

## Research Article

# DeeBERT-S3WD: Three-Way Multigranularity Decision for Interactive Information Sentiment Analysis Research

Jun Wang <sup>1,2</sup>, Xiangnan Zhang,<sup>1,2</sup> Guanyuan Yu <sup>3,4</sup>, Yuexin Chen,<sup>1,2</sup> and Shiyuan Rao<sup>1,2</sup>

<sup>1</sup>Department of Artificial Intelligence, School of Economic Information Engineering, Southwestern University of Finance and Economics, Chengdu 611130, China

<sup>2</sup>Financial Intelligence & Financial Engineering Key Laboratory of Sichuan Province, Southwestern University of Finance and Economics, Chengdu 611130, China

<sup>3</sup>Joint Lab of Data Science and Business Intelligence, School of Statistics, Southwestern University of Finance and Economics, Chengdu 611130, China

<sup>4</sup>XWBank, Chengdu 611130, China

Correspondence should be addressed to Guanyuan Yu; yuguanyuan@swufe.edu.cn

Received 18 May 2022; Revised 25 August 2022; Accepted 10 September 2022; Published 10 October 2022

Academic Editor: Yuchen Li

Copyright © 2022 Jun Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Interactive information between listed companies and investors is an emerging form of information disclosure in the stock market and has a crucial impact on stock market analysis. Interactive information interferes with investors' decision-making by influencing their sentiments, significantly affecting the stock market's health. Due to the unique nature of interactive information, the existing approaches to dynamic interactive information sentiment analysis rarely consider a multistep study under the tradeoff of cost and accuracy. To address this problem, we propose a novel unified framework combining DeeBERT with the sequential three-way decision based on the early exiting mechanism and continually mining uncertain regions for interactive information sentiment analysis. Specifically, we treat the question-answer pair interaction information as an entire sample and leverage DeeBERT to allow the samples to exit early without traversing through all the Transformer layers. Subsequently, the uncertain samples are selected at each Transformer layer to be reinvestigated at the next granularity of time-evolving data. Besides, we utilize a manual modification method to assign the determined samples to training sets to update the model. Lastly, a series of comparative experiments demonstrate that our proposed model has outstanding performance in terms of time efficiency and interactive information sentiment index.

## 1. Introduction

Information asymmetry in the Chinese stock market can cause huge losses to investors every year [1, 2]. To alleviate this problem, the China Securities Regulatory Commission (CSRC) and the Shenzhen Stock Exchange (SZSE) have established many digital interactive media, such as HudongYi [3]. Digital interactive media is an emerging interactive technology for communicating between investors and the senior management of listed companies. Typically, digital interactive media exhibit three characteristics: prompt question and answer via short texts, prompt information dissemination, and rich sentiment information.

Many studies in economics and finance have proven that social media content can directly reflect, influence, and even spread investor sentiments [4–8]. According to the irrational investor hypothesis, individual investors' sentiment directly impacts their investment decisions [9], and plenty of irrational investment behaviors can result in dramatic fluctuations in stock prices [4, 6, 7, 10]. Consequently, quantifying and analyzing the sentiment of digital interactive media plays a crucial role in probing investor behaviors and predicting stock prices. This topic has received much attention from researchers who have developed many sentiment classification models, such as the emotion space model [11], the heterogeneous information fusion model [12], and the

contextual entropy model [13]. Nevertheless, these models suffer from the following two main drawbacks.

- (i) *Low Accuracy.* There are many problems with short texts in digital interactive media, such as the feature sparsity of short texts, the usage of nonstandard words, and irregular grammar. These problems make it difficult for traditional models to learn short texts, thereby leading to low sentiment classification accuracy rates.
- (ii) *High Time Cost.* Classifying the sentiment of interactive information is a time-consuming decision problem that traditional models hardly consider. However, investment decisions in the stock market rely heavily on the promptness of sentiment analysis. Therefore, it is challenging to apply conventional models to practical investment decisions.

To solve the problems existing in the traditional models, this study proposes DeeBERT-S3WD, which combines DeeBERT [14] with the sequential three-way decision (S3WD) [15] to quantify and analyze the sentiment of interactive information in digital interactive media. Specifically, DeeBERT is an enhanced BERT, which uses a dynamic early exiting mechanism to accelerate the inference and decision-making process of the original BERT. That is, DeeBERT introduces an additional decision-making function to each layer of the original BERT. When the decision outcome hits a predetermined threshold, DeeBERT immediately exits the subsequent decision process. DeeBERT preserves the powerful representation learning capability of BERT for texts and can solve problems, such as the feature sparsity of short texts, the usage of nonstandard words, and irregular grammar. Besides, DeeBERT accelerates the decision process via its dynamic early exiting mechanism and can solve the second problem mentioned above. Consequently, DeeBERT is ideally suited for the sentiment classification of interactive information in digital interactive media. However, DeeBERT slightly sacrifices the decision accuracy due to its dynamic early exiting mechanism. To address the shortcoming of DeeBERT, we introduce the concept of S3WD to DeeBERT; that is, we add an uncertain set to cache samples that the current layer of DeeBERT cannot evaluate. These samples in the uncertain set will be further learned and judged by the subsequent layers of DeeBERT.

The ablation study demonstrates that DeeBERT-S3WD may significantly reduce the time cost of interactive sentiment analysis while maintaining its accuracy. In addition, substantial experimental results have shown that the sentiment index performs well in econometric models and machine learning algorithms. In conclusion, this research makes the following three unique contributions:

- (i) Based on the character-level embedding technique, DeeBERT-S3WD successfully addresses the representation learning challenges of short texts, leading to improved mining of the interactive information's features.

- (ii) DeeBERT-S3WD employs the concept of dynamic early exiting to establish a balance between accuracy and inference time, which partially compensates for the time burden problems caused by dynamic training.
- (iii) DeeBERT-S3WD incorporates the S3WD concept into DeeBERT, thereby enhancing the effectiveness and efficiency of classifying interactive information sentiments. The key techniques of this study have been implemented and deployed on the Stock++ (Stock++ is a quantitative analysis system for securities' market risk, which can be accessed at the website <https://intelligentstock.cn/sub-vue/home>) system [16].

We organize the remaining sections of this study as follows. In Section 2, we briefly introduce information sentiment classification and the concept of multigranularity computing with the three-way decision. In Section 3, we propose DeeBERT-S3WD for classifying the sentiment of interactive information in digital interactive media. Section 4 reports the experimental results, and Section 5 concludes this study.

## 2. Related Work

With the emergence of digital interactive media platforms, investors and listed companies have gained a direct channel to interact and communicate. As a brand-new factor influencing the volatility of the securities market, research on the impact of interactive information on the risk volatility of the Chinese stock market is urgent and essential. The classification of information sentiment has long been of interest to academics, and related research has steadily increased over the past few years. However, there is a lack of research on quantifying the sentiment of interactive information in digital interactive media. Therefore, this study provides a systematic review of relevant research on the classification of information sentiment.

### 2.1. The Traditional Classification of Information Sentiment.

The influence of Internet information on stock markets has been increasing considerably because of its exponentially increased volume and rapid dissemination. As a result, the researchers' attention has gradually shifted to the sentiment factor of Internet content. Due to the limitations of text mining techniques in the early stages, the amount of Internet information was used as a sentiment indicator. For instance, [17] discovered that the amount of news on Dow Jones and Company was directly related to the aggregate indicators of stock market activity, such as trading volume and market returns, and investors tended to react slowly to negative news. However, utilizing news amounts was insufficient to reflect the influence of news.

To overcome these restrictions, researchers resorted to text mining techniques to extract valuable information from media [18]. As a pilot study, [19] measured the negative (positive) sentiment polarity of an article in terms of the fraction of negative (positive) emotional words in the

document in which the sentiment words are determined by the general emotion word dictionary Harvard-IV-4. Because of their simplicity and easy implementation, these proportion-based sentiment analysis methods were widely adapted to study the influence of textual information. However, this proportion-based approach determines the sentiment polarity solely by general emotional words without considering the syntax of sentences, which could make an opposite analysis of sentiment polarity for the sentences containing negative prepositional phrases like “not” and “without.” In addition, general sentiment words may not be emotional in finance [20]. For example, the general negative sentiment word “tire” was typically used to identify a specific firm.

Furthermore, an emotionless word can be a sentiment in finance. For example, “bear” originally refers to a carnivorous mammal, indicating widespread pessimism in finance, such as a “bear market.” To improve the precision of the sentiment analysis models, it is necessary to determine the documents’ sentiments in terms of financial sentiment words rather than general sentiment words. Therefore, some researchers resorted to more complicated and advanced sentiment analysis techniques [19, 21–24]. For example, [25] proposed a statistical model to detect finance-oriented sentiment words and represented news articles with financial sentiment words to study their influence on stock markets.

### 2.2. *The Modern Classification of Information Sentiment.*

With the substantial increase in computational power, deep learning has begun to receive more attention, providing a new way to analyze information sentiment. In 2018, the Google AI team released a new language model, BERT, which brought a breakthrough in pretrained models in natural language processing and achieved state-of-the-art performance on various natural language processing tasks. Yu and Jiang [26] proposed a multimodal BERT architecture for multimodal target-oriented sentiment classification, and Zhang and Zhang [27] proposed a novel neural network. The main innovations of this approach lie in improving user, item, and word representation learning. They utilized BERT to obtain contextualized word representations and seamlessly feed them into their model to replace the original representations. Lei et al. [28] considered context-aware sentiment analysis as a sequence classification task and proposed a BERT-based hierarchical sequence classification model. It is evident that the BERT has good performance in handling sentiment classification.

