electronics

MDPI

Article

Research on Joint Extraction Model of Financial Product Opinion and Entities Based on RoBERTa

Jiang Liao and Hanxiao Shi *

School of Management and E-Business, Zhejiang Gongshang University, Hangzhou 310018, China; 20020200018@pop.zjgsu.edu.cn
* Correspondence: hxshi@zjgsu.edu.cn

Abstract: With the rapid development of the Internet, and its enormous impact on all aspects of life, traditional financial companies increasingly focus on the user's online reviews, aiming to promote competitiveness and quality of service in the products of this enterprise. Due to the difficulty of extracting comment text compared with structured data itself, coupled with the fact that it is too colloquial, the traditional model insufficiently semantically represents sentences, resulting in unsatisfactory extraction results. Therefore, this paper selects RoBERTa, a pre-trained language model that has exhibited an excellent performance in recent years, and proposes a joint model of financial product opinion and entities extraction based on RoBERTa multi-layer fusion for the two tasks of opinion and entities extraction. The experimental results show that the performance of the proposed joint model on the financial reviews dataset is significantly better than that of the single model.

Keywords: sentiment analysis; entities extraction; RoBERTa; TextCNN; joint model

Citation: Liao, J.; Shi, H. Research on Joint Extraction Model of Financial Product Opinion and Entities Based on RoBERTa. *Electronics* **2022**, *11*, 1345. https://doi.org/10.3390/electronics11091345

Academic Editors: Daniel Hládek, Matúš Pleva, Piotr Szczuko and Andrej Zgank

Received: 1 April 2022
Accepted: 22 April 2022
Published: 23 April 2022

1. Introduction

With the continuous popularity of the Internet, people generate a large amount of data all the time, but most of them are unstructured data such as text, images, and videos. When many users experience the products and services provided by enterprises, they will share their experiences in some online communities, such as Weibo, Zhihu, Xiaohongshu, etc. There are huge economic benefits hidden behind these behavioral data, which have a strong guiding significance for enterprises to improve product management, accurately grasp customer needs, optimize marketing strategies, and improve the core competitiveness of enterprises. However, in the ocean of data, it is very expensive to search and obtain valuable information manually. With the wide application of deep learning, it is possible to explore and use Natural Language Processing (NLP)-related technologies to complete the automatic extraction of review information, which will reduce the cost of data acquisition for enterprises, bring convenience to the improvement and optimization of subsequent products and services, and help enterprises make effective decisions.

This paper extracts the opinion and entities of financial products through the short public reviews of Zhihu, investment and finance, and other related forum communities provided by the financial innovation special track of the 2021 CCF Big Data and Computational Intelligence Competition. It is mainly divided into two sub-tasks: one is sentiment analysis, for example, Bank of China's loan approval is very fast, which shows that the emotional label given by users is positive; the other is named entity recognition, in which the bank and loan are entities.

For these two sub-tasks, this paper considers that traditional models face difficulties in feature selection and cannot effectively represent sentence semantics. In recent years, with the emergence of pre-trained language models such as Bidirectional Encoder Representations from Transformers (BERT) [1], due to their excellent semantic expression ability, can fully capture the context information of sentences and achieve the effect of

the state of the art in many tasks. Considering the correlation between the two sub-tasks, this paper proposes the joint modeling method based on RoBERTa [2], which is the one of most famous pre-trained language models. This is a multi-task learning method that can let the model complement mutual knowledge in joint learning and reduce the complexity of the task. Compared with the single task extraction, the effect of the joint model has significantly improved.

The paper is organized as follows: Section 2 reviews the latest related work on financial text sentiment analysis and Named Entity Recognition. Section 3 gives a detailed information framework for the joint extraction model of opinion and entities. Section 4 reviews the configuration of the proposed experiments. Section 5 presents the results and analysis. Finally, Section 6 gives the conclusions and directions for future research.

2. Related Works

This section mainly introduces the related work in the domain of financial text sentiment analysis and named entity recognition.

