach was adopted to evaluate the performance of training and testing. The average profit points obtained for the test period with a fixed transaction cost of 0.005 are listed in Table 4.

The experimental results obtained in the previous section were used as a benchmark to verify whether the influence model could help agents to generate profit. Because of the limitation of

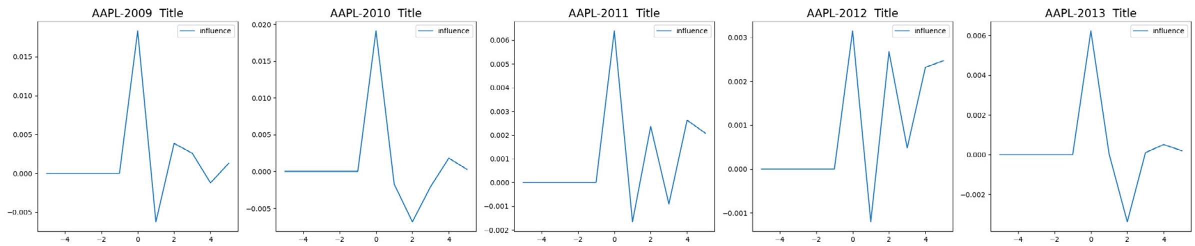


Fig. 7. Title influence distribution of AAPL used in reinforcement trading from 2009 to 2013.

Table 4
Comparison of various influence models of reinforcement trading in the information technology sector.

COST = 0.005		The annual profit point (%)				
Company	Type	2009	2010	2011	2012	2013
APPL	DRNN	11.2632%	10.4651%	18.4984%	130.3419%	29.9984%
	MDRNNT	12.3962%	11.8721%	**35.6084%**	119.2878%	26.9159%
	MDRNNA	11.8353%	7.8253%	26.6356%	**163.5456%**	31.7973%
	Influ_T	13.8051%	**17.1897%**	13.4845%	26.6769%	**52.7704%**
	Influ_A	10.9410%	10.1124%	10.0047%	28.3888%	35.6801%
	Influ_S	**19.0962%**	15.8009%	24.2656%	112.8342%	39.8262%
GOOG	DRNN	14.4255%	12.3436%	39.8729%	38.5812%	14.0174%
	MDRNNT	19.5308%	18.4508%	48.2641%	60.9802%	21.8241%
	MDRNNA	20.4464%	**21.1583%**	37.2920%	37.2704%	28.7837%
	Influ_T	20.5293%	11.3878%	13.7729%	57.5539%	34.3188%
	Influ_A	34.9218%	17.5405%	**89.6654%**	50.9337%	**55.1281%**
	Influ_S	**35.0111%**	14.0450%	76.8058%	**62.4491%**	43.9462%
MSFT	DRNN	5.6545%	4.3718%	8.4107%	8.4242%	8.4404%
	MDRNNT	7.2374%	7.7046%	9.2550%	12.3988%	12.9176%
	MDRNNA	5.6017%	7.5706%	8.7760%	8.8836%	9.2936%
	Influ_T	**22.3579%**	**9.5179%**	5.5437%	10.2606%	13.2768%
	Influ_A	15.4618%	9.3717%	9.8065%	10.0904%	12.9239%
	Influ_S	13.5674%	8.4910%	**11.8882%**	**14.3409%**	**21.4225%**

the dataset, the training and testing period was shortened (2009 to 2013) for the sake of influence production. In addition to the article influence and synthesized influence, title influence was investigated, which was obtained using the same settings used to obtain article influence. The average profit points obtained using the influence model in reinforcement trading are listed in Tables 4 (Information Technology) and 5 (other sectors).

To distinguish between the MDRNN using the influence model and the original MDRNN, Influ_T, Influ_A, and Influ_S represent MDRNNs using title, article, and synthesized influence, respectively. MDRNNT and Influ_T refers to the neural network that uses the sentiment scores of news article titles. MDRNNA and Influ_A refers to the neural network using the sentiment scores of article contents. Influ_S refers to the neural network using not only the sentiment scores of article contents but also the titles.

To ensure the robustness of the experiment, new data on financial news (2014 to 2020) are collected, and the results are listed in Table 6. In Table 6, only the title influence is considered because news articles are usually difficult to obtain and clean; therefore, only DRNN, MDRNN_T, and Influ_T are implemented here. In Table 6, AAPL and GOOG are selected from the S&P 500 index as trading targets because they are leaders in the information technology sector and more information is available on them compared with other companies.