BERT-LARGE has up to 340 million parameters, whereas BERT-BASE has 110 million parameters. Even though Google has officially released a pretraining model, the fine-tuning time is still long. The inference time of Fast-BERT [29] may change dynamically to meet various needs. In addition, Fast-BERT is fine-tuned with a novel self-distillation process with enhanced computational performance and little performance degradation. Distilled BiLSTM [30] was proposed to streamline BERT-LARGE into a single layer of BiLSTM, therefore decreasing the number of parameters by 100 times and boosting the speed by 15 times. DeeBERT

[14] uses the concept of an early exiting to accelerate BERT inference. If the sample can be classified in the shallow layer, there is no need to traverse all Transformer layers. DeeBERT slightly sacrifices the decision accuracy due to its dynamic early exiting mechanism.

Our study is inspired by DeeBERT, in which the early exiting mechanism saves time and enables researchers to react more sensitively to information from financial markets. However, they are more concerned with model shrinkage, inference speed improvement, and lack of attention to model accuracy. We invoke the idea of sequential three-branching, in which the samples are divided into positive, negative, and uncertain at each inference stage of the model. The uncertain samples will continue to be classified in the subsequent steps.

*2.3. The Multigranularity Computing.* Multiple layers and views comprise the multigranularity structure [31]. A granule creates a granular hierarchy and forms an entire granular structure. The granular structure may abstractly describe complex systems, which can then be decomposed or synthesized via granular computing. Using granular computing in three-way choices can improve the interpretability of the decision phases. In cognitive science, Yao [32] proposed a description of “three” based on the three-way decision, including “three steps,” “three elements,” and “three levels,” and further proposed a trisecting-acting-outcome (TAO) model by considering the cognitive problem in “three.” The TAO model reflects a basic structural framework of the broad sense of the three-way decision process. The objects are classified into three categories via trisecting, and decisions are made in acting. Finally, the results of the effectiveness of the “trisecting” and “acting” process are evaluated in “outcome.” The process provides interpretability in classifying uncertain problems and the thinking in “three.” The TAO model provides a feasible interpretation form for multigranularity classification and abstract representation.

With the advent of dynamic decision-making, the traditional single-step (static) decision process has found it difficult to satisfy practical requirements, whereas the necessity for multistep decisions has increased. Complex decision-making processes should be analyzed using multigranularity to get more intelligible and interpretable results. Several researchers in recent years have extensively explored it. For example, a new S3WD model based on a penalty function to improve the classification accuracy by modifying cost parameters was proposed [33], and a cotraining method was incorporated into S3WD to label new samples with higher confidence [34]; a novel dynamic update method of three-way granular concepts was discussed [35], and other research works described granular computing and S3WD [15, 36–39] and three-way granular computing based on multilevel and multiview structure [40–42]. Therefore, in the interactive information sentiment analysis in this study, the application of the multigranularity method combined with S3WD ideas can be well reflected in the sequence classification task of DeeBERT.

*2.4. The Sequential Three-Way Decision.* To improve the accuracy of the DeeBERT model, we combined it with the sequential three-way decision models (S3WD), which balance efficiency and decision time. Yao [43, 44] introduced three-way decisions in 2010 as a valuable method for handling uncertain problems and processing information based on a decision-theoretic rough set (DTRS) and Bayesian risk decision-making [45]. It divides the analyzed object into positive, negative, and uncertain [44]. Yao [15] was the first to propose the S3WD model, a multilayer framework leading to faster decision-making at a reduced cost. Based on the DTRS architecture, many S3WD models were created. For example, Li et al. [37] presented a cost-sensitive sequential 3WD model for image classification. The cost-sensitive S3WD technique uses the decision-theoretic rough sets (DTRS) model, while conventional 3WD employs a static strategy. This model made an excellent and low-cost decision based on high-quality facial images.

The optimal decision depends on the minimum decision cost according to the currently available information. In the case of insufficient information, the decision may fall into a boundary domain with high-cost loss. Thus, it is necessary to search for an appropriate sequential decision strategy to balance the decision and test costs, leading to a lower overall decision. In the past decade, S3WD has been widely applied to different fields in decision-making. Li et al. [37] proposed a sequential three-way decision method based on granular computing for cost-sensitive face recognition. Zhang et al. [46] proposed a cost-sensitive combination technique using sequential three-way decisions in sentiment classification. Li et al. [47] presented a DNN-based sequential granular feature extraction method to extract a hierarchical granular structure from the input images.

Some recent papers focused on introducing the S3WD-based sentiment analysis methods. For example, Basiri et al. [48] proposed two deep fusion models based on three-way decision theory to analyze drug reviews. Yang et al. [49] suggested a temporal-spatial three-way multigranularity learning framework for dynamic text sentiment categorization to address dynamic data uncertainty. Zhang et al. [46, 50] introduced the 3W-CNN model for an upgraded three-way convolutional neural network to improve sentiment classification accuracy.

Nevertheless, few researchers applied S3WD in the financial area. For example, Shen et al. [51] proposed a novel three-stage reject inference learning framework by using transfer learning and three-way decision theory in credit scoring. Maldonado et al. [52] provided an approach for applying three-way decisions for credit scoring problems that are expensive and time-consuming.

In addition, S3WD models have less applicability for identifying interactive stock market information sentiment. S3WD models may improve classification efficacy and precision; this study represents an approach that combines the DeeBERT model with S3WD models to decrease time cost while ensuring accuracy dramatically. Using the S3WD model to increase classification accuracy, DeeBERT can

successfully address the short text and class-imbalanced issues. Consequently, this combination is suitable for balancing efficiency and effectiveness.

### 3. DeeBERT-S3WD and Experiments

This study classifies the sentiment of interactive messages on digital platforms. Therefore, based on previous research, we design an intelligent framework of DeeBERT and sequential three-way decisions to solve this problem efficiently. First, we obtain investors' inquiries over digital platforms using web crawler technology. Next, we use the DeeBERT-S3WD model to capture the sentiment tendencies of interactive information within the massive data available through the digitized media. Finally, we utilize different predictive analysis models to investigate the effect of interactive information sentiment on the stock market.

*3.1. The Research Data.* This study collected data from China's three largest digital interactive platforms: Panorama Network (Panorama Network (<https://www.p5w.net>)) was founded in 1999 as an independent website of securities' finance and economy in Mainland China, and Panorama has accumulated more than 900,000 pages of data and information reserve over the past ten years, which can provide users with public information of listed companies in the first time), eHuDong (eHuDong (<https://sns.sseinfo.com>)) is a communication platform established by the Shanghai Stock Exchange and used by all participants for free. The website aims to guide and promote information communication between listed companies, investors, and other market participants), and HudongYi (HudongYi (<https://irm.cninfo.com.cn>)) is an interactive platform for investor relations of listed companies launched by the Shenzhen Stock Exchange<sup>4</sup>. The primary purpose of HudongYi is to meet the communication and interaction needs between listed companies and investors. On the platform, investors can ask questions directly to listed companies, which can respond [53]. The daily number of visitors might reach 2 million. Unlike conventional media, these platforms have forums for each listed company. In other words, users can publish questions on discussion boards, and representatives of each company make formal replies. Since there is no publicly available dataset of interactive information for the digital interactive platform in academia, this study extracted 1.62 million interactive information (question-answer pairs) from 1,765 listed companies between 01/2012 and 12/2018 via a distributed web crawler system from Panorama Network, eHuDong, and HudongYi. Next, we completed the data preprocessing, such as interpolating missing values and eliminating duplicated values, based on Spark distributed technology. In addition, question-answer interactions represent the entire information disclosure process. The question-answer interactive information is manually labeled with sentiment classes. To reduce the workload of sentiment annotation, we resorted to a professional Chinese financial sentiment lexicon created by the previous work [54].

3.2. *The Intelligent Collaborative System Framework.* The intelligent framework for classifying the sentiment of interactive information in this study is shown in Figure 1, which mainly includes some essential parts, such as document preprocessing and document vectorization.

3.2.1. *The Documents Preprocessing.* Data cleaning is an indispensable step in the whole process of data analysis. The quality of the results is directly related to the effect of the model and the conclusion. Therefore, before starting the experiment, we must preprocess the documents obtained. The main steps of document preprocessing include removing missing data, replacing duplicate values, and manual sentiment annotation of interactive information (question-answer pairs) documents.

3.2.2. *The DeeBERT-S3WD Model.* The DeeBERT-S3WD model proposed in this study is an optimization improvement of the BERT model. The BERT model is a combination of pretraining and downstream task models, which means that the BERT model is still used when doing downstream tasks, and it naturally supports text classification tasks. However, the inference speed of traditional BERT is slow. It is difficult to effectively solve the sparsity and feature dispersion problems of short texts and lacks adaptability in the face of massive data processing tasks. In addition, it is necessary to balance misclassification cost and time cost to improve the accuracy and obtain the optimal decision in dynamic and uncertain decision-making, especially in text classification tasks for the securities market. Therefore, this study proposes an intelligent framework for textual sentiment classification that combines dynamic early exiting DeeBERT with the sequential three-way decision to analyze the sentiment analysis of interactive information in digital platforms.