2.1. Financial Text Sentiment Analysis

As the basic task of natural language processing, sentiment analysis, also known as tendency analysis and opinion mining, refers to the process of analyzing, processing, and judging the attitude of the subjective text with emotional polarity, and has been widely used in public opinion analysis, recommendation systems and other aspects. Through the analysis of [3,4], it can be concluded that the method of sentiment analysis has developed from the early statistical machine learning to the current deep learning, and has been applied in different fields. With the development of the Internet, people are accustomed to expressing their opinion on related stocks online, and measuring investor sentiment is an important research direction in financial forecasting. Based on the method of emotion dictionary, Bartov et al. [5] analyzed the tweets in the trading days before the earnings announcement, used three different word lists to measure the emotions contained in tweets, and then Sohangir et al. [6] used a domain dictionary to quickly extract the tweets in financial social media (such as Stock Twits). Recently, Mehdi et al. [7] proposed a hybrid method to construct a vocabulary specially used for financial market sentiment analysis to improve the sentiment analysis of the financial environment. Using the establishment of sentiment dictionary, this method can achieve better performance, but it lacks domain transfer and has great domain limitations.

Sentiment analysis based on traditional machine learning only needs to identify the emotional polarity of the text, which can be regarded as a supervised classification problem. Chiong et al. [8] extracted sentiment-related features from financial news and proposed Support Vector Machines (SVM) based on the particle swarm optimization algorithm for financial market prediction. Bourezk et al. [9] used the Naive Bayes algorithm to capture the public's investment sentiment. Gupta et al. [10] analyzed the emotions in financial online community reviews by combining text feature methods such as word bag and Bigram with algorithms such as Logistic Regression, SVM, and the results showed that sentiment has an impact on stock price changes. Compared with traditional sentiment dictionaries, the machine learning method can extract features automatically, and the classification performance is thus further improved.

With the rapid development of computer hardware, deep learning has become a new direction of machine learning. Gite et al. [11] used the emotions carried by news headlines combined with the Long Short-Term Memory (LSTM) model to predict stock prices. Subsequently, Jing et al. [12] used a deep Convolutional Neural Network (CNN) to classify the hidden emotions of investors and then used LSTM model to analyze the technical indicators and sentiment analysis results of the stock market. The experiment showed that the proposed mixed model with emotion significantly outperforms the single model. Wu et al. [13] proposed a stock price prediction method combining multiple data sources and investor sentiment and verified its effectiveness on real datasets. Zhao et al. [14]

propose a sentiment analysis and key entity detection approach based on BERT, which is applied in online financial text mining and public opinion analysis in social media.

Obviously, most of the current financial text sentiment analysis model frameworks are based on deep learning, whose excellent semantic representation ability makes the performance better than traditional machine learning. It is helpful for investment institutions to make reasonable decisions by introducing a sentiment index of stock investors into stock price forecast and evaluating stock value more comprehensively.

2.2. Named Entity Recognition

Named Entity Recognition (NER) refers to extracting entities with special meanings such as institutions, people, opinions, and geographic locations from text-based unstructured data. NER plays an important role in NLP applications such as text comprehension, information retrieval, question answering, and knowledge base. The unsupervised learning methods design rules based on domain-specific entity words and syntactic lexical patterns to extract entities in text. Khaing et al. [15] proposed a financial news extraction model based on named entities and syntactic features. Burdick et al. [16] developed two financial institution NER tools by using the rule-based text extraction platform. Due to the particularity of the financial field, the accuracy of dictionary-based methods often depends on the rules and the number of dictionaries, so there are great limitations.

At present, NER is mainly realized based on machine learning and deep learning algorithms. Commonly used models include the Hidden Markov Model (HMM), Support Vector Machines (SVM) and Conditional Random Field (CRF), etc. Wang et al. [17] identified stock names in financial dictionaries and proposed a named recognition method of CRF that combined mutual information, boundary entropy, and context features, the experimental results showed that the accuracy of the model in finance-related datasets was significantly improved. BiLSTM-CRF is the most common model architecture in NER of deep learning; this architecture was first proposed by Huang et al. [18], which achieved optimal accuracy on relevant datasets. Jayakumar et al. [19] extracted financial entities from documents by introducing a custom financial named entity recognizer. Zhou et al. [20] proposed a two-way Gated Recurrent Unit (GRU) attention mechanism joint model for the problem of financial entity relationship extraction, which solved the difficulty of manual feature selection to a certain extent and improved the computational efficiency. Due to the expensive data labor cost of supervised learning, Liu et al. [21] proposed a semi-supervised method based on BERT and bootstrapping for financial entity recognition.

In recent years, with the deepening of text sentiment analysis and entity extraction in the financial field, remarkable results have been achieved, providing diversified decision support for investors or platforms. However, it is found that most of the models focus on a single aspect of the task, while ignoring the correlation between the two tasks, which leads to more time and economic cost. Therefore, this paper proposes a model of joint extraction of opinion and entities.