First, the performance of the MDRNN with and without the influence model was compared. In the information technology (APPL, GOOG, MSFT), industrial (GE), financial (MCO), and health care (PFE) sectors, all the multimodal agents outperformed the DRNN agents in all test years. The Influ_T model outperformed the original MDRNNT in most test years for all sectors except the industrial (GE) sector. The experimental performance of Influ_A and MDRNN were similar. For all sectors and years, no particular type of influence model outperformed the others, but Influ_T and Influ_S usually outperformed Influ_A. Most of the MDRNNs with the influence models outperformed the original MDRNN; however, the DRNN performed well in the consumer-related sector (AMZN). This might be due to inappropriately related titles or because titles related to AMZN were unsuitable for determining the price trend.

The average profit points of different models were calculated and their performance was compared. On average, the original MDRNN and all influence models outperformed the DRNN; the best model, MDRNN_T, outperformed DRNN by 19.58%. To verify the effectiveness of the influence model, the MDRNN and the influence model were compared. In the information technology sector (APPL, GOOG, MSFT), although the goodness between Influ_T and MDRNNT was not obvious, Influ_S outperformed the MDRNNT by 59.34%. In addition, Influ_S outperformed the MDRNNA by 28.13%, demonstrating that sentiment influence models can help agents access better market insights and thus generate more profit. The results obtained for the information technology sector were consistent with those of most of sectors. However, for AMZN, the profits obtained using the MDRNNT and Influ_T model were less than those obtained using the DRNN. The MDRNNT and Influ_T model achieved 24.30% and 4.30% fewer points than the DRNN did, respectively. This comparison demonstrates that news title sentiment was unsuitable for assisting agents in profit generation. Even with the influence model, Influ_T yielded lower profits, indicating that some sentiments were unsuitable as a means to help agents.

The performance of the different models was compared for trading stocks in six sectors. The MDRNN outperformed the DRNN, and the influence models, particularly Influ_T, outperformed the original MDRNN. Furthermore, with the aid of an influence model, news sentiments can help agents understand the market and thus generating more profits.

Table 5
Comparison of various influence models of reinforcement trading in other sectors.

COST = 0.005		The annual profit point (%)				
Company	Type	2009	2010	2011	2012	2013
AMZN	DRNN	5.6599%	11.5285%	**22.1985%**	22.1985%	**21.9208%**
	MDRNNT	4.2110%	5.0559%	18.4693%	14.5504%	20.9268%
	MDRNNA	0.0184%	4.6899%	10.0139%	0.3069%	17.1070%
	Influ_T	4.7441%	**12.9686%**	14.7524%	**27.8211%**	19.6224%
	Influ_A	**23.3486%**	0.3900%	7.2396%	1.5178%	13.9217%
	Influ_S	4.6616%	7.0448%	14.0669%	8.3898%	10.3738%
GE	DRNN	1.5338%	1.6685%	4.5417%	3.2224%	−0.0636%
	MDRNNT	2.5604%	4.1838%	8.7828%	6.2166%	1.9672%
	MDRNNA	1.7228%	2.7684%	3.0078%	3.3338%	1.2648%
	Influ_T	−18.4913%	0.1727%	0.7356%	6.2849%	**3.3597%**
	Influ_A	**2.6756%**	3.1629%	**13.4742%**	7.4614%	1.8488%
	Influ_S	2.2769%	**5.1087%**	11.4319%	**8.1137%**	−0.0441%
MCO	DRNN	−0.0157%	0.0139%	−0.1624%	−0.0265%	−0.0195%
	MDRNNT	2.8781%	11.1906%	26.6891%	26.2669%	27.1926%
	MDRNNA	1.8437%	10.4899%	21.9866%	24.1675%	28.6115%
	Influ_T	**7.2989%**	**21.7187%**	**27.7149%**	27.3377%	**37.1425%**
	Influ_A	2.9290%	20.5196%	26.0133%	**29.8520%**	31.5003%
	Influ_S	0.0003%	−0.0083%	−0.0756%	6.8771%	−0.1209%
CVX	DRNN	19.8739%	25.6106%	**65.9395%**	43.9960%	35.5992%
	MDRNNT	32.0239%	28.7096%	60.8494%	55.7649%	33.0083%
	MDRNNA	22.5267%	26.2239%	49.6275%	55.6680%	32.9474%
	Influ_T	**33.0660%**	**42.8066%**	31.2926%	**64.7508%**	**43.6223%**
	Influ_A	30.5766%	32.4271%	1.6353%	57.0766%	29.1950%
	Influ_S	20.4261%	27.0593%	67.0132%	47.5695%	24.6960%
PFE	DRNN	5.9260%	2.2159%	6.8725%	7.6724%	7.3597%
	MDRNNT	4.5678%	1.4285%	5.6162%	6.6362%	6.6541%
	MDRNNA	5.8724%	3.1015%	6.5179%	7.2715%	7.1820%
	Influ_T	**14.0247%**	1.6570%	0.8313%	5.2184%	6.3651%
	Influ_A	6.1225%	2.4607%	7.7878%	7.1826%	6.9498%
	Influ_S	7.8650%	**4.1858%**	**9.6839%**	**9.4118%**	**9.0657%**