(1) *Multigranularity Sentiment Classification.* The sequential three-way decision provides a valuable and meaningful simulation of the decision-making process. This study combines this model with multigranularity sentiment classification for interactive information sentiment analysis. In the S3WD model, a delayed decision is more reasonable when information is unavailable, and a precise decision cannot be reached quickly. What is more, an incorrect decision may result in a higher cost. The more available information can be used, the more precise granular feature can be obtained. Thus, a delayed decision in the earlier steps could be a low-cost decision with high performance [47].

A multistep (sequence) decision-making procedure is hierarchical and time-consuming. Hierarchical granularity can be seen as a supplement to sequential decision-making. The combination of S3WD and GrC is an efficient method to classify uncertain objects and achieve lower-cost results. The definition of three regions can be described in the following dynamic scenarios.

*Definition 1.* The multilevel structure,  $GS = \{GS_1, GS_2, \dots, GS_n\}$ , and the  $i$ th level of granular structure,  $GS_i = \{U_i, C_i, v_i\}$ , are given, where  $U_i$ ,  $C_i$ , and  $v_i$ ,

respectively, represent a finite and nonempty set of objects, conditional attributes, and evaluation functions. In the sequential decision-making process, the  $i$ th level of three regions can be defined as

$$\begin{aligned} \text{Region}_{1i}(v(x)_i) &= \{x \in U_i | v(x)_i \geq \alpha_i\}, \\ \text{Region}_{2i}(v(x)_i) &= \{x \in U_i | \alpha_i < v(x)_i < \beta_i\}, \\ \text{Region}_{3i}(v(x)_i) &= \{x \in U_i | v(x)_i \leq \beta_i\}, \end{aligned} \quad (1)$$

where  $(\alpha_i, \beta_i)$  is a pair of thresholds ( $0 \leq \beta_i < \alpha_i \leq 1$ ) under the  $i$ th step granularity  $GS_i$ .

The S3WD models are built based on information tables, including attribute and value sets. But interactive information sensitive analysis covers a data set rather than included values. To address this problem, a more general S3WD model was proposed by [37], which is not restricted to rough sets and information tables, where an S3WD model is formulated as follows.

*Definition 2.* An S3WD set  $ST = (ST_1, ST_2, \dots, ST_i)$  is given, where  $T_i = \{a_p^i, a_B^i, a_N^i\}$  is the  $i$ th step in the decision process. The subset  $X = \{x_1, x_2, \dots, x_i\}$  and the multigranular set  $GS = \{GS_1, GS_2, \dots, GS_i\}$  are given, where the  $i$ th granular set is  $GS_i = \{g_1, g_2, \dots, g_i\} = \{[x]_1, [x]_2, \dots, [x]_i\}$ . Thus, the sequential decision process is described as follows [55, 56]:

$$\begin{aligned} ST &= (ST_1, ST_2, \dots, ST_i) \\ &= (\phi(GS_1), \phi(GS_2), \dots, \phi(GS_i)), \end{aligned} \quad (2)$$

where  $\phi(GS_i) = \{\phi(g_1), \phi(g_2), \dots, \phi(g_i)\}$  is the optimal three-way decision with the minimum risk cost in the  $i$ th step of decision-making.

The S3WD models consider not only the decision cost but also the test cost because high-quality granular features tend to obtain a low misclassification cost with high time consumption, which increases the total cost of a decision. Therefore, it is necessary to balance the total cost and correctly evaluate a decision by considering the test cost [47]. In this study, the test cost of interactive information sentiment analysis denotes the  $i$ th base classifiers in the DeeBERT model, including the time cost of the first and next base classifiers [46].

In the S3WD models, consider a state set with only two states  $\Omega = \{C, \bar{C}\}$ , indicating that the event belongs to  $C$  and does not belong to  $C$ . Given the decision actions set as the positive, negative, and boundary decision under the  $i$ th step,  $A = \{a_p^i, a_N^i, a_B^i\}$ . In standard scenarios, different decisions will produce various losses, so a function loss value  $\lambda$  is derived, and the corresponding decision loss matrix [45] is as follows.

In Table 1,  $\lambda_{PP}$ ,  $\lambda_{BP}$ , and  $\lambda_{NP}$  is the corresponding loss values when  $x \in C$  and the three decisions of  $a_p^i$ ,  $a_B^i$ , and  $a_N^i$  are adopted, while the other loss values  $\lambda_{PN}$ ,  $\lambda_{BN}$ , and  $\lambda_{NN}$  are defined when  $x \in \bar{C}$  and  $a_p^i$ ,  $a_B^i$ , and  $a_N^i$  in the  $i$ th step decision-making process. The relationship between the three loss values is

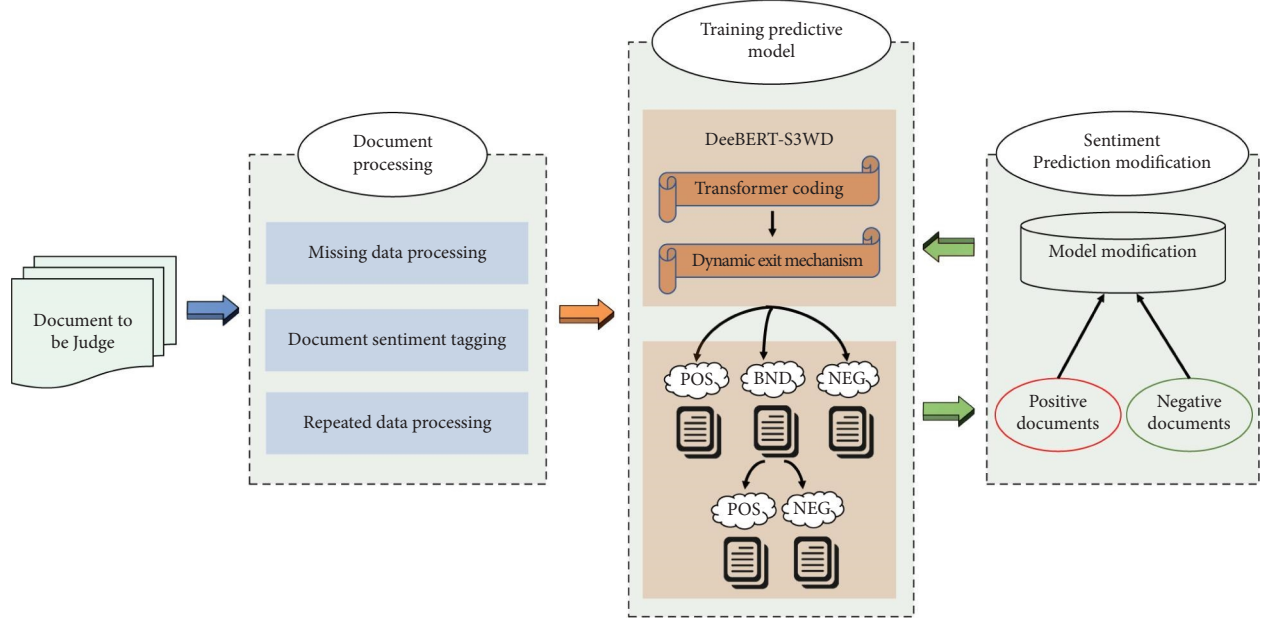


FIGURE 1: The framework of interactive information sentiment analysis.

$$\lambda_{PP} < \lambda_{BP} < \lambda_{NP}, \lambda_{PN} < \lambda_{BN} < \lambda_{NN}. \quad (3)$$

According to the risk loss matrix and the Bayesian decision method of minimum risk, the expected loss of three different decisions is calculated. Given the expected risk corresponding to the three decisions of  $a_p, a_B, a_N$  in the  $i$ th granular layer,

$$\begin{aligned} R(a_p^i | [x]_i) &= \lambda_{PP} \Pr(C | [x]_i) + \lambda_{PN} \Pr(C | [x]_i), \\ R(a_B^i | [x]_i) &= \lambda_{BP} \Pr(C | [x]_i) + \lambda_{BN} \Pr(C | [x]_i), \\ R(a_N^i | [x]_i) &= \lambda_{NP} \Pr(C | [x]_i) + \lambda_{NN} \Pr(C | [x]_i). \end{aligned} \quad (4)$$

Compute the minimum decision cost according to three expected decision costs  $R(a_p | [x]_i), R(a_B | [x]_i),$  and  $R(a_N | [x]_i)$  [55]. The optimal selection S3WD decision can be calculated as follows:

$$\phi([x]_i) = \arg \min_{a \in \{a_p^i, a_B^i, a_N^i\}} R(a | [x]_i). \quad (5)$$

With the continuing of each single-step decision, decision cost will increase gradually. The decision cost can be computed in a single  $t$ th step decision  $ST_t$ :

$$\text{cost}(x, ST_t) = \min_{i \in \{P, B, N\}} \sum_{j \in \{P, N\}} \lambda_{ij} P(X_j | GS_t), \quad (6)$$

where  $\{\lambda_{ij}\}_{3 \times 2}$  ( $i \in \{P, B, N\}, j \in \{P, N\}$ ) is a decision cost matrix.