3. The Framework of Models

3.1. Introduction to Related Models

3.1.1. Pre-Trained Language Model

Since it is difficult for traditional word vectors (such as Word2vec and Glove) to describe the semantic information of a word in different contexts, in order to solve this problem, researchers propose a context-dependent word vector, namely dynamic word vector. With the pre-trained language model BERT proposed by Devlin et al., which utilizes large-scale unlabeled text data to generate rich contextual representations, it refreshes the best records for 11 NLP tasks and shows strong performance, so NLP entered the era of "Fine-tuning". As an improved version of BERT, RoBERTa [22] adopted more training data, bigger batches, and longer pre-training steps than BERT in the training method, deleted the next sentence prediction task and adopted the strategy of dynamic mask. Each time a sequence is input to the training model, a new mask form will be generated. As data

are continuously input, the model will adapt to different mask strategies. Considering the characteristics of Chinese words, Cui et al. [2] optimized the Chinese pre-trained model officially released by Google and used Whole Word Masking (WWM) technology to mask all the Chinese characters that makeup words, so it can better capture semantic information, improve the generalization ability of the model, and report that RoBERTa has a significant effect on multiple tasks compared to BERT or ERNIE [23].

When inputting a sentence as $X = (x_1, x_2, \ldots, x_n)$, n is the length of the sentence. For the classification and sequence annotation tasks of single sentences, a special character, [CLS], will be inserted at the beginning of the sentence and a special character, [SEP], will be inserted at the end of the sentence as a marker. Therefore, when the input sequence is entered into the model, $X = (cls, x_1, x_2, \ldots, x_n, sep)$, the output of the model is $H = (h_0, h_1, \ldots, h_t), t = n + 1$. For the classification task, the contextual representation h_0 labeled by [CLS] is directly projected to the label prediction layer for classification, while for sequence labeling task the output hidden states $h_1, h_2, \ldots, h_n$ are used for classification on the entity extraction layer. The model is pre-trained on a large-scale untagged text corpus, thus providing a powerful context representation that can be used quickly in downstream tasks with only fine tuning and is significantly better than traditional language models.

3.1.2. TextCNN

Convolutional Neural Networks (CNN) can extract local features. The local features of text are sliding windows composed of several words. CNN can combine and filter the input text features to obtain semantic information of different levels. Before the pre-trained language model was developed, TextCNN [24,25] showed excellent ability in feature extraction of short text, with a significant classification effect and fast calculation speed. Therefore, this paper uses it as an opinion recognition model in a joint extraction task. In TextCNN, the input layer uses all the encoder layer word vectors in the pre-trained language model. To ensure the integrity of words, the width of the convolution kernel is consistent with the dimensions of the word vector. The convolution operation is as shown in Formulas (1) and (2), and then the max-pooling layer is used as shown in Formula (3) to retain the most representative features, and finally, the fully connected layer is used for classification.

$$c_i = f(w \cdot x_{i:i+h-1} + b) \tag{1}$$

$$c = [c_1, c_2, \ldots, c_{n-h+1}] \tag{2}$$

$$\hat{c} = max\{c\} \tag{3}$$

where w is the weight matrix, x is the word vector, b is the bias, and c_i is the feature generated by the text under the window size $(i, i + h - 1)$ through nonlinear functions, such as hyperbolic tangent.

3.2. Joint Extraction Model

In order to solve the extraction of opinion and entities of financial products, considering that multi-task joint learning [26,27] has shown its advantages over single-task models in various fields in recent years, the joint model is used in this paper. The baseline model in this paper uses the joint model proposed by Chen et al. [28], which is fine-tuned by BERT. On the one hand, it is output to the linear layer for intent classification, and on the other hand, it is output to the CRF layer for slot filling.

Because of RoBERTa's powerful feature extraction and semantic representation capabilities, it can be quickly migrated to the joint model of product opinion and entities extraction in this paper. The model framework proposed is mainly composed of a RoBERTa pre-trained language model, TextCNN and CRF, as shown in Figure 1.

