Fuzzy Systems and Data Mining IX
A.J. Tallón-Ballesteros and R. Beltrán-Barba (Eds.)

doi:10.3233/FAIA231046

The Impact of Sarcasm Contained in Online Media Disinformation Headlines on Enterprise Value: A Study of a Leading Internet Company

Ting WU[a], Guang YU[a, 1], Qiushi WANG[b], Chang LIU[b] and Qiuping CHEN[c]
[a] *School of Management, Harbin Institute of Technology, P.R.China*
[b] *School of Mathematics, Harbin Institute of Technology, P.R.China*
[c] *Zhiwei Research Institute, Beijing, P.R.China*

Abstract. Disinformation in online media has the profound impact on enterprise value. And faced with massive online information, online readers usually only browse through headlines. Therefore, even few negative words including sarcasm within headlines can still impact companies. Given the increasing importance of headlines in online media, it is crucial to examine sarcasm contained in disinformation headlines impact on enterprise value. However, that has received comparatively less attention in the existing literature. Therefore, the primary aim of this article is to address this gap by examining the impact of sarcasm contained in online media disinformation headlines on enterprise value. This article takes a leading Internet company as an example, and combines enterprise value related theory, machine learning, deep learning, pre-training model to mine sarcasm contained in disinformation headlines. Moreover, it uses regression analysis methods to study whether sarcasm contained in online media disinformation headlines will have an impact on the closing price, and further study whether it will exert enterprise value. The results indicate that sarcasm contained in disinformation headlines has a significant impact on enterprise value. This study enriches the relevant research on the impact of sarcasm contained in disinformation headlines on enterprise value, and provides support for companies to reduce enterprise value losses caused by online disinformation.

Keywords. Disinformation, Enterprise value, Sarcasm, Pre-training model

1. Introduction

As an important type of information pollution, fake news in online media has had serious impacts on national security, social stability, business management and other aspects [1-5]. Bodies of research mainly focus on the fake news influence on political system and democratic order [4, 6-11], and some studies pay attention to the field of health care [12]. It's also important to recognize the impact of fake news in online media on companies' overall well-being. In recent years, researchers have gradually focused on the impact of

[1] Corresponding author, Guang YU, School of Management, Harbin Institute of Technology, 92 Xidazhi Street, Nangang District, Harbin, P.R.China; E-mail: yug@hit.edu.cn
This research was funded by the National Natural Science Foundation of China, Grant No. 72074060.

fake news on companies. Some researchers are interesting in the impact of fake news on companies' brand image and employees, and pay attention to companies' solutions to fake news[13-15]. But there are few studies which paid attention to its impact on enterprise value. Wu YAN et al. take the clarification reports of Listed Companies in Shanghai and Shenzhen stock markets in 2000-2009 as samples. They find that the release and dissemination of fraudulent information affects the stock volatility, and there is an obvious information manipulation phenomenon [16]. Chih-chien WANG et al. conduct a study on a Taiwanese company and find that its uses false news to affect its stock price and trade volume [17]. In these above enterprise value studies, information is released by companies themselves.

It should be noted that people usually get useful or interesting information through headlines and tweets (we collectively call them headlines) efficiently. Benjamin D. Horne et al. show that the use of specific nouns in headlines is of great significance for the public to distinguish true information from fake news [18]. Sarcasm also attracts people's attention and affects people's judgment. Maynard D et al. study the effect that sarcasm has on sentiment in tweets, and carry out a number of rules to improve the accuracy of sentiment analysis [19]. Therefore, even few negative words including sarcasm within headlines can still impact companies. Given the increasing importance of headlines in online media, it is crucial to examine their impact on enterprise value.

However, as highlighted by previous studies, it is noteworthy that the impact of sarcasm within fake news headlines in online media on enterprise value has received comparatively less attention in the existing literature. In addition, although there is a wealth of research on fake news, it cannot fully describe the complex and changeable phenomenon of information pollution. A concept of information disorder is proposed to describe information pollution, and three different types of information disorder are defined, namely mis-, dis-, and mal-information. Claire Wardle et al. describe the differences between three types by using dimensions of harm and falseness [20]. The details are shown in Table 1. This study is interesting in disinformation which is false and intended to be harmful to the target.