Table 6
Annual profit points obtained using the influence model for reinforcement trading, 2014–2020.

COST = 0.005	The annual profit point (%)					
Company	GOOG			AAPL		
Type	DRNN	MDRNNT	Influ_T	DRNN	MDRNNT	Influ_T
2014	22.6059%	18.9232%	**29.4664**%	−4.1674%	−3.0589%	**−0.2800**%
2015	−83.9292%	−53.9101%	**−26.9652**%	0.1172%	7.3758%	1**0.1375**%
2016	9.7030%	**38.9255**%	11.0247%	0.6874%	3.7951%	**6.0411**%
2017	−45.2615%	9.1132%	**73.0486**%	−1.1735%	**1.2718**%	0.3040%
2018	−91.5589%	−70.4893%	**−45.2615**%	0.0126%	0.8193%	**7.3076**%
2019	108.1357%	114.3067%	**148.2159**%	1.7829%	26.8960%	**51.5719**%
2020	11.2554%	71.1675%	**79.0040**%	−1.3636%	2.8761%	**4.2680**%

6. Conclusion

In this study, a deep multimodal RL trading system was developed based on RL, multimodal learning, and sentiment analysis to solve investment decision problems. The key contributions of this research are as follows. First, the multimodal integration of price information and news sentiment allowed the proposed RL agents to generate more profits. In this study, the multimodal deep recurrent networks were responsible for learning price and news information related to the market from various modalities. Besides, different types of sentiment were combined into the trading system which shows the novelty of this paper. The agents made trading decisions based on the joint representation, which simultaneously reflected the trends in price and news sentiment. The experiment results showed that multimodal agents performed better than the original agents by 44.69% and 38.55% for companies in Information Technology and for companies in other sectors, respectively.

Next, considering news publication and news propagation and establishing influence models to evaluate the how news impacts the market. The news publication day was used as the anchor point to measure the influence that real-world events have on the market. The consideration of the change of time is novel among the related researches. Through our experiments conducted with multiple companies in the S&P 500 index, ±5 days is found to be the adequate influence-day range. Different companies were found to have different types of influence distributions, which may be an indicator of a company's degree of information disclosure. Then, the influence model was incorporated into the original MDRNN. The influence model enhanced the RL agents' ability to generate profits. They were able to increase profits by 44.69% for companies in the information technology sector and by 22.15% for companies in other sectors compared with the original MDRNN. The results demonstrated that the news influence and propagation model enabled agents to more effectively comprehend the market and thus generate more profits, especially in the information technology sector. Companies in information technology, such as APPL and GOOG, are a frequent topic of public discussion. This may explain why the sentiment-based models outperformed others in this sector. In addition, because the data of financial news is public, it is not difficult for the investors who want to gain excess profit to implement this paper.

Despite the positive results achieved in this study, room for improvement remains. First, the approach which was used to convert news articles into sentiment scores was developed for use in social media research rather than finance. More suitable methods may be available for converting financial news into sentiment scores. Second, only price and financial news was input to the proposed model as features. Other information such as social media data and government policies should also be considered as inputs to the MDRNN model. If these inputs can be adequately quantified, the RL agents may be able to consider more comprehensive market situations. Third, total profit was used as a reward function to update the MDRNN. Other complex reward functions that consider risks, such as the Sharpe ratio, could be adopted to measure the performance of agents. Finally, many RL algorithms, such as deep deterministic policy gradient and actor-critic could be applied. In future work, a goal was established to apply different updating rules to allow agents to generate more profits under the same structure of the proposed system.

