

# Research on the Prediction of Public Opinion Reversal Based on Machine Learning in New Media Context

Shudi Lou\*

School of Management, Harbin Institute of Technology, Harbin, China.

2022111145@stu.hit.edu.cn

**Abstract.** With the Internet development in the new media era, frequent public opinion reversal events have been a challenge that cyberspace governance should tackle urgently. Predicting the public opinion reversal events in time can help relevant departments grasp the trend of public opinion and formulate effective coping strategies. Based on the typical public opinion events on Weibo from 2022 to 2024, this paper extracts multi-dimensional variables such as communication characteristics, agenda setting degree, and emotional tendency. Also, two text vectorization methods are adopted for calculation: TF-IDF and Doc2Vec, and machine learning methods are introduced for modeling prediction and evaluation. According to the empirical results, Doc2Vec performs better than TF-IDF, and Doc2Vec's deep learning model performs best, which can improve the recognition accuracy of public opinion reversal events. The research offers technical insights into constructing a prediction model of public opinion reversal, and supports the development of public opinion prediction.

**Keywords:** Online Public Opinion, Public Opinion Reversal, Machine Learning, Deep Learning, Agenda Setting, Emotional Analysis.

## 1. Introduction

In the era of the Internet and new media, the speed of information dissemination is rising rapidly with the increasing penetration rate of the Internet. According to the *54th Statistical Report on China's Internet Development* released by CNNIC [1], as of June 2024, the number of Internet users in China is nearly 1.1 billion, an increase of 7.42 million compared with December 2023, and the Internet penetration rate reaches 78.0%. Meanwhile, the threshold for information dissemination continues to decrease, and social media platforms such as Weibo, Zhihu, and Douban have become vital places for public opinion to germinate and ferment. Taking Weibo as an example, the *2024 Weibo User Consumption Trend Report* [2] released by Sina Weibo in 2024 indicates that as of September 2024, Weibo has 587 million monthly active users and 257 million daily active users, covering a huge user group composed of multiple subjects such as media celebrities, brand merchants, and netizens. It is a critical frontier for the generation and dissemination of public opinion.

The dissemination of public opinion in the new media era depends more on the creative output, emotional attitude and behavior patterns of the majority of netizens. Luo Yuting [3] et al. pointed out that the public opinion reversal is essentially the result of the changing public value judgment and emotional attitude. Besides, the interaction of views with different orientations promotes the public opinion evolution in different directions. Due to netizens' non-professionalism, the public opinion is often mixed with superficial, one-sided and highly emotional information. However, some mainstream media and Internet celebrities often deliberately amplify these immature views to gain the popularity, which leads to extreme public opinions. With the passage of time, the information related to the incident has been improved, and the emotions of netizens tend to be rational. After the truth of the incident has been restored, public opinion has developed in the opposite direction, resulting in public opinion reversal. Frequent reversal events will weaken the media credibility, bring strong anxiety and panic to the public, and cause serious secondary harm to the victims. Some bad events even extend offline and turn into vicious social events. This destroys the network ecological environment and hinders the construction of a harmonious society.

Selecting Weibo to collect data, this paper uses qualitative and quantitative research to extract event features. Apart from constructing a mathematical model, it predicts the possibility of public opinion reversal with the help of machine learning algorithms and builds an evaluation system. Thus,

a new model and data support is provided for predicting public opinion reversal, which can better predict the trend of public opinion, formulate and implement effective intervention strategies in time, and reduce adverse consequences caused by the public opinion reversal.

## 2. Research Status

Presently, the research on public opinion reversal mainly focuses on its causes, evolution mechanism and coping strategies. Based on previous research, this paper holds that public opinion reversal is a crucial change in the mainstream public opinion attitude or judgment in a certain node during the spread of public opinion events.

### 2.1 Causes of Public Opinion Reversal

As an increasingly common social communication phenomenon in the new media era, the causes of public opinion reversal are complex and multi-dimensional. By integrating existing research results, this paper summarizes the main causes of public opinion reversal into the following three aspects:

#### (1) Ambiguity of Media Information Expression

In the early stage of spreading the incident, due to incomplete information and unreliable sources, it is often difficult for the public to fully understand the truth. Jiang Jiabin<sup>[4]</sup> believed that the media would deliberately blur some information and mislead public sentiment in initial news reports. Against the information asymmetry, the media affects the communication path and the orientation of public cognition by controlling the topics, the frequency of information source release and the expression tone in the early stage, which is a crucial factor causing the public opinion reversal.