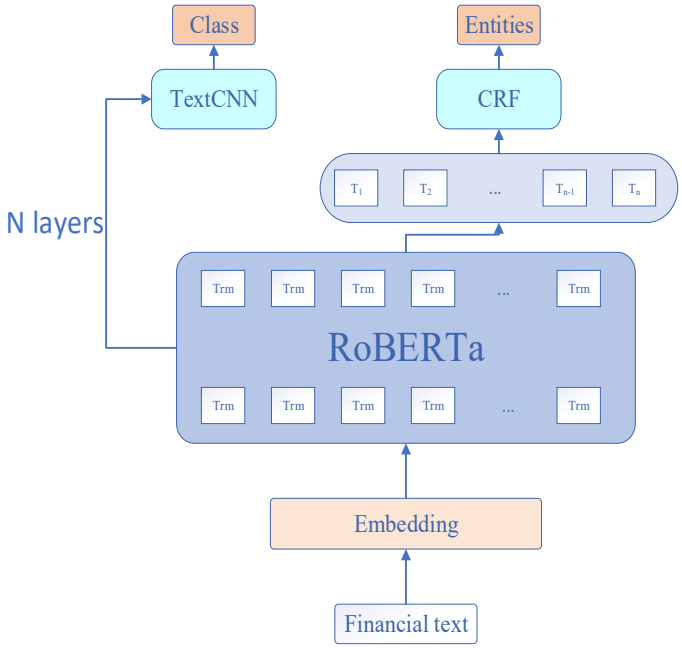

Figure 1. Joint extraction model framework for financial product opinion and entities.

As can be seen from the figure, the financial text outputs the input representation v of RoBERTa through the embedding layer, and then inputs it into the RoBERTa encoder. In the RoBERTa encoder, the input representation v is encoded by multi-layer Transformer and fully learns the semantic association between each word in the sentence with the help of self-attention mechanism. Finally, the contextual semantic representation of the sentence is obtained. In the final model output stage, on the one hand, this paper integrates multi-layer information to output opinion in the TextCNN layer; on the other hand, it outputs entities in the CRF layer.

The joint model proposed in this paper does not directly predict the product opinion according to the last hidden state $\boldsymbol{h}$ of the special character [CLS] at the beginning of a sentence, such as Formula (4). Instead, it is inspired by Jawahar G et al. [29] to capture the different feature information of the sentence in different layers of BERT and other pre-trained language models. The more network layers, the more advanced semantic features are captured. Therefore, the outputs of different layers are extracted and combined together as the input of the TextCNN model in Figure 1, as shown in Formula (5).

$$y^i = softmax\left(W^i h + b^i\right) \tag{4}$$

$$cls\text{-}embeddings = concat(\sum_{i}^{N} hidden\text{-}state_i) \tag{5}$$

where W^i is the weight matrix, h is the final hidden state of the special character [CLS], b^i is the bias, y^i is the classification result, $hidden\text{-}state_i$ is the output hidden state of each layer, $cls\text{-}embeddings$ is the fusion result of hidden state of each layer.

Because the prediction of entity label depends on surrounding words, in the entity extraction part, compared with HMM, CRF can better capture global information and avoid label offset and other problems. Therefore, the last hidden states $H^{entity} = (h_1, h_2, \ldots, h_n)$ of the pre-trained language model are selected, which are input to CRF layer and decoded by the Viterbi algorithm to obtain the entities in the last sentence. In terms of the loss function of the model, the two sub-task losses of the model are weighted and summed (as

shown in Formula (6)), which are trained as the total loss, and the parameters are updated at the same time.

$$total\_loss = \alpha * loss_{class} + \beta * loss_{entity} \quad (6)$$

where α, β are hyper-parameters, which defaults to 1.

The objective function of the financial product opinion and entities joint extraction model is shown in Formula (7), to maximize the probability $p\left(y^{class}, y^{entity} \middle| x\right)$, it is assumed that y^{class} and y^{entity} are independent.

$$p\left(y^{class}, y^{entity} \mid \boldsymbol{x}\right) = p\left(y^{class} \mid \boldsymbol{x}\right) \prod_{n=1}^{N} p\left(y_n^{entity} \mid \boldsymbol{x}\right) \quad (7)$$

where x is the input text vector, y_n^{entity} is the predicted value of every hidden state excluding [CLS], y^{class} and y^{entity} are the predicted results of opinion and entities, respectively.

4. Experiment

4.1. Dateset

This paper aims at the extraction of product opinion and entities in the financial field, so the data provided by the sub-task "Product review Opinion Extraction" of 2021 CCF BDCI were selected from the public review short texts of Zhihu, investment and wealth management, and other relevant forum communities. The data contain three fields in total: Text, BIO_anno, and Class, which, respectively, represent the original text of the review, entity annotation in BIO format, and sentiment classification, as shown in Table 1.