In a cost-sensitive sequential three-way decision, the  $i$ th step decision is acquired by minimizing the decision cost through the  $i$ th granular feature information. The training time for a DeeBERT is related to the training loop and decision step, which is computed as the test cost. The total cost in the  $i$ th step can be calculated according to the Bayesian decision procedure [37, 46, 47]:

$$\begin{aligned} \text{cost}(a_p^i | GS_i) &= \lambda_{PP} \Pr(P | GS_i) + \lambda_{PN} \Pr(N | GS_i) + \sum_{j=1}^i t_{\text{cost}_j}, \\ \text{cost}(a_N^i | GS_i) &= \lambda_{NN} \Pr(P | GS_i) + \lambda_{NP} \Pr(P | GS_i) + \sum_{j=1}^i t_{\text{cost}_j}, \\ \text{cost}(a_B^i | GS_i) &= \lambda_{BB} \Pr(P | GS_i) + \lambda_{BN} \Pr(N | GS_i) + \sum_{j=1}^i t_{\text{cost}_j}, \end{aligned} \quad (7)$$

where  $\text{cost}(GS_i) = \sum_{j=1}^i t_{\text{cost}_j}$  denotes the test cost. Based on the decision costs and the decision selection with the minimum cost as the optimal decision for the  $i$ th step,

$$\begin{aligned} ST_i &= \phi(GS_i) \\ &= \arg \min_{a \in \{a_p^i, a_B^i, a_N^i\}} \text{cost}(a | GS_i). \end{aligned} \quad (8)$$

The optimal results of sequential decisions are determined by both the misclassification costs and test costs. If the granular features are not precise enough, the decision results in a high misclassification cost and needs more training time to obtain available and valuable information. Thus, the minimal cost is a crucial criterion in the termination step in the sequential decision process [47].

(2) *DeeBERT-S3WD Predicting Model.* This section describes the process of interactive information sentiment classification using the BERT-sequential three-way prediction model. BERT has achieved good results as a large-scale pretraining model. The requirement of BERT for input text makes it more suitable for handling short textual data. The interactive information of investors in the market is usually short text, so BERT is useful. However, the inference speed of BERT makes it notorious for its massive number of participants, which makes its training and inference time substantially longer.

DeeBERT [14] uses a dynamic inference mechanism to determine whether the sample has enough information to exit at each Transformer layer except the last one. If satisfied, it can leave early without going through all Transformer layers so that the sample exits early to speed up BERT inference, dramatically improving the inference speed at the expense of a small amount of accuracy. However, the input data of DeeBERT is static. After training, the model is validated on the testing set, and the accuracy and inference time are output. The information in the stock market changes rapidly, with a constant stream of information input daily. A reasonable model should be able to make decisions based on the latest information when judging each sample rather than using unchanging and ancient information.

Therefore, this study proposes a new inference decision system based on the improved DeeBERT, named DeeBERT-S3WD, drawing on previous research. The traditional DeeBERT makes decisions at each Transformer layer except the last one, exits early if there is enough confidence, continues to enter if there is not enough information, and outputs directly at the last Transformer layer if the sample finally enters the previous Transformer layer. DeeBERT-S3WD uses three-branch decisions at each Transformer layer in the inference stage; that is, the results are divided into the positive, negative, and boundary domains. The samples in the boundary domain will go to the next Transformer layer for judgment. DeeBERT-S3WD still makes judgments at the last layer, and if the sample is confident enough to exit early, then the sample is not retained. We manually modify these early exiting samples to make them into training set training models. Let the retained samples wait for the model update to enter the next batch of samples to be tested. These steps are shown in Figure 2 and Algorithm 1.

(3) *Dynamic Early Exiting Mechanism.* We use Algorithm 2 to illustrate the dynamic early exiting mechanism. In this study, a judgment is added to each Transformer layer, which allows the sample to exit early without going through all the layers. One of the last Transformer layers continues to make judgments with sufficient and insufficient confidence. The sample is retained and tested with the next fresh batch of test sets after waiting for the model to be updated.

A smaller threshold enables more samples to make decisions at a deeper level, resulting in a higher accuracy of the sample's output per level but a lower speed. A larger threshold allows more samples to be made at a shallow level, resulting in a decrease in the accuracy of the sample output per level but an increase in speed. With this criterion, we hope to find a suitable set of  $\lambda_{PP}, \lambda_{BP}, \lambda_{NP}, \lambda_{PN}, \lambda_{BN},$  and  $\lambda_{NN}$  that minimizes the sum of misclassification cost and total time cost.

### 3.2.3. Validate the Applicational Effect of the DeeBERT-S3WD Model

(1) *Regression Model for Index Tests.* Here, we use the regression model to test the effectiveness of interactive information sentiment classification. Thus, we adopt the most

representative Fama-French 3-factor model [57] in the field of finance in our study:

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_i(R_{m,t} - R_{f,t}) + \delta_i SMB_t + \theta_i HML_t + \varepsilon_{i,t}, \quad (9)$$

where  $R_{i,t}$  is the expected return rates of a portfolio on the  $t$ th day,  $R_{f,t}$  is the risk-free return rates on the  $t$ th day, and  $R_{m,t}$  is the return of the market portfolio on the  $t$ th day.  $SMB_t$  is the scale factor, which stands for small market capitalization – big market capitalization on the  $t$ th day, and  $HML_t$  is the net asset market value ratio factor, indicating the differences between high book-to-market ratio minus low book-to-market ratio on the  $t$ th day.  $SMB_t$  and  $HML_t$  are used to measure the historic excess returns of small capitals over big capitals and value stocks over growth stocks.

We add the interactive information sentiment indicator  $S$  into the Fama-French 3-factor model to evaluate the effect of textual information on stocks as suggested by [19], and we hope that the indicators of sentiment refined by DeeBERT-S3WD will have good performance in our experiments.

$$R_{i,t} = R_{f,t} + \alpha_i + \beta_i(R_{m,t} - R_{f,t}) + \delta_i SMB_t + \theta_i HML_t + \varphi_i S + \varepsilon_{i,t}. \quad (10)$$

(2) *Machine Learning Models for Index Tests.* To further verify the effectiveness of the quantified indicators of interactive information sentiment, we also design some comparative experiments based on the computer science field. Specifically, we add interactive information sentiment indicators to traditional machine learning models (Bayes, SVM) and deep learning models (CNN, LSTM). By observing the effect changing of the machine learning model before and after the classification of interactive information sentiment indicators added, we can directly get the differences in models' ability to capture the future trends of the stock prices. Therefore, we can obtain the performance of quantitative interactive information sentiment indicators in the practical stock market.

## 4. Experimental Results

4.1. *Classification Performance of DeeBERT-S3WD.* In this section, we evaluate the effectiveness of DeeBERT-S3WD in analyzing the sentiment of interactive information. In particular, we select SVM, CNN, DeeBERT, BERT, RoBERTa [58], and DeBERTa [59] as baseline models and test them on a large-scale corpus containing 100,000 interaction texts with sentiment labels (positive or negative). As shown in Table 2, DeeBERT-S3WD outperforms SVM, CNN, and DeeBERT but is inferior to the state-of-the-art models, such as BERT, RoBERTa, and DeBERTa. This is because we build DeeBERT-S3WD based on the dynamic early existing mechanism of S3WD, which sacrifices accuracy for inference time.

In addition, since the interactive information in the digital platforms is class-imbalanced datasets and short text information, this may lead to underperformance in sentiment analysis. Therefore, we decide to compare the performance of

```

Input: data.
1 while  $data \neq \emptyset$  do
2   dataset =  $data \cup uncertain$ ;
3   uncertain  $\leftarrow \emptyset$ ;
4   for  $x \in dataset$  do
5     for  $i=0$  to  $n$  do
6        $z_i = f_i(x_i, \theta)$ ;
7       if  $S3W D(z_i) > \lambda_1$  then
8          $x \in positive$ ;
9         break;
10      end
11     if  $S3W D(z_i) < \lambda_2$  then
12        $x \in negative$ ;
13       break;
14     end
15   end
16    $x \in uncertain$ .
17 end
18 end

```

ALGORITHM 1: Dynamic early exiting mechanism.

```

1 Initiate use a large-scale text training set for training BERT;
2 repeat
3   Dataset Dataset =  $data \cup uncertain$ ;
4   Classify Three subsets of Dataset;
5   Correction Manual correction of the positive and negative sets;
6   Update the model with a certain set;
7 until  $data = \emptyset$ ;

```

ALGORITHM 2: The model operating mechanism.

TABLE 1: Corresponding decision loss matrix.

	C	C
$a_p^i$	$\lambda_{PP}$	$\lambda_{PN}$
$a_B^i$	$\lambda_{BP}$	$\lambda_{BN}$
$a_N^i$	$\lambda_{NP}$	$\lambda_{NN}$

the DeeBERT-S3WD-based approaches with the traditional financial methods (proportion-based approaches). In the conventional financial field, the proportion-based approach proposed by [19] is one of the most classic ones. The primary measure of media content is the standardized fraction of negative (or positive) words in each news story. In this experiment, the final accuracy is the average of the 10-fold cross-validation. As Table 3 displays, the DeeBERT-S3WD-based sentiment analysis approaches can almost reach a human-level judgment and suggest a supreme performance compared to the proportion-based approaches.