CRediT authorship contribution statement

Yu-Fu Chen: Conceptualization, Methodology, Software, Investigation, Writing – original draft. **Szu-Hao Huang:** Conceptualization, Data curation, Writing – review & editing, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported in part by the Ministry of Science and Technology, Taiwan, under Contract MOST 110-2221-E-A49 -101 and Contract MOST 110-2622-8-009 -014 -TM1; and in part by the Financial Technology (FinTech) Innovation Research Center, National Yang Ming Chiao Tung University.

References

[1] J. Moody, M. Saffell, Learning to trade via direct reinforcement, IEEE Trans. Neural Netw. 12 (4) (2001) 875–889, http://dx.doi.org/10.1109/72.935097.
[2] Y. Deng, F. Bao, Y. Kong, Z. Ren, Q. Dai, Deep direct reinforcement learning for financial signal representation and trading, IEEE Trans. Neural Netw. Learn. Syst. 28 (3) (2016) 653–664, http://dx.doi.org/10.1109/TNNLS.2016.2522401.
[3] C. Kearney, S. Liu, Textual sentiment in finance: A survey of methods and models, Int. Rev. Financ. Anal. 33 (2014) 171–185, http://dx.doi.org/10.1016/j.irfa.2014.02.006.
[4] S.W. Chan, M.W. Chong, Sentiment analysis in financial texts, Decis. Support Syst. 94 (2017) 53–64, http://dx.doi.org/10.1016/j.dss.2016.10.006.
[5] J. Kalyani, P. Bharathi, P. Jyothi, et al., Stock trend prediction using news sentiment analysis, 2016, http://dx.doi.org/10.21474/IJAR01/3749, arXiv preprint arXiv:1607.01958.
[6] J. Li, H. Bu, J. Wu, Sentiment-aware stock market prediction: A deep learning method, in: 2017 International Conference on Service Systems and Service Management, IEEE, 2017, pp. 1–6, http://dx.doi.org/10.1109/ICSSSM.2017.7996306.
[7] R. Akita, A. Yoshihara, T. Matsubara, K. Uehara, Deep learning for stock prediction using numerical and textual information, in: 2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS), IEEE, 2016, pp. 1–6, http://dx.doi.org/10.1109/ICIS.2016.7550882.
[8] R. Feldman, Techniques and applications for sentiment analysis, Commun. ACM 56 (4) (2013) 82–89, http://dx.doi.org/10.1145/2436256.2436274.
[9] R. Ren, D.D. Wu, T. Liu, Forecasting stock market movement direction using sentiment analysis and support vector machine, IEEE Syst. J. 13 (1) (2018) 760–770, http://dx.doi.org/10.1109/JSYST.2018.2794462.
[10] M. Maia, A. Freitas, S. Handschuh, Finsslx: A sentiment analysis model for the financial domain using text simplification, in: 2018 IEEE 12th International Conference on Semantic Computing (ICSC), IEEE, 2018, pp. 318–319, http://dx.doi.org/10.1109/ICSC.2018.00065.
[11] S.I. Lee, S.J. Yoo, Multimodal deep learning for finance: integrating and forecasting international stock markets, J. Supercomput. 76 (10) (2020) 8294–8312, http://dx.doi.org/10.1007/s11227-019-03101-3.
[12] Q. Li, J. Tan, J. Wang, H. Chen, A multimodal event-driven lstm model for stock prediction using online news, IEEE Trans. Knowl. Data Eng. (2020) http://dx.doi.org/10.1109/TKDE.2020.2968894.
[13] R. Sawhney, P. Mathur, A. Mangal, P. Khanna, R.R. Shah, R. Zimmermann, Multimodal multi-task financial risk forecasting, in: Proceedings of the 28th ACM International Conference on Multimedia, 2020, pp. 456–465, http://dx.doi.org/10.1145/3394171.3413752.
[14] T. Baltrušaitis, C. Ahuja, L.-P. Morency, Multimodal machine learning: A survey and taxonomy, IEEE Trans. Pattern Anal. Mach. Intell. 41 (2) (2018) 423–443, http://dx.doi.org/10.1109/TPAMI.2018.2798607.
[15] H. Gao, J. Mao, J. Zhou, Z. Huang, L. Wang, W. Xu, Are you talking to a machine? dataset and methods for multilingual image question answering, 2015, arXiv preprint arXiv:1505.05612.