Table 1. Examples of training data.

Text	Do you have industrial and commercial deposits? First card for direct application? That's a good amount. (楼主有工商存款吗? 直接申请的首卡? 额度不错呀)
BIO_anno	O O O B-BANK I-BANK O O O O O O O O O B-PRODUCT I-PRODUCT O B-reviews_N I-reviews_N B-reviews_ADJ I-reviews_ADJ O
Class	1 (Positive)

The task provided a total of 7528 pieces of training data, with positive, negative, and neutral sentiment labels accounting for 3.83%, 10.56%, and 85.61%, respectively, and the average text length is 25. The entity annotation of the dataset includes four categories of bank, product, reviews noun, and reviews adjective, as shown in Table 2.

Table 2. Examples of product entity categories.

Entity Category	
Bank	Industrial and Commercial Bank of China (中国工商银行), The bank of China (中国银行)
Product	Consumer loan (消费贷), The credit card (信用卡)
reviews_ N.	Rate (利率), lines (额度), Credit record (信用记录)
reviews_Adj.	Too slow (太慢), Good (不错)

4.2. Experimental Requirements and Evaluation Indicators

In order to realize the whole extraction process of financial text opinion and entities, this paper fine-tunes RoBERTa and other pre-trained language models. The initial default parameters are set as sequence length = 200, batch size = 16, dropout rate = 0.2, kernel sizes = [2, 3, 4], number of filters = 3, cross-entropy loss function, adamW optimizer, using differential learning rate: 2×10^{-5} on RoBERTa module, 2×10^{-3} on other modules. During the experiment, this paper uses early stopping to terminate the training and retain the optimal parameters and results under the condition that the loss does not decrease

many times. The dataset is randomly divided into 6775 training set and 753 test set (neutral samples accounted for 83.79%), and 10-fold cross-validation was used.

In this paper, the evaluation index of sub-task 1 product opinion extraction is class accuracy (Formula (8)). The number of False positives (FP), False negatives (FN), and True positives (TP) in the confusion matrix were used to compute Precision (Formula (9)), Recall (Formula (10)) and *F*1-score (Formula (11)), and *F*1-score was used for the evaluation index of entity extraction for sub-task 2. Sentence accuracy (Formula (12)) was adopted for input sentences if and only if both opinion and entities extraction were correct.

$$class\text{-}acc. = \frac{y_{pred}}{y_{true}} \tag{8}$$

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

$$F1\text{-}score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{11}$$

$$sentence\text{-}acc. = \frac{\text{number of opinion and entities that are correct}}{\text{total number of data}} \tag{12}$$

5. Results Analysis

In order to illustrate the effectiveness of the joint model proposed in this paper compared with the single-task model, the joint model was divided into two sub-models, used RoBERTa-wwm-ext for 10-fold cross-validation of the two sub-tasks of opinion and entities extraction, respectively, and reported the average results on the test set. The results are shown in Table 3.

Table 3. Comparison results of single model and joint model.

Model	Class Acc. (%)	Entity F1 (%)	Sent. Acc. (%)
Single	86.31	90.67	67.64
Joint (Ours)	88.75 (+2.44)	90.51 (–0.16)	69.99 (+2.35)

It can be seen from the table that the joint model proposed in this paper is 2.44%, 2.35% higher than the single model in opinion extraction accuracy and sentence accuracy. Although the score of entities extraction has dropped slightly, it should be noted that sentence accuracy is the gold standard. Then, validating the performance improvement with statistical significance test for experiments, where t-test is performed to measure whether the results from the proposed model are significantly better than ones from single model. The results of classification and sentence accuracy indicate that improvement is significant with $p < 0.05$. So, the joint model has a significant improvement in sentence accuracy compared with the single model, indicating that the sharing of model parameters is conducive to reducing the propagation of error and complementing knowledge among different tasks.

In the next experiments, three pre-trained language models BERTbase, ERNIE1.0 and RoBERTa-wwm-ext are used, respectively, in the baseline joint model [28] and the joint model proposed in this paper in the test set and report the average results on the test set. The results are shown in Table 4.

Table 4. Results under different pre-trained language models.