In particular, the DeeBERT-S3WD-based approaches outperform the proportion-based approaches by 8% on positive sentiment classification and 6% on negative sentiment classification. Note that we favor the proportion-based approach by utilizing financial sentiment words instead of general sentiment words adopted by [19]. Approximately three-quarters of all negative words in a general emotion

word dictionary (Harvard-IV-4) are not considered negative in a financial context [20]. Therefore, we use the financial Sentiment word list provided by [25] and revise the proportion-based approach to enhance the effect.

**4.2. Decision Efficiency of DeeBERT-S3WD.** In fact, time efficiency is also critical for the proposed method, especially utilizing the model in particles. Here, we conduct a time detection experiment to evaluate the time-sensitive efficiency. Particularly, we pay more attention to the actual working time of the model, so we ignore the time required for manual correction and model update.

There are two reasons for this: First, we assume the manual correction happens after the closing, and other models are at rest during the manual correction phase and the DeeBERT-S3WD update phase. Second, the manually corrected samples are only newly added samples daily. Manual correction does not require much time. Thus, our work can ensure that the model decision and update can be completed before the next opening.

As shown in Table 4, DeeBERT-S3WD is more applicable to the actual financial market since it allows a fast sentiment analysis while guaranteeing high accuracy. DeeBERT-S3WD uses S3WD to calculate the threshold sensibly, allowing samples to exit early without going through all encoder layers.



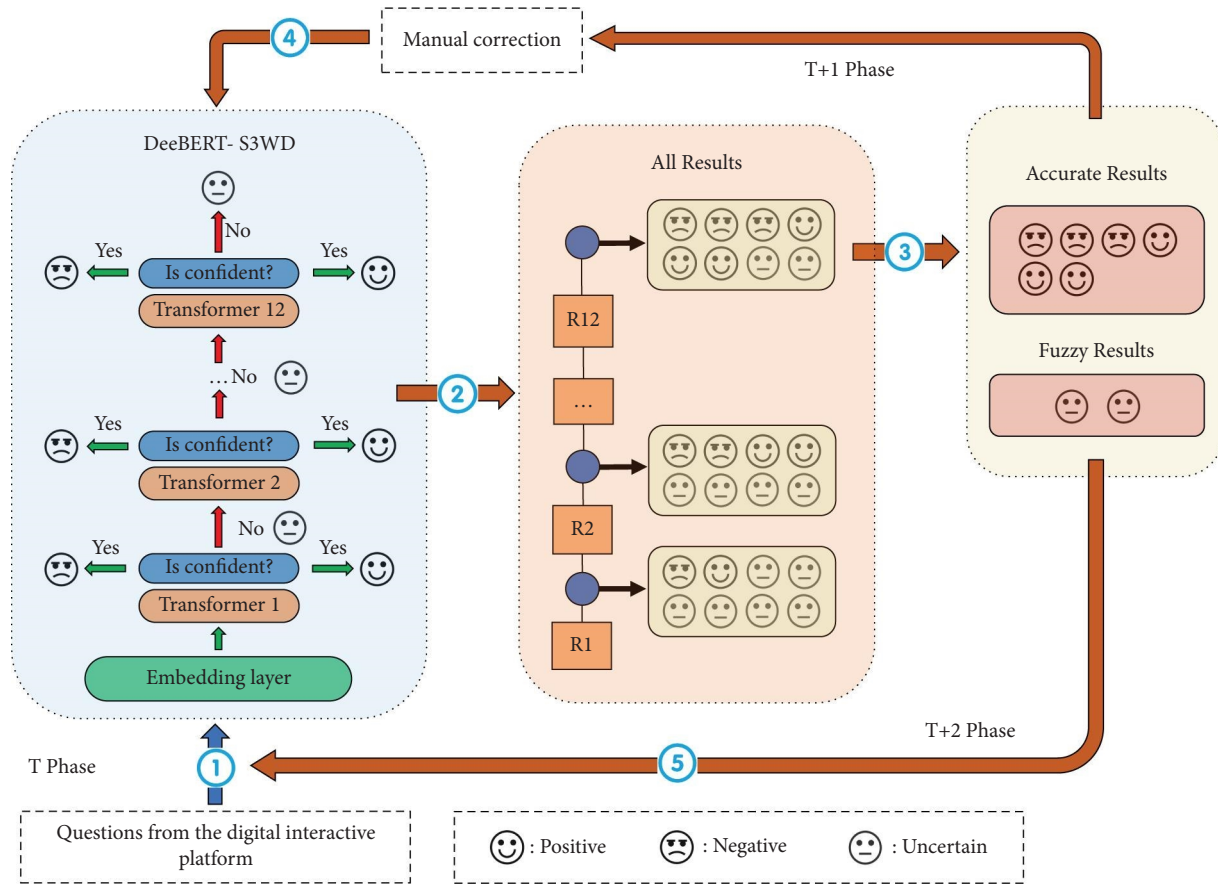


FIGURE 2: Overview of DeeBERT-S3WD. (1) Data from digital interactive media are preprocessed and then fed into the model. (2) This is the model’s specification. As the number of layers grows, the number of uncertain samples will decrease. (3) At the final layer, the uncertain and certain samples will go in opposite directions. (4) The determined samples will be used as training sets for model updates following manual adjustment. (5) Uncertain samples will be maintained and incorporated into the model with the next batch of data after the model’s update.

TABLE 2: Sentiment classification results.

Method	Class	Total accuracy	Accuracy	$F_1$
SVM	Positive	0.7694	0.7837	0.7549
	Negative		0.7551	0.7396
CNN	Positive	0.7759	0.8063	0.7834
	Negative		0.7455	0.7188
DeeBERT	Positive	0.8023	0.7953	0.7849
	Negative		0.8092	0.8214
DeeBERT-S3WD	Positive	0.8177	0.8090	0.7951
	Negative		0.8264	0.846
BERT	Positive	0.818	0.8102	0.8231
	Negative		0.8257	0.8325
RoBERTa	Positive	0.8301	0.8215	0.8336
	Negative		0.8387	0.8597
DeBERTa	Positive	0.8346	0.8324	0.8437
	Negative		0.8367	0.8664

### 4.3. Application Effectiveness of DeeBERT-S3WD Model

4.3.1. Effectiveness of Sentiment Classification Index in Econometric Model. To verify the validity of the classification results of the interactive information sentiment classification results, we apply the interactive sentiment analysis

results produced by DeeBERT-S3WD to the Fama-French 3-factor model. In our first test experiments, we can explore the performance of interactive information sentiment analysis results to evaluate the performance of sentiment classification approaches. To estimate model parameters, we collect stock price data of the SSE 50 (the SSE 50 investment

TABLE 3: 10-fold cross-validation results of interactive information sentiment analysis.

Emotional attributes	Human judgment	Total	Proportion-based approaches		DeeBERT-S3WD-based approaches	
			Prediction results	ACC (%)	Prediction results	ACC (%)
Positive	67486	1000	48981	72.58	54596	80.90
Negative	32514		24200	74.43	26870	82.64

TABLE 4: Sentiment detection time.

Method	Total accuracy	Total time (s)
SVM	0.7694	89.6
CNN	0.7759	110.6
DeeBERT	0.8023	77
DeeBERT-S3WD	0.8177	73.2
BERT	0.818	91
RoBERTa	0.8301	94
DeBERTa	0.8346	103

TABLE 5: Effect of interactive information sentiment analysis indicator in the regression model.

	Before			After		
	Estimate	<i>t</i> -value	Pr ( $> t $ )	Estimate	<i>t</i> -value	Pr ( $> t $ )
<i>R</i>	1.382***	193.362	0.001	1.176***	154.44	0.001
SMB	0.524***	37.771	0.001	0.493***	35.707	0.001
HML	-0.072**	-3.837	0.013	-0.046**	-2.935	0.042
<i>S</i>				-0.011***	-11.896	0.001
Cons	-0.317***	-79.639	0.001	-0.183***	-62.372	0.001

TABLE 6: List of SSE 50 companies.

Stock code					
600519.SH	600000.SH	600028.SH	601169.SH	601211.SH	
600958.SH	600048.SH	601985.SH	601668.SH	601006.SH	
601857.SH	601088.SH	601688.SH	600019.SH	600104.SH	
600050.SH	601398.SH	600919.SH	600547.SH	600999.SH	
600036.SH	601878.SH	601766.SH	601390.SH	601601.SH	
601186.SH	601881.SH	603993.SH	601988.SH	600837.SH	
600029.SH	600887.SH	601229.SH	601328.SH	601288.SH	
601989.SH	600111.SH	600340.SH	601336.SH	601628.SH	
601166.SH	601669.SH	600518.SH	601800.SH	600030.SH	
600309.SH	601318.SH	600606.SH	600016.SH	601818.SH	

index is a sample stock selected from the most representative 50 stocks with large scale and good liquidity in the Shanghai stock market according to scientific and objective methods so as to comprehensively reflect the overall situation of a group of leading enterprises with the most market influence in the Shanghai stock market. For a list, see Table 5 in the appendix) index from January 2018 to January 2019 (with a one-month time window). The results are shown in Table 6.