[16] H. De Vries, F. Strub, S. Chandar, O. Pietquin, H. Larochelle, A. Courville, Guesswhat?! visual object discovery through multi-modal dialogue, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 5503–5512, http://dx.doi.org/10.1109/CVPR.2017.475.
[17] M. Liang, Z. Li, T. Chen, J. Zeng, Integrative data analysis of multi-platform cancer data with a multimodal deep learning approach, IEEE/ACM Trans. Comput. Biol. Bioinform. 12 (4) (2014) 928–937, http://dx.doi.org/10.1109/TCBB.2014.2377729.
[18] H.-I. Suk, S.-W. Lee, D. Shen, A.D.N. Initiative, et al., Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis, NeuroImage 101 (2014) 569–582, http://dx.doi.org/10.1016/j.neuroimage.2014.06.077.
[19] M. SteadieSeifi, N.P. Dellaert, W. Nuijten, T. Van Woensel, R. Raoufi, Multimodal freight transportation planning: A literature review, European J. Oper. Res. 233 (1) (2014) 1–15, http://dx.doi.org/10.1016/j.ejor.2013.06.055.
[20] P. Modesti, A. Sciomachen, A utility measure for finding multiobjective shortest paths in urban multimodal transportation networks, European J. Oper. Res. 111 (3) (1998) 495–508, http://dx.doi.org/10.1016/S0377-2217(97)00376-7.
[21] K. Ma, J. Wang, V. Singh, B. Tamersoy, Y.-J. Chang, A. Wimmer, T. Chen, Multimodal image registration with deep context reinforcement learning, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2017, pp. 240–248, http://dx.doi.org/10.1007/978-3-319-66182-7_28.
[22] M. Chen, S. Wang, P.P. Liang, T. Baltrušaitis, A. Zadeh, L.-P. Morency, Multimodal sentiment analysis with word-level fusion and reinforcement learning, in: Proceedings of the 19th ACM International Conference on Multimodal Interaction, 2017, pp. 163–171, http://dx.doi.org/10.1145/3136755.3136801.
[23] H. Yuan, Y. Tang, W. Xu, R.Y.K. Lau, Exploring the influence of multimodal social media data on stock performance: an empirical perspective and analysis, Internet Research (2021) http://dx.doi.org/10.1108/INTR-11-2019-0461.
[24] A. Agarwal, Sentiment analysis of financial news, in: 2020 12th International Conference on Computational Intelligence and Communication Networks (CICN), IEEE, 2020, pp. 312–315, http://dx.doi.org/10.1109/CICN49253.2020.9242579.
[25] P. Aguilar, C. Ghirelli, M. Pacce, A. Urtasun, Can news help measure economic sentiment? An application in COVID-19 times, Econom. Lett. 199 (2021) 109730, http://dx.doi.org/10.1016/j.econlet.2021.109730.
[26] Y. Chen, X. Wang, A hybrid stock trading system using genetic network programming and mean conditional value-at-risk, European J. Oper. Res. 240 (3) (2015) 861–871, http://dx.doi.org/10.1016/j.ejor.2014.07.034.
[27] Z. Xiong, X.-Y. Liu, S. Zhong, H. Yang, A. Walid, Practical deep reinforcement learning approach for stock trading, 2018, http://dx.doi.org/10.2139/ssrn.3690996, arXiv preprint arXiv:1811.07522.
[28] Y. Li, W. Zheng, Z. Zheng, Deep robust reinforcement learning for practical algorithmic trading, IEEE Access 7 (2019) 108014–108022, http://dx.doi.org/10.1109/ACCESS.2019.2932789.
[29] L. Conegundes, A.C.M. Pereira, Beating the stock market with a deep reinforcement learning day trading system, in: 2020 International Joint Conference on Neural Networks (IJCNN), IEEE, 2020, pp. 1–8, http://dx.doi.org/10.1109/IJCNN48605.2020.9206938.
[30] T. Théate, D. Ernst, An application of deep reinforcement learning to algorithmic trading, Expert Syst. Appl. 173 (2021) 114632, http://dx.doi.org/10.1016/j.eswa.2021.114632.
[31] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, A.Y. Ng, Multimodal deep learning, in: ICML, 2011, pp. 689–696.
[32] P. Atrey, M. Hossain, A. El Saddik, M. Kankanhalli, Multimodal fusion for multimedia 2https, sites. google. com/site/fredyuhuiwang/home analysis: a survey, Multimed. Syst. 16 (6) (2010) 345–379, http://dx.doi.org/10.1007/s00530-010-0182-0.