Table 1. Three types of information disorder

	Falseness?	Harm?
Dis-information	Yes	Yes
Mis-information	Yes	No
Mal-information	No	Yes

Therefore, the primary aim of this study is to address this gap by examining the impact of sarcasm contained in online media disinformation headlines on enterprise value. This study takes a leading Internet company in China as an example. The data comes from Zhiwei Research Institute [21] and China Stock Market & Accounting Research Database [22], respectively. We collect data from January 2018 to October 2021. And we classify sarcasm contained in disinformation headlines by machine learning, deep learning and pre-training algorithms. Further, we use statistical methods, to study the impact of sarcasm contained in disinformation headlines on enterprise value.

Firstly, we use python data processing module and natural language processing technology for disinformation headlines preprocessing and word vectorization.

Secondly, we classify sarcasm contained in disinformation headlines. Sarcasm is considered as an implicit form of sentiment, and conveys the opposite of the intended meaning [23]. We label headlines as sarcasm when they do not express attitudes and sentiments directly. We select a training dataset of disinformation headlines, and label headlines that do not directly express attitudes and emotions as sarcasm. Then machine learning, deep learning and pre-training algorithms are used to train model. The best performing algorithm is selected for classification model. And it is used to classify unlabeled headlines.

Finally, this paper adopts regression model to analyze the impact of sarcasm contained in disinformation headlines on enterprise value. This paper defines dependent variable as company stock closing price. According to the study of YAN Wu et al. [16], and volume-price relationship, control variables are defined as Hong Kong base interest rate, Hong Kong's Hang Seng Index, and trading volume. Independent variable is defined as whether sarcasm contained in disinformation headlines. Then we establish regression models of the impact of sarcasm contained in disinformation headlines on company stock closing price.

The structure of this paper is as follows: the second part describes the study methods and analyzes the experimental results, including data collection and preprocessing, classification of sarcasm in disinformation headlines, as well as sarcasm impact analysis. The third part analyzes the result. Finally, we conclude and discuss this study.

2. Methodology

Enterprise value consists of two parts: equity capital value and debt value. When the capital structure of a company remains unchanged, the greater overall enterprise value, the greater equity capital value, and the higher stock value. Stock value determines stock price, so enterprise value determines stock price, namely market value. While for debt value, when risk is small, the fluctuation of market price is very small. Stock price of a company reflects enterprise value to a certain extent[24-26]. In this study, closing price is used to replace stock value, and is taken as dependent variable to measure enterprise value.

On the other hand, sarcasm is defined as a more aggressive type of irony. It usually expresses dissatisfaction, criticism, or opposition through mock and ridicule. Aggressive tone, expression, and intention make it a suitable irony type to conceal negative information. [27] So we take sarcasm contained in disinformation headlines as independent variable. In addition, according to volume-price relationship, local base interest rate, Hong Kong's Hang Seng Index and trading volume are taken as control variables because they all affect stock closing price [16].

2.1. Data collection and preprocessing

2.1.1. Data collection

This paper collects the leading Internet company disinformation headlines in online media from Zhiwei Research Institute [21]. Zhiwei Research Institute is an intelligence big data company, and provides leading companies with multi-level services including intelligence perception, intelligence analysis and intelligence think tank. There are two elements in judging whether it is disinformation [20]. One is whether the author intends to cause harm to a company, and the other is whether the information is false. We collect disinformation headlines in five most representative online media platforms (including Network Media, Headline, Weibo, WeChat and Self-media) from January 2018 to

October 2021. And data dimension includes disinformation headlines, release time, release platform, full-text link, relevant information of posters, etc. Firstly, all the company relevant reports headlines on five online media platforms are obtained through keyword search. Secondly, we preliminarily screen company negative reports dataset based on account background check, account historical posting pattern, and expert judgment. This dataset contains all information intended to cause harm to the company. Finally, we judge the information as falseness by two standards. One is that the company issues a response and provides corresponding evidence to prove that the information is wrong. The second is that the company doesn't respond, but in the public information, there is no factual basis to support the information.

Then we get dependent variable and control variables from CSMAR database [22]. CSMAR Database is an economic and financial database combining with the actual situation of China. It comes from the needs of academic research, drawing on the professional standards of CRSP and databases of the University of Chicago. The database covers the fields of economy, listed companies, stocks, Hong Kong stocks and so on, which provides a great help for the research of colleges and universities. The heading Internet company is listed on the Hong Kong Stock Exchange. We collect dependent variable and control variables from CSMAR database from CSMAR Database. Data dimensions include closing price, opening price, trade volume, stock yield, closing price and opening price of Hang Seng Index, etc. Finally, the Hong Kong local interest rate from 2018 to 2021 is obtained from Hong Kong Monetary Authority[28].