In recent years, researchers have found that using sentiment words or sentiment indicators has captured most media images and extracted the core ideas of text information [6, 9, 54, 60, 61]. Therefore, sentiment analysis, as the classification of media information, has become the mainstream method in the academic world. Table 5 shows the

performance of the interactive information sentiment variables in the econometric model before and after they are added to the model. The interactive information sentiment analysis variable has excellent statistical significance and passes the robustness test. Specifically, the DeeBERT-S3WD-based approach has a better performance in the econometric regression model, and the interactive information sentiment (*S*) variable has excellent statistical significance ( $p < 0.01$ ). Therefore, the results prove that the quantified sentiment indicators extracted through the DeeBERT-S3WD-based approaches are validated in the detection of regression models, validating the effectiveness of interactive information sentiment tendency indicators. Many previous studies have proved the practical application ability of media sentiment indicators, which is consistent with the results of this study.

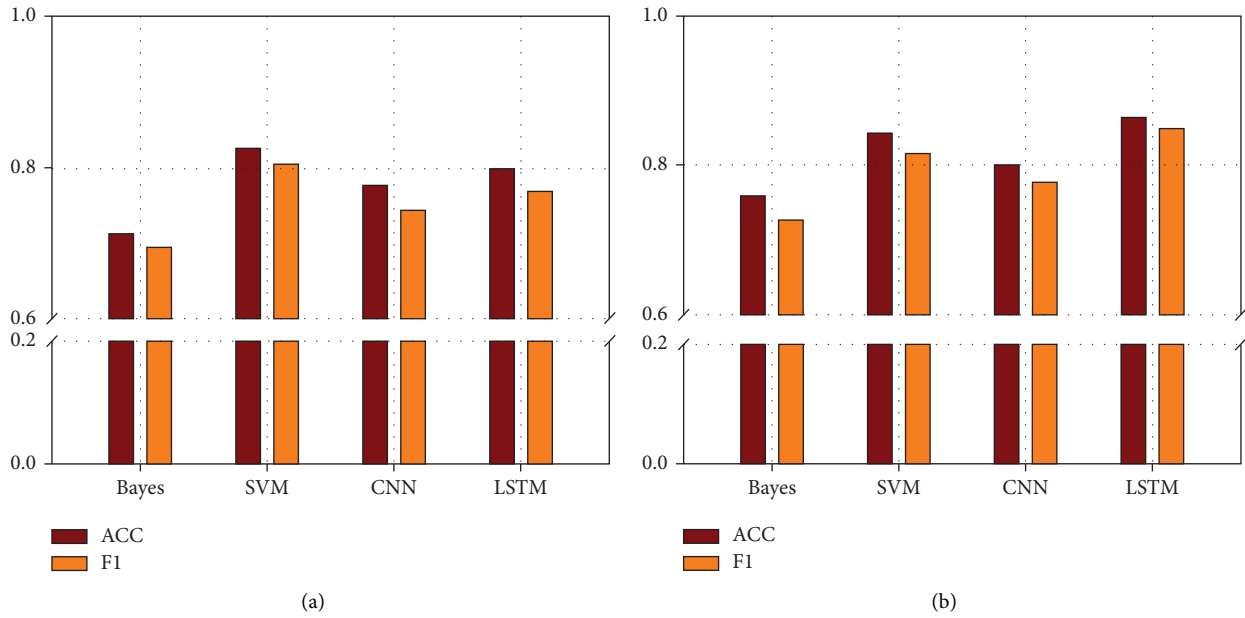


FIGURE 3: Effectiveness of sentiment classification index in machine learning models.

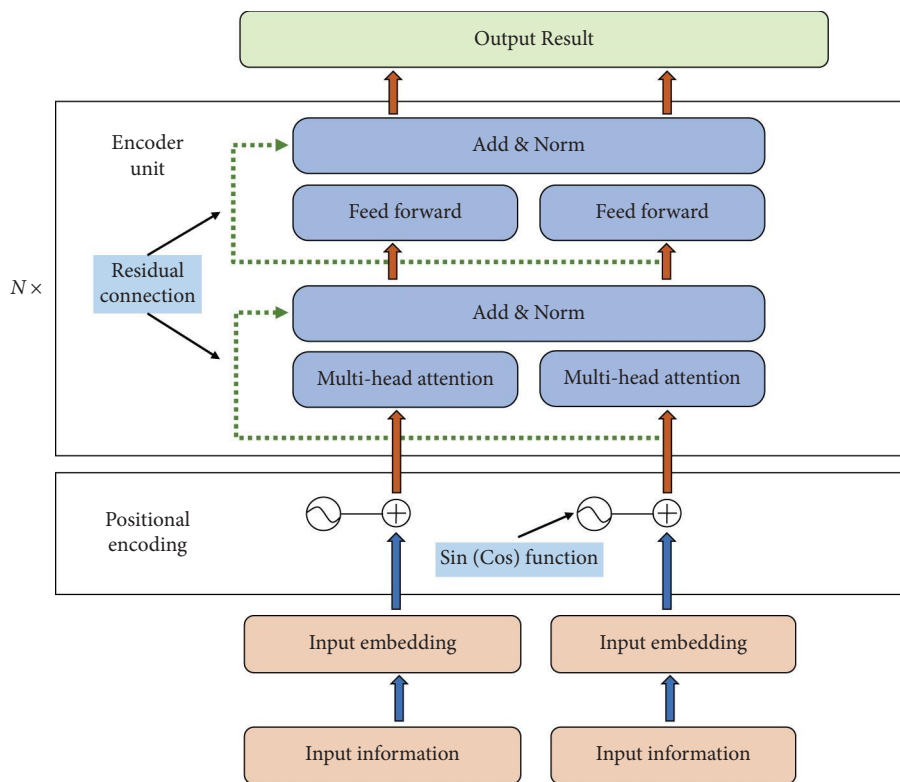


FIGURE 4: Transformer encoder unit.

4.3.2. *Effectiveness of Sentiment Classification Index in Machine Learning Models.* We further design experiments with machine learning models to verify the effectiveness of the classified indicators of interactive information senti-

ment variable ( $S$ ) to traditional machine learning models (Bayes and SVM) and deep learning models (CNN and LSTM) to predict stock price future direction. Figure 3 suggests the experimental results.

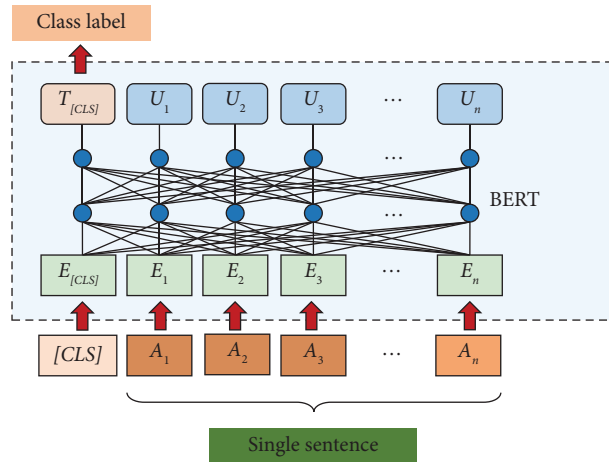


FIGURE 5: Fine-tuning of BERT-based text classification tasks.

In Figure 3(b), we can see the experiment results of traditional machine learning models (Bayes and SVM) and deep learning models (CNN and LSTM) with only fundamental variables. SVM model has higher prediction ACC (81%) and  $F_1$  than other models. In Figure 3(a), after adding classify interactive information sentiment variable, all the models performed better in both ACC and  $F_1$  respect. The results show the following: (1) After adding sentiment variables to traditional machine learning and deep learning models, the ACC and  $F_1$  also have significantly improved, proving sentiment variables' influential role. (2) Compared with traditional machine learning models, the quantified interactive information sentiment variable has a more significant impact on deep learning models, and the effect is more obvious. We can prove that the predictive effect of the quantitative sentiment variable on all machine learning models has a significant enhancement effect, and the enhancement effect performs better in deep learning models.

## 5. Conclusion

Information is essential to the development of the stock market. Investment decisions rely on the truthfulness, precision, timeliness, and accessibility of information. The emerging digital interactive platforms provide a channel of communication between investors and listed companies, which significantly impacts the securities market. In this study, we develop a classification and analysis framework called DeeBERT-S3WD for interactive information attitudes. The extensive experiments have shown the following: (i) The DeeBERT-S3WD outperforms other machine learning models in class-imbalanced datasets, short texts, and decision time. (ii) Compared to the proportion-based approaches in finance, the DeeBERT-S3WD performs better in terms of accuracy. (iii) The sentiment classification index from the DeeBERT-S3WD can significantly improve the effectiveness of econometric and machine learning models. Finally, the DeeBERT-S3WD has been implemented and deployed on the Stock++, a quantitative analysis system for securities market risk [16].

## Appendix

### A. BERT Pretraining Tasks

The BERT sets up two unsupervised target tasks to obtain word and sentence-level epistasis, respectively.