[33] Z.-z. Lan, L. Bao, S.-I. Yu, W. Liu, A.G. Hauptmann, Multimedia classification and event detection using double fusion, Multimedia Tools Appl. 71 (1) (2014) 333–347, http://dx.doi.org/10.1007/s11042-013-1391-2.
[34] Z. Wu, L. Cai, H. Meng, Multi-level fusion of audio and visual features for speaker identification, in: International Conference on Biometrics, Springer, 2006, pp. 493–499, http://dx.doi.org/10.1007/11608288_66.
[35] A.H. Qureshi, Y. Nakamura, Y. Yoshikawa, H. Ishiguro, Robot gains social intelligence through multimodal deep reinforcement learning, in: 2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids), IEEE, 2016, pp. 745–751, http://dx.doi.org/10.1109/HUMANOIDS.2016.7803357.
[36] Y.A. Solangi, Z.A. Solangi, S. Aarain, A. Abro, G.A. Mallah, A. Shah, Review on natural language processing (NLP) and its toolkits for opinion mining and sentiment analysis, in: 2018 IEEE 5th International Conference on Engineering Technologies and Applied Sciences (ICETAS), IEEE, 2018, pp. 1–4, http://dx.doi.org/10.1109/ICETAS.2018.8629198.
[37] N. Oliveira, P. Cortez, N. Areal, Stock market sentiment lexicon acquisition using microblogging data and statistical measures, Decis. Support Syst. 85 (2016) 62–73, http://dx.doi.org/10.1016/j.dss.2016.02.013.
[38] M.-F. Tsai, C.-J. Wang, On the risk prediction and analysis of soft information in finance reports, European J. Oper. Res. 257 (1) (2017) 243–250, http://dx.doi.org/10.1016/j.ejor.2016.06.069.
[39] A. Pilvere-Javorska, I. Pilvere, European nordic countries stock market listed companies': Factor and cluster analysis approach, Emerg. Sci. J. 4 (6) (2020) 443–453, http://dx.doi.org/10.28991/esj-2020-01244.
[40] A. Tiganescu, B. Grecu, I.G. Craifaleanu, Dynamic identification for representative building typologies: Three case studies from bucharest area, Civil Eng. J. 6 (3) (2020) 418–430, http://dx.doi.org/10.28991/cej-2020-03091480.
[41] A. Seyedimany, Stock price reactions on NASDAQ stock exchange for special dividend announcements, Emerg. Sci. J. 3 (6) (2019) 382–388, http://dx.doi.org/10.28991/esj-2019-01200.
[42] C. Hutto, E. Gilbert, Vader: A parsimonious rule-based model for sentiment analysis of social media text, in: Proceedings of the International AAAI Conference on Web and Social Media, Vol. 8, 2014, p. 1.
[43] H. Mei, J. Eisner, The neural hawkes process: A neurally self-modulating multivariate point process, 2016, arXiv preprint arXiv:1612.09328.
[44] S.Y. Yang, A. Liu, J. Chen, A. Hawkes, Applications of a multivariate Hawkes process to joint modeling of sentiment and market return events, Quant. Finance 18 (2) (2018) 295–310, http://dx.doi.org/10.1080/14697688.2017.1403156.
[45] M.S. Rozeff, M.A. Zaman, Overreaction and insider trading: Evidence from growth and value portfolios, J. Finance 53 (2) (1998) 701–716, http://dx.doi.org/10.1111/0022-1082.275500.
[46] E.A. Kardinata, N.A. Rakhmawati, N.A. Zuhroh, Ontology-based sentiment analysis on news title, in: 2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT), IEEE, 2021, pp. 360–364.
[47] L. Cui, S. Wang, D. Lee, Same: sentiment-aware multi-modal embedding for detecting fake news, in: Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2019, pp. 41–48.
[48] R.K. Bakshi, N. Kaur, R. Kaur, G. Kaur, Opinion mining and sentiment analysis, in: 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), IEEE, 2016, pp. 452–455.
[49] P. Cziraki, E. Lyandres, R. Michaely, What do insiders know? Evidence from insider trading around share repurchases and SEOs, J. Corporate Finance (2019) 101544, http://dx.doi.org/10.1016/j.jcorpfin.2019.101544.
[50] Q. Xu, V. Chang, C.-H. Hsu, Event study and principal component analysis based on sentiment analysis–a combined methodology to study the stock market with an empirical study, Inform. Syst. Front. 22 (5) (2020) 1021–1037, http://dx.doi.org/10.1007/s10796-020-10024-5.