Model	Class Acc. (%)	Entity *F*1 (%)	Sent. Acc. (%)
Joint(BERT)	87.45	90.32	67.82
Ours(BERT)	88.54	90.44	68.94(+1.12)
Joint(ERNIE)	87.98	90.73	68.80
Ours(ERNIE)	88.60	90.86	69.84(+1.04)
Joint(RoBERTa)	87.89	90.57	68.41
Ours(RoBERTa)	88.75	90.51	69.99(+1.58)

It can be clearly seen from the table that the Sentence Accuracy of the three pre-trained language models proposed in this paper is 68.94%, 69.84%, and 69.99%, respectively, which are significantly improved compared with the baseline model, being 1.12%, 1.04%, and 1.58%, respectively. Since this paper uses the multi-layer fusion method as the input vector of TextCNN, the accuracy of the opinion extraction is improved significantly, especially by 1.58% under the RoBERTa pre-trained language model, which further explains the effectiveness of the model improvement in this paper.

6. Conclusions

With the rapid development of the Internet, traditional financial enterprises pay more and more attention to users' online reviews. Full mining review information plays a crucial role in enterprises improving products and services and managers' making reasonable decisions. However, regarding the online reviews of financial products, previous research only focused on the extraction of the two single tasks of opinion and entities, respectively. Because the joint model with shared parameters is beneficial to reduce the propagation of errors and complement each other's knowledge, the pre-trained language model has excellent feature extraction and semantic representation capabilities. Therefore, this paper proposes a RoBERTa-based financial product opinion and entities joint extraction model. The experimental results on relevant dataset show that the proposed joint model has significantly improved performance compared with the single model, especially in capturing product opinion. Of course, the point of opinion extraction in this paper is only limited to the sentence level, and there may be multiple points of opinion in some sentences, so further fine-grained analysis of entities such as products or services are needed. In addition, the proposed joint model is also applicable to other fields.

Author Contributions: Conceptualization, J.L. and H.S.; methodology, J.L. and H.S.; validation, J.L.; writing—original draft preparation, J.L.; writing—review and editing, H.S.; supervision, H.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partly supported by the Zhejiang Provincial Natural Science Foundation of China (No. LY19F020007).

Data Availability Statement: Publicly available datasets were analyzed in this study. These data can be found here: https://www.datafountain.cn/competitions/529, accessed date on 23 April 2022.

Acknowledgments: The authors would like to thank the reviewers for their invaluable comments and suggestions, which greatly helped to improve the presentation of this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the NAACL-HLT 2019, Minneapolis, MN, USA, 2–7 June 2019; pp. 4171–4186.
2. Cui, Y.; Che, W.; Liu, T.; Qin, B.; Yang, Z. Pre-training with whole word masking for chinese bert. *IEEE/ACM Trans. Audio Speech Lang. Processing* **2021**, *29*, 3504–3514. [CrossRef]
3. Batra, S.; Rao, D. Entity based sentiment analysis on twitter. *Science* **2010**, *9*, 1–12.