2.1.2. Data preprocessing

Natural language processing technology is used for disinformation headlines extraction, preprocessing and word segmentation. First of all, we use pandas tool set (https://pandas.pydata.org/) to extract the company disinformation headlines and release time, and extract the company stock closing price, trade volume and closing price of Hang Seng Index, and store them together with the Hong Kong local interest rate.

Secondly, there are missing, abnormal and noise data in the disinformation headlines. This paper uses pandas tool set to clean them. Specifically, we fill in missing data, replace abnormal data with average value, and delete the noise data. There are 8713 disinformation headlines available after cleaning.

Finally, Jieba Word Segmentation Library (https://pypi.org/project/jieba/) of GitHub opensource community is used for disinformation headlines word segmentation. Jieba Word Segmentation Library is a third-party library of Chinese word segmentation with outstanding performance. It uses Chinese word segmentation to determine the probability of association between Chinese characters to generate the word segmentation results. This article uses Jieba Word Segmentation Library to generate headlines word segmentation results, and after removing stop words, the disinformation headlines dictionary is formed.

2.2. Sarcasm classification models

2.2.1. Dictionary construction based on word2vec

Before classification, we need to vectorize disinformation headlines. This study uses Word2vec word vector model released by Google in 2013. Word2Vec learns semantic knowledge in an unsupervised manner from a large number of text corpora. Because it has achieved good results [29-31], it is widely used in natural language processing (NLP). Among them, CBOW model uses the context word to predict the head word to obtain word vector, which can accurately obtain the semantic relationship between words and words [29, 30]. This study uses CBOW model to vectorize disinformation headlines.

2.2.2. Sarcasm classification algorithms

In this paper, disinformation headlines attitudes are divided into three categories: sarcasm, direct smear, and others. According to the definition of sarcasm, we label disinformation headlines as sarcasm which contain mock, ridicule and strong emotions. And we label disinformation headlines as direct smear which contain straightforward blackening and strong emotions. And the remaining headlines are labeled as others. This paper uses machine learning, deep learning and pre-training model, namely Decision tree [32], Random forest [33], K-Nearest Neighbor (KNN) [34], Long Short-Term Memory (LSTM) [35], and Chinese-roberta-wwm-ext [36] to conduct disinformation headlines attitude analysis. This paper compares these five algorithms and selects the best for classification prediction.

Decision tree is a tree structure, in which an internal node represents a judgment on attribute, a branch represents a judgment result, and a leaf node represents a classification result. It is established according to minimizing loss function, and usually includes three steps, namely feature selection, decision tree generation and pruning [32]. Decision tree is easy to interpret and can achieve intuitive visualization, but is prone to overfitting.

Random forest can be seen as a set of decision trees, and it largely solves the overfitting problem. It uses a decision tree as a learner and introduces random attribute selection during the training of a decision tree. Each decision tree in the Random forest is judged separately, and then final classification result is selected by voting [33]. Random forest has good generalization performance and high accuracy, and is less prone to overfitting due to the introduction of randomness.

KNN is a special machine learning algorithm with supervised learning. It achieves classification by measuring the distance between different eigenvalues. The core idea of this method is that if most of the k nearest samples in the feature space belong to a certain category, then the sample also belongs to this category [34]. KNN is easy to understand, and is simpler than other algorithms because of not requiring parameter estimation and training. But its time complexity and spatial complexity are higher than others.

LSTM is a popular deep learning algorithm. It is a structural variant of Recurrent Neural Network (RNN), which can well solve the long-term dependence problem. Similar to RNN, LSTM is also a chain recurrent network structure. The difference is that there is only one network layer in the standard RNN network unit, and the LSTM has four network layers. The core of LSTM is the cell state which is controlled by input gate, forget gate and output gate. Among them, the most important is forget gate, which determines which previous memories will be retained or removed, so the LSTM has the function of long-term memory [35].