- (i) Masked LM implements pretraining of bidirectional language models: Unlike other language models such as Word2Vec, which require the prediction of all words in the input sequence, Masked LM randomly selects 15% of words in the input data for masked operation and predicts this 15% of words by contextual words to avoid the influence of the next words on the current word. Thus, this is a true “two-way” operation. Among the 15% of the masked words, 80% are replaced by the “[MASK]” symbol, 10% are replaced by randomly selected words from the corpus, and the remaining 10% are kept without transformation.
- (ii) Next sentence prediction: This method is used to determine whether two sentences (A, B) are related to each other as a binary classification task. 50% of the training data (A, B) is the real next sentence as the positive example, and the remaining 50% of the (A, B) is randomly selected as the negative example for training. However, neural networks, such as LSTM and GRU, have a serial problem due to their own temporal class models, which greatly limits the training speed, while BERT adopts the Transform coding structure and uses self-attention to compute the relationship between words in parallel to solve this problem.

### B. Encoder Unit of BERT Model

BERT consists of a stack of multiple Transformer encoders, which represent the document as a vector and are fed into the Transformer encoding unit. The Transformer discards the circular network structure of RNN and models a piece of

text based entirely on the attention mechanism. The encoding unit is shown in Figure 4.

To address the problem that the self-attentive mechanism cannot extract timing features, the Transformer adopts a position embedding approach by defining sin and cos functions for even and odd positions, respectively, to add timing information, as shown in the following equation.

$$PE(\text{pos}, 2i) \\ PE(\text{pos}, 2i + 1) = \cos\left(\frac{\text{pos}}{10000^{(2i/d_{\text{model}})}}\right), \quad (\text{A.1})$$

where pos is the position of the current word in the sentence, the BERT model takes the sum of the word vector, the text vector obtained by embedding, and the position vector obtained by positional encoding as the model input.

Subsequently, the obtained vectors are fed into the main module of the Transformer encoding. The basic formula is shown in the following equation:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (\text{A.2})$$

where  $Q$ ,  $K$ , and  $V$  are all input word vector matrices.

The fundamental concept of this method is to calculate the interrelationships of each word in a sentence for all words in the sentence and then consider that these word-to-word interrelationships reflect the relatedness and importance of different words in the sentence. These interrelations are used to adjust the weights of each word to obtain a new representation of each word. This new representation implies not only the word itself but also the relationship between other words and the word and is, therefore, a more global representation than a simple word vector.

To avoid the limited information extracted by self-attention, to extend the ability of the model to focus on different locations, and to increase the representation subspace of attention units, the Transformer adopts the Multi-Head model, as shown in the following equation:

$$\text{MultiHead}(Q, K, V) \\ \text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V). \quad (\text{A.3})$$

Here,  $W^o$  is the additional weight matrix that makes  $Q$ ,  $K$ , and  $V$  larger, splits a portion of them equally to each head, and merges the vectors obtained from each head.

In order to ensure efficient gradient transfer in deep networks, the Transformer adds residual network and layer normalization after merging vectors to address the problem of the missing gradient in deep models effectively, and it can break network symmetry, improve network degradation problems, accelerate convergence, and normalize optimization space.

$$x_{i+1} \\ LN(x_i) = \alpha \times \frac{x_i - \mu_L}{\sqrt{\sigma_L^2 + \varepsilon}} + \beta, \quad (\text{A.4})$$

where  $\alpha$  and  $\beta$  are the parameters to be learned and  $\mu$  and  $\sigma$  are the mean and standard deviation of the input layer. The Feed Forward Network (FFN) layer is added for spatial transformation, FFN contains two layers of a linear transformation, and the activation function in the middle is ReLu.

$$FFN(x) = \max(0, xW_1 + b_1)W_2 + b_2. \quad (\text{A.5})$$

Finally, the output vector of the last two steps is combined with another residual operation and normalized to perform the operations in the BERT pretraining task.

### C. Fine-Tuning of BERT-Based Single-Text Classification Tasks

For the text classification task, as shown in Figure 5, the BERT model inputs a [CLS] symbol in front of the text and uses the output vector corresponding to this symbol as the semantic representation of the whole text. A simple softmax classifier is added to the top of BERT to predict the probability of the specific label  $c$ .

$$p(c|h) = \text{softmax}(Wh). \quad (\text{A.6})$$

Here,  $W$  is the task-specific parameter matrix. Once the label probabilities are obtained, there is a basis for multi-granularity sentiment classification of the three decisions.

### Data Availability

The data used to support the findings of the study are available at <https://drive.google.com/drive/folders/1Et1iNHLgmFIRoEXpJ9UIHNUgATagG6aj?usp=sharing>.

### Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

### Acknowledgments

This study was supported by the National Natural Science Foundation of China (NSFC) (72071160 and 62072379), Fundamental Research Funds for the Central Universities (kjcx20210103 and JBK2103016), Financial Intelligence and Financial Engineering Key Lab of Sichuan Province, Chengdu SWUFE Jiaozi Institute of Fintech Innovation Co., Ltd. (cgzh20210204), Research Program of Science and Technology at Universities of Inner Mongolia Autonomous Region (2021GG0164), and Financial Innovation Center of the Southwestern University of Finance and Economics, Joint Lab of Data Science and Business Intelligence at Southwestern University of Finance and Economics.

### References

- [1] A. Madhavan, D. Porter, and D. Weaver, "Should securities markets be transparent," *Journal of Financial Markets*, vol. 8, no. 3, pp. 265–287, 2005.

- [2] P. B. Henry, "Stock market liberalization, economic reform, and emerging market equity prices," *The Journal of Finance*, vol. 55, no. 2, pp. 529–564, 2000.
- [3] Q. Meng, Q. Huang, J. Zhang, and S. Wang, "Does investors' communication with listed firms have information? a study on "hudongyi" online communication system," *China Economic Quarterly*, vol. 19, no. 2, pp. 637–662, 2020.
- [4] J. B. De Long, A. Shleifer, L. H. Summers, and R. J. Waldmann, "Noise trader risk in financial markets," *Journal of Political Economy*, vol. 98, no. 4, pp. 703–738, 1990.
- [5] N. Barberis, A. Shleifer, and R. Vishny, "A model of investor sentiment. We are grateful to the NSF for financial support, and to oliver blanchard, alon brav, john campbell (a referee), john cochrane, edward glaeser, J.B. Heaton, danny kahneman, david laibson, owen lamont, drazen prelec, jay ritter (a referee), ken singleton, dick thaler, an anonymous referee, and the editor, bill schwert, for comments.1," *Journal of Financial Economics*, vol. 49, no. 3, pp. 307–343, 1998.
- [6] P. C. Tetlock, "Giving content to investor sentiment: the role of media in the stock market," *The Journal of Finance*, vol. 62, no. 3, pp. 1139–1168, 2007.
- [7] T. Preis, H. S. Moat, and H. E. Stanley, "Quantifying trading behavior in financial markets using google trends," *Scientific Reports*, vol. 3, no. 1, pp. 1684–1686, 2013.
- [8] Z. Da, J. Engelberg, and P. Gao, "The sum of all fears investor sentiment and asset prices," *Review of Financial Studies*, vol. 28, no. 1, pp. 1–32, 2015.
- [9] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *Journal of Computational Science*, vol. 2, no. 1, pp. 1–8, 2011.
- [10] D. M. Cutler, J. M. Poterba, and L. H. Summers, "Speculative dynamics," *The Review of Economic Studies*, vol. 58, no. 3, pp. 529–546, 1991.
- [11] B. Luo, J. Zeng, and J. Duan, "Emotion space model for classifying opinions in stock message board," *Expert Systems with Applications*, vol. 44, pp. 138–146, 2016.
- [12] X. Zhang, Y. Zhang, S. Wang, Y. Yao, B. Fang, and P. S. Yu, "Improving stock market prediction via heterogeneous information fusion," *Knowledge-Based Systems*, vol. 143, pp. 236–247, 2018.
- [13] L.-C. Yu, J.-L. Wu, P.-C. Chang, and H.-S. Chu, "Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news," *Knowledge-Based Systems*, vol. 41, pp. 89–97, 2013.
- [14] J. Xin, R. Tang, J. J. Lee, Y. liang Yu, and J. Lin, "Deebert: dynamic early exiting for accelerating bert inference," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 2246–2251, Venue, July 2020.
- [15] Y. Y. Yao, "Granular computing and sequential three-way decisions," in *Proceedings of the International Conference on Rough Sets and Knowledge Technology*, Springer, Berlin, Germany, October 2013.
- [16] J. Wang, A. M. Alfosoool, J. Su, X. Fu, and J. Tan, "An intelligent system for rumor recognition and rumor sentiment judgment," in *Proceedings of the 2020 International Conference on Computing, Networking and Communications*, pp. 309–313, IEEE, Big Island, HI, USA, Feb 2020.
- [17] M. L. Mitchell and J. H. Mulherin, "The impact of public information on the stock market," *The Journal of Finance*, vol. 49, no. 3, pp. 923–950, 1994.
- [18] P. C. Tetlock, M. Saar-Tsechansky, and S. Macskassy, "More than words: quantifying language to measure firms' fundamentals," *The Journal of Finance*, vol. 63, no. 3, pp. 1437–1467, 2008.
- [19] R. Johnson and T. Zhang, "Effective use of word order for text categorization with convolutional neural networks," in *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 103–112, Venue, May 2015.
- [20] T. Loughran and B. McDonald, "When is a liability not a liability? textual analysis, dictionaries, and 10-ks," *The Journal of Finance*, vol. 66, no. 1, pp. 35–65, 2011.
- [21] D. Liu, T. R. Li, and D. C. Liang, "Incorporating logistic regression to decision-theoretic rough sets for classifications," *International Journal of Approximate Reasoning*, vol. 55, no. 1, pp. 197–210, 2014.
- [22] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," vol. 1, no. 12, CS224N Project Report, Stanford, CA, USA, 2009.
- [23] P. Nakov, A. Ritter, S. Rosenthal, F. Sebastiani, and V. Stoyanov, "Semeval-2016 task 4: sentiment analysis in twitter," in *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pp. 1–18, San Diego, CA, USA, June 2016.
- [24] X. M. Luo, J. Zhang, and W. J. Duan, "Social media and firm equity value," *Information Systems Research*, vol. 24, no. 1, pp. 146–163, 2013.
- [25] Q. Li, T. J. Wang, P. Li, L. Liu, Q. X. Gong, and Y. Z. Chen, "The effect of news and public mood on stock movements," *Information Sciences*, vol. 278, pp. 826–840, 2014.
- [26] J. F. Yu and J. Jiang, "Adapting bert for target-oriented multimodal sentiment classification," in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, pp. 5408–5414, Macao, China, 2019.
- [27] Y. H. Zhang and W. Zhang, "Improved representations for personalized document-level sentiment classification," in *Proceedings of the International Conference on Database Systems for Advanced Applications*, pp. 769–785, Jeju, Korea, September 2020.
- [28] J. H. Lei, Q. Zhang, J. S. Wang, and H. L. Luo, "Bert based hierarchical sequence classification for context-aware microblog sentiment analysis," in *Proceedings of the International Conference on Neural Information Processing*, pp. 376–386, Sydney, NSW, Australia, 2019.
- [29] W. J. Liu, P. Zhou, Z. R. Wang, Z. Zhao, H. T. Deng, and Q. Ju, "Fastbert: a self-distilling bert with adaptive inference time," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 6035–6044, Venue, 2020.
- [30] R. Tang, Y. Lu, L. Q. Liu, L. L. Mou, O. Vechtomova, and J. Lin, "Distilling task-specific knowledge from bert into simple neural networks," 2019, <https://arxiv.org/abs/1903.12136>.
- [31] Y. Y. Yao, "A triarchic theory of granular computing," *Granular Computing*, vol. 1, no. 2, pp. 145–157, 2016.
- [32] Y. Y. Yao, "Three-way granular computing, rough sets, and formal concept analysis," *International Journal of Approximate Reasoning*, vol. 116, pp. 106–125, 2020.
- [33] Q. H. Zhang, G. H. Pang, and G. Y. Wang, "A novel sequential three-way decisions model based on penalty function," *Knowledge-Based Systems*, vol. 192, Article ID 105350, 2020.
- [34] D. Dai, H. X. Li, X. Y. Jia, X. Z. Zhou, B. Huang, and S. N. Liang, "A co-training approach for sequential three-way decisions," *International Journal of Machine Learning and Cybernetics*, vol. 11, no. 5, pp. 1129–1139, 2020.
- [35] B. H. Long, W. H. Xu, X. Y. Zhang, and L. Yang, "The dynamic update method of attribute-induced three-way granular concept in formal contexts," *International Journal of Approximate Reasoning*, vol. 126, pp. 228–248, 2020.