4. Zhang, L.; Wang, S.; Liu, B. Deep learning for sentiment analysis: A survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2018**, *8*, e1253. [CrossRef]
5. Bartov, E.; Faurel, L.; Mohanram, P.S. Can Twitter help predict firm-level earnings and stock returns? *Account. Rev.* **2018**, *93*, 25–57. [CrossRef]
6. Mishev, K.; Gjorgjevikj, A.; Vodenska, I.; Chitkushev, L.T.; Trajanov, D. Evaluation of sentiment analysis in finance: From lexicons to transformers. *IEEE Access* **2020**, *8*, 131662–131682. [CrossRef]
7. Yekrangi, M.; Abdolvand, N. Financial markets sentiment analysis: Developing a specialized Lexicon. *J. Intell. Inf. Syst.* **2021**, *57*, 127–146. [CrossRef]
8. Chiong, R.; Fan, Z.; Hu, Z.; Adam, M.T.; Lutz, B.; Neumann, D. A sentiment analysis-based machine learning approach for financial market prediction via news disclosures. In Proceedings of the Genetic and Evolutionary Computation Conference Companion, Kyoto, Japan, 15–19 July 2018; pp. 278–279.
9. Bourezk, H.; Raji, A.; Acha, N.; Barka, H. Analyzing Moroccan Stock Market using Machine Learning and Sentiment Analysis. In Proceedings of the 2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), Meknes, Morocco, 16–19 April 2020; pp. 1–5.
10. Gupta, R.; Chen, M. Sentiment analysis for stock price prediction. In Proceedings of the 2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), Shenzhen, China, 6–8 August 2020; pp. 213–218.
11. Gite, S.; Khatavkar, H.; Kotecha, K.; Srivastava, S.; Maheshwari, P.; Pandey, N. Explainable stock prices prediction from financial news articles using sentiment analysis. *PeerJ Comput. Sci.* **2021**, *7*, e340. [CrossRef] [PubMed]
12. Jing, N.; Wu, Z.; Wang, H. A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction. *Expert Syst. Appl.* **2021**, *178*, 115019. [CrossRef]
13. Wu, S.; Liu, Y.; Zou, Z.; Weng, T.-H. S_I_LSTM: Stock price prediction based on multiple data sources and sentiment analysis. *Connect. Sci.* **2022**, *34*, 44–62. [CrossRef]
14. Zhao, L.; Li, L.; Zheng, X.; Zhang, J. A BERT based sentiment analysis and key entity detection approach for online financial texts. In Proceedings of the 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Dalian, China, 5–7 May 2021; pp. 1233–1238.
15. Khaing, E.T.; Thein, M.M.; Lwin, M.M. Stock trend extraction using rule-based and syntactic feature-based relationships between named entities. In Proceedings of the 2019 International Conference on Advanced Information Technologies (ICAIT), Yangon, Myanmar, 6–7 November 2019; pp. 78–83.
16. Burdick, D.; De, S.; Raschid, L.; Shao, M.; Xu, Z.; Zotkina, E. resMBS: Constructing a financial supply chain from prospectus. In Proceedings of the Second International Workshop on Data Science for Macro-Modeling, San Francisco, CA, USA, 26 June–1 July 2016; pp. 1–6.
17. Wang, S.; Xu, R.; Liu, B.; Gui, L.; Zhou, Y. Financial named entity recognition based on conditional random fields and information entropy. In Proceedings of the 2014 International Conference on Machine Learning and Cybernetics, Lanzhou, China, 13–16 July 2014; pp. 838–843.
18. Huang, Z.; Xu, W.; Yu, K. Bidirectional LSTM-CRF models for sequence tagging. *arXiv* **2015**, arXiv:1508.01991.
19. Jayakumar, H.; Krishnakumar, M.S.; Peddagopu, V.V.V.; Sridhar, R. RNN based question answer generation and ranking for financial documents using financial NER. *Sādhanā* **2020**, *45*, 1–10. [CrossRef]
20. Zhou, Z.; Zhang, H. Research on entity relationship extraction in financial and economic field based on deep learning. In Proceedings of the 2018 IEEE 4th International Conference on Computer and Communications (ICCC), Chengdu, China, 7–10 December 2018; pp. 2430–2435.
21. Liu, Y.; Li, X.; Shi, J.; Zhang, L.; Li, J. Named entity recognition using a semi-supervised model based on bert and bootstrapping. In Proceedings of the China Conference on Knowledge Graph and Semantic Computing, Nanchang, China, 12–15 November 2020; pp. 54–63.
22. Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. Roberta: A robustly optimized bert pretraining approach. *arXiv* **2019**, arXiv:1907.11692.
23. Sun, Y.; Wang, S.; Li, Y.; Feng, S.; Chen, X.; Zhang, H.; Tian, X.; Zhu, D.; Tian, H.; Wu, H. Ernie: Enhanced representation through knowledge integration. *arXiv* **2019**, arXiv:1904.09223.
24. Kim, Y. Convolutional Neural Networks for Sentence Classification. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1746–1751.
25. Zhang, Y.; Wallace, B. A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. *arXiv* **2015**, arXiv:1510.03820.
26. Kendall, A.; Gal, Y.; Cipolla, R. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7482–7491.
27. Qin, L.; Liu, T.; Che, W.; Kang, B.; Zhao, S.; Liu, T. A co-interactive transformer for joint slot filling and intent detection. In Proceedings of the ICASSP 2021—2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 6–11 June 2021; pp. 8193–8197.

28. Chen, Q.; Zhuo, Z.; Wang, W. Bert for joint intent classification and slot filling. *arXiv* **2019**, arXiv:1902.10909.
29. Jawahar, G.; Sagot, B.; Seddah, D. What Does BERT Learn about the Structure of Language? In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July 2019; pp. 3651–3657.