Due to the presence of forget gate in LSTM, it still cannot effectively handle the matter of long-distance semantic dependence. Transformer based on self-attention structure [37] can effectively solve the above problem. It is one of the best performing structures in natural language processing (NLP). In recent years, pre-training models based on transformer, such as Bidirectional Encoder Representations from Transformers (BERT) [38], a robustly optimized BERT pretraining approach (RoBERTa) [39], have achieved the best performance in NLP. Meanwhile, Chinese pre-training model performs better in Chinese NLP tasks than Chinese model of BERT, as it can better reflect the semantic connections of Chinese. This study chooses a representative Chinese pre-training model, namely the whole-word mask RoBERTa-wwm-ext, which proposed by the Joint Laboratory of HIT and iFLYTEK Research (HFL) in 2019 [36]. Its performance is significantly better than that of Chinese model of BERT. This model is based on Google's WWM technology, Facebook's RoBERTa pre-training model, and uses HIT Language Technology Platform (LTP) as a word segmentation tool.

2.2.3. Sarcasm classification performance

Firstly, in order to ensure data balance, a hierarchical sampling method is used to construct the training dataset. Among them, sarcasm, direct smear and other types is 1: 1: 1. There are 1274 data after sampling. Secondly, we use 10-folds cross validation to evaluate Decision tree, Random forest and KNN classification accuracy. For LSTM and pre-training model, training set and development set are divided according to the ratio of 7:3. This section evaluates the classification performance four classification algorithms except pre-training model in four words vector length (namely, 50, 100, 150 and 200 dimensions). For pre-training model, we use its built-in word vectorization function to generate word vectors. The classification accuracies are shown in Table 2.

As we can see, the classification accuracy of Chinese-roberta-wwm-ext is the highest, which is 75.64%. Therefore, this article will use Chinese-roberta-wwm-ext to classify unlabeled sarcasm in disinformation headlines.

Table 2. Classification accuracies of sarcasm in disinformation headlines

word vector length	50	100	150	200
Decision tree	62.06%	65.67%	64.18%	66.22%
Random forest	69.05%	70.15%	69.91%	68.50%
KNN	71.78%	71.09%	71.88%	70.22%
LSTM	72.61%	73.64%	72.09%	74.16%
sentence vector length	768			
Chinese-roberta-wwm-ext	75.64%			

2.3. Sarcasm impact analysis

Next, a linear regression model is used to analyze the impact of sarcasm contained in disinformation headlines on company stock closing price. As mentioned above, dependent variable is company stock closing price, control variables are Hong Kong base interest rate, Hong Kong's Hang Seng Index, and trading volume, and independent variable is that whether sarcasm contained in disinformation headlines. We establish a regression model as Eqs. (1):

$$CP = \alpha_1 HIS + \alpha_2 BIR + \alpha_3 Vol + \alpha_4 Sar \quad (1)$$

Among them, CP is company stock closing price, HIS is Hang Seng Index, BIR is base interest rate, Vol is trading volume, and Sar is that whether sarcasm contained in disinformation headlines.

3. Results

In this study, Model 1 investigates the impact of Hong Kong base interest rate and Hang Seng Index on the company stock closing price. On this basis, Model 2 and Model 3 respectively add trading volume and that whether sarcasm contained in disinformation headlines to Model 1. Finally, Model 4 considers the impact of all variables. The linear regression results are shown in Table 3.

It can be seen from Table 3 that the R^2 of four models are about 0.8, indicating that four models have a good fitting on the sample data, and the models have a good interpretation ability. Secondly, the coefficient of Hang Seng Index and Hong Kong base interest rate are significant at the level of 0.01. The coefficient of trading volume is also significant at the level of 0.01. But that whether sarcasm contained in disinformation headlines (Sar) is significant at the level of 0.1. In addition, the coefficient of Sar is positive, indicating that the more sarcasm, the greater the negative impact it will have on stock closing prices. That is, sarcasm has a greater impact on stock closing prices than direct smear.

Table 3. Linear regression results

	Model 1	Model 2	Model 3	Model 4
HSI	0.020863*** (26.45)	0.0210923*** (27.07)	0.0209015*** (26.51)	0.0211376*** (27.17)
BIR	-135.139*** (-60.72)	-136.5661*** (-61.75)	-135.3533*** (-60.77)	-136.8203*** (-61.82)
Vol	-	-5.03*** (-5.24)	-	-5.09*** (-5,31)
Sar	-	-	0.227413* (1.64)	0.2519367* (1.84)
R-squared	0.7998	0.8055	0.8004	0.8062
Adjusted R-squared	0.7994	0.8049	0.7997	0.8054
F-statistic	1873.81	1293.60	1252.33	973.50

Note: * indicates that the 0.1 level is significant; ** indicates that the 0.05 level is significant; *** indicates that the 0.01 level is significant; the value in parentheses is the estimated T statistic; '-' indicates that the corresponding explanatory variable is not included in the model.