- [36] J. Yang, G. Y. Wang, Q. H. Zhang, Y. H. Chen, and T. H. Xu, "Optimal granularity selection based on cost-sensitive sequential three-way decisions with rough fuzzy sets," *Knowledge-Based Systems*, vol. 163, pp. 131–144, 2019.
- [37] H. X. Li, L. B. Zhang, B. Huang, and X. Z. Zhou, "Sequential three-way decision and granulation for cost-sensitive face recognition," *Knowledge-Based Systems*, vol. 91, pp. 241–251, 2016.
- [38] A. V. Savchenko, "Fast multi-class recognition of piecewise regular objects based on sequential three-way decisions and granular computing," *Knowledge-Based Systems*, vol. 91, pp. 252–262, 2016.
- [39] X. Yang, Y. Y. Zhang, H. Fujita, D. Liu, and T. R. Li, "Local temporal-spatial multi-granularity learning for sequential three-way granular computing," *Information Sciences*, vol. 541, pp. 75–97, 2020.
- [40] D. Liu, X. Yang, and T. R. Li, "Three-way decisions: beyond rough sets and granular computing," *International Journal of Machine Learning and Cybernetics*, vol. 11, no. 5, pp. 989–1002, 2020.
- [41] X. Yang, T. R. Li, D. Liu, and H. Fujita, "A multilevel neighborhood sequential decision approach of three-way granular computing," *Information Sciences*, vol. 538, pp. 119–141, 2020.
- [42] H. R. Ju, W. Pedrycz, H. X. Li, W. P. Ding, X. B. Yang, and X. Z. Zhou, "Sequential three-way classifier with justifiable granularity," *Knowledge-Based Systems*, vol. 163, pp. 103–119, 2019.
- [43] Y. Y. Yao, "The superiority of three-way decisions in probabilistic rough set models," *Information Sciences*, vol. 181, no. 6, pp. 1080–1096, 2011.
- [44] Y. Y. Yao, "Three-way decisions with probabilistic rough sets," *Information Sciences*, vol. 180, no. 3, pp. 341–353, 2010.
- [45] Y. Y. Yao, "Three-way decision: an interpretation of rules in rough set theory," in *Proceedings of the 4th International Conference on Rough Sets and Knowledge Technology*, Berlin, Germany, July 2009.
- [46] Y. B. Zhang, D. Q. Miao, J. Q. Wang, and Z. F. Zhang, "A cost-sensitive three-way combination technique for ensemble learning in sentiment classification," *International Journal of Approximate Reasoning*, vol. 105, pp. 85–97, 2019.
- [47] H. X. Li, L. B. Zhang, X. Z. Zhou, and B. Huang, "Cost-sensitive sequential three-way decision modeling using a deep neural network," *International Journal of Approximate Reasoning*, vol. 85, pp. 68–78, 2017.
- [48] M. E. Basiri, M. Abdar, M. A. Cifci, S. Nemati, and U. R. Acharya, "A novel method for sentiment classification of drug reviews using fusion of deep and machine learning techniques," *Knowledge-Based Systems*, vol. 198, Article ID 105949, 2020.
- [49] X. Yang, Y. Li, Q. Li, D. Liu, and T. Li, "Temporal-spatial three-way granular computing for dynamic text sentiment classification," *Information Sciences*, vol. 596, pp. 551–566, 2022.
- [50] Y. Zhang, Z. Zhang, D. Miao, and J. Wang, "Three-way enhanced convolutional neural networks for sentence-level sentiment classification," *Information Sciences*, vol. 477, pp. 55–64, 2019.
- [51] F. Shen, X. C. Zhao, and G. Kou, "Three-stage reject inference learning framework for credit scoring using unsupervised transfer learning and three-way decision theory," *Decision Support Systems*, vol. 137, Article ID 113366, 2020.
- [52] S. Maldonado, G. Peters, and R. Weber, "Credit scoring using three-way decisions with probabilistic rough sets," *Information Sciences*, vol. 507, pp. 700–714, 2020.
- [53] J. Wang, Z. Xie, Q. Li et al., "Effect of digitalized rumor clarification on stock markets," *Emerging Markets Finance and Trade*, vol. 55, no. 2, pp. 450–474, 2019.
- [54] Q. Li, T. Wang, Q. Gong, Y. Chen, Z. Lin, and S.-k. Song, "Media-aware quantitative trading based on public web information," *Decision Support Systems*, vol. 61, pp. 93–105, 2014.
- [55] H. X. Li, X. Z. Zhou, B. Huang, and D. Liu, "Cost-sensitive three-way decision: a sequential strategy," in *Proceedings of the International Conference on Rough Sets and Knowledge Technology*, Springer, Berlin, Germany, October 2013.
- [56] L. B. Zhang, H. X. Li, X. Z. Zhou, and B. Huang, "Sequential three-way decision based on multi-granular autoencoder features," *Information Sciences*, vol. 507, pp. 630–643, 2020.
- [57] E. F. Fama and K. R. French, "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, vol. 33, no. 1, pp. 3–56, 1993.
- [58] Y. Liu, M. Ott, N. Goyal et al., "Roberta: A robustly optimized bert pre-training approach," 2019, <https://arxiv.org/abs/1907.11692>.
- [59] P. He, X. Liu, J. Gao, and W. Chen, "Deberta: decoding-enhanced bert with disentangled attention," in *Proceedings of the International Conference on Learning Representations*, Vienna, Austria, 2021.
- [60] T. G. Andersen, T. Bollerslev, F. X. Diebold, and C. Vega, "Real-time price discovery in global stock, bond and foreign exchange markets," *Journal of International Economics*, vol. 73, no. 2, pp. 251–277, 2007.
- [61] C.-H. Chen, P.-Y. Chen, and J. Chun-Wei Lin, "An ensemble classifier for stock trend prediction using sentence-level Chinese news," *Sentiment and Technical Indicators*, vol. 7, pp. 53–64, 2022.