4. Conclusion

This paper takes a leading Internet company in China as example. And examines the impact of sarcasm contained in online media disinformation headlines on enterprise value by adopting computer algorithms and using statistical methods. Firstly, we select the best performing algorithm from machine learning, deep learning, and pre-training models to classify sarcasm contained in disinformation headlines. Secondly, we use linear regression analysis to explore the impact of sarcasm on the company stock closing price.

After analysis, we find that compared to machine learning and deep learning algorithms, pre-training model performs better in sarcasm classification. At the same time, we also verify that compared to direct smear, sarcasm contained in disinformation headlines has a more significant negative impact on the company stock closing price.

Further analysis shows that sarcasm contained in online media disinformation headlines has a certain negative impact on enterprise value.

This study emphasizes the importance of negative words such as sarcasm in online media disinformation headlines in influencing factors of enterprise value and managing corporate public opinion, and enriches the theory of enterprise value. Practically, on the basis of paying attention to the impact of disinformation, companies should pay more attention to the significant impact of negative words such as sarcasm, and take timely and effective public relations measures minimize the impact of disinformation on enterprise value. Also, this study provides new methods such as pre-training model to mine potential influencing factors of enterprise value. And the data comes from specific company application scenarios, which enriches the experience of combining research with corporate practice.

Although the current research results are positive, this study still has limitations. This paper only considers a specific company. Future research can compare multiple companies in different industries, and carry out comparative analysis of the impact of sarcasm contained in disinformation on enterprise value. And this paper only focuses on disinformation headlines, follow-up research can take the full text of disinformation into consideration. In addition, sarcasm is a type of emotion, and it may remind us to pay attention to the potential impact of online readers other negative psychological changes on companies.

References

[1] Meel P, Vishwakarma DK. Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities. Expert Systems with Applications 2020;153:112986.

[2] Lazer D, Baum MA, Benkler Y, Berinsky AJ, Greenhill KM, Menczer F, et al. The science of fake news. Science 2018;359(6380):1094-6.

[3] Kapantai E, Christopoulou A, Berberidis C, Peristeras V. A systematic literature review on disinformation: Toward a unified taxonomical framework. New Media & Society 2021;23(5):1301-26.

[4] Aimeur E, Amri S, Brassard G. Fake news, disinformation and misinformation in social media: a review. Social Network Analysis and Mining 2023;13(1).

[5] White A. Review essay: fake news, and online misinformation and disinformation Lie machines, by Philip N. Howard, New Haven and Oxford, Yale University Press, 2020, xviii+221 pp., 20 pound (hardback), ISBN 978-0-300-25020-6. Information Communication & Society 2022;25(11):1669-75.

[6] Bennett WL, Livingston S. The disinformation order: Disruptive communication and the decline of democratic institutions. Eur J Commun 2018;33(2):122-39.

[7] Bovet A, Makse HA. Influence of fake news in Twitter during the 2016 US presidential election. Nat Commun 2019;10:14.

[8] Murphy G, Loftus EF, Grady RH, Levine LJ, Greene CM. False Memories for Fake News During Ireland's Abortion Referendum. Psychol Sci 2019;30(10):1449-59.

[9] Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science 2018;359(6380):1146.

[10] Samaras G. The Capitol Riots. Digital Media, Disinformation, and Democracy Under Attack. International Journal of Press-Politics 2023.

[11] Wang Y. Politics of Disinformation: The Influence of Fake News on the Public Sphere. Mass Communication and Society 2023.

[12] Iacobucci G. Vaccination: "fake news" on social media may be harming UK uptake, report warns. BMJ-British Medical Journal 2019;364:1.

[13] Lee LW, Hannah D, McCarthy IP. Do your employees think your slogan is "fake news?" A framework for understanding the impact of fake company slogans on employees. Journal of Product and Brand Management 2020;29(2):199-208.

[14] Martin-Herrera I, Micaletto Belda JP. Effects of disinformation on the brand image and the reaction

of three companies in the food sector to a communication crisis: Starbucks, Mercadona and Burger King. Obra Digital-Revista De Comunicacion 2021(20):49-66.
[15] Obada D-R, Dabija D-C. The Mediation Effects of Social Media Usage and Sharing Fake News about Companies. Behavioral Sciences 2022;12(10).
[16] Wu Y, Dong CY. Study of Stock Volatility Affected by Fraudulent Information:Evidence from Shanghai and Shenzhen Stock Exchanges. Contemporary Finance & Economics 2010.
[17] Wang CC, Chiang CY. Analyzing Online Fake Business News Communication and the Influence on Stock Price: A Real Case in Taiwan. Journal of Information Technology Applications & Management 2019;26.
[18] Horne BD, Adali S. This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News. 2017.
[19] Maynard DG, Greenwood MA. Who cares about sarcastic tweets? Investigating the impact of sarcasm on sentiment analysis. *Language Resources and Evaluation Conference (LREC).* 2014.
[20] Wardle C, Derakhshan H. Information disorder: Toward an interdisciplinary framework for research and policymaking. Council of Europe Strasbourg; 2017.
[21] ZhiweiData. Zhiwei Data Research Institute. 2012:www.research.zhiweidata.com.
[22] CSMAR. China Stock Market & Accounting Research Database. 2020:www.cn.gtadata.com.
[23] Sundararajan K, Palanisamy A. Multi-rule based ensemble feature selection model for sarcasm type detection in twitter. Computational intelligence and neuroscience 2020;2020.
[24] Feng Y, Lai K-h, Zhu Q. Legitimacy in operations: How sustainability certification announcements by Chinese listed enterprises influence their market value? International Journal of Production Economics 2020;224.
[25] Jonek-Kowalska I. How do turbulent sectoral conditions sector influence the value of coal mining enterprises? Perspectives from the Central-Eastern Europe coal mining industry. Resources Policy 2018;55:103-12.
[26] Zhang F, Fang H, Wang X. Impact of Carbon Prices on Corporate Value: The Case of China's Thermal Listed Enterprises. Sustainability 2018;10(9).
[27] Frenda S, Teresa Cignarella A, Basile V, Bosco C, Patti V, Rosso P. The unbearable hurtfulness of sarcasm. Expert Systems with Applications 2022;193.
[28] Hong Kong Monetary Authority. 2023:www.hkma.gov.hk.
[29] Mikolov T, Chen K, Corrado G, Dean J. Efficient Estimation of Word Representations in Vector Space. Computer Science 2013.
[30] Mikolov T, Sutskever I, Kai C, Corrado G, Dean J. Distributed Representations of Words and Phrases and their Compositionality. *arXiv.* 2013.
[31] Baroni M, Dinu G, Kruszewski G. Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. *52nd Annual Meeting of the Association-for-Computational-Linguistics (ACL).* Baltimore, MD; 2014:238-47.
[32] Safavian SR, Landgrebe D. A SURVEY OF DECISION TREE CLASSIFIER METHODOLOGY. Ieee Transactions on Systems Man and Cybernetics 1991;21(3):660-74.
[33] Breiman L. Random forests. Machine Learning 2001;45(1):5-32.
[34] Denoeux T. A K-NEAREST NEIGHBOR CLASSIFICATION RULE-BASED ON DEMPSTER-SHAFER THEORY. Ieee Transactions on Systems Man and Cybernetics 1995;25(5):804-13.
[35] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation 1997;9(8):1735-80.
[36] Cui Y, Che W, Liu T, Qin B, Yang Z. Pre-Training With Whole Word Masking for Chinese BERT. Ieee-Acm Transactions on Audio Speech and Language Processing 2021;29:3504-14.
[37] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention Is All You Need. *31st Annual Conference on Neural Information Processing Systems (NIPS).* 30. Long Beach, CA; 2017.
[38] Devlin J, Chang M-W, Lee K, Toutanova K, Assoc Computat L. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Conference of the North-American-Chapter of the Association-for-Computational-Linguistics - Human Language Technologies (NAACL-HLT).* Minneapolis, MN; 2019:4171-86.
[39] Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al. RoBERTa: A Robustly Optimized BERT Pretraining Approach. Information Systems Research 2019.