#### (2) Fragmentation Trend of Information Dissemination

In the new media era, the information dissemination of social platforms features typical fragmentation, and the public tends to obtain information from short and straightforward expressions such as short videos and screenshots. According to Huang Hongye<sup>[5]</sup>, there are fewer pieces of news in the new media era, with the audience forced to accept much fragmented information in this process. Nowadays, the breadth of information for the audience is far greater than the depth, and the dissemination of information fragmentation will mislead them and aggravate the public opposition.

#### (3) Diversified Trend of Public Opinion Participants' Structure

The spiral theory of silence put forward by Neumann<sup>[6]</sup> holds that when individuals perceive the mainstream of public opinion, they tend to follow the crowd out of fear of isolation, which triggers the gradual silence of minority opinions and the effect of an "opinion echo". Li Yong<sup>[7]</sup> et al. believed that the essence of public opinion reversal is an "anti-silence spiral", which depends on: the heterogeneity of backbone molecules, the threshold effect of information stimulation and the nonlinearity of time dimension. In the new media era, more users participate in the public opinion development, and the collision of multiple views will prompt the originally silent dissidents to speak out and promote the opposite public opinions.

### 2.2 Evolution Mechanism of Public Opinion Reversal

The public opinion reversal is a changing evolution, which has apparent stages and dynamics. Based on the existing research, public opinion reversal events often have a clear life cycle division, and most of their turning points are accompanied by the impact of new supplementary information or the exposure of authoritative evidence. This key information greatly reverses the public cognitive foundation and mainstream emotional tendencies, leading to a sharp change of the public opinion.

He Jiejun [8] et al. analyzed the characteristics of emotional evolution from the spatiotemporal dimensions, divided the public opinion reversal into six life cycle development stages, and mentioned their differences in emotional evolution in time and space. Zhang Xinyu [9] divided public opinion reversal into three configuration types through fsQCA, and further clarified the constituent factors and transmission paths of various types of reversal events. From a macro aspect, the public opinion reversal is a dynamic evolution driven by the interaction of information, media and public emotion,

which can be explained by the social burning theory. Social burning theory is an analogy model of social physics proposed by Niu Wenyuan [10] based on the burning phenomenon in nature, which emphasizes the burning-supporting role of social conflict, media misleading and emotional stimulation in the evolution of public opinion. Based on the social combustion theory, this paper regards controversial events or public opinion events as combustible materials, media intervention methods and reporting directions as ignition sources, and public emotions as motivators for public opinion warming and public opinion fermentation.

### 2.3 Coping Strategies for Public Opinion Reversal

Public opinion reversal event results from many factors, and its coping strategy should start from the perspective of multi-agent coordination. As the key nodes in the chain of public opinion development, communicators and audiences affect the trend of public opinion, which are the key objects when formulating governance strategies.

When it comes to communicators, it is necessary to strengthen the media's sense of responsibility, standardize reporting behavior and enhance the media's credibility. Jiang Yesha [11] et al. proposed to strengthen the responsibility of online media, build industry regulatory rules, and enhance government's credibility and social governance ability. Media reports should follow the professional ethics of journalism, ensure the balance of information, and avoid one-sided information dissemination for the sake of blog traffic. In terms of the audience, public media literacy is directly related to the communication path and evolution direction of public opinion. An Lu [12] et al. held that the public should improve recognition ability, not easily believe vague evidence, and reduce the negative effects during the public opinion dissemination. Furthermore, the prevention and control of the public opinion reversal should proceed from the overall situation and build an effective collaborative governance model. Bi Hongyin [13] et al. proposed to build a governance system of multi-party cooperation from aspects of grass-roots government, media personnel, netizens and new network technologies. Based on the symbiosis theory, Zhang Liu [14] et al. believed that netizens, media and the government should jointly govern public opinion reversal.

### 2.4 Prediction of Public Opinion Reversal

The prediction of public opinion reversal aims to classify public opinion reversal events. With the maturity of machine learning, machine learning algorithms are increasingly used in the prediction of public opinion reversal, and the research methods manifest the trend of multi-dimensional and multi-model fusion. Wang Nan [15] et al. built a deep learning method based on the Doc2EV model and hierarchical mask attention to generate event feature vectors for prediction. In addition to identifying influencing factors of We Media's public opinion reversal from aspects of platform control, information accuracy, subject criticism and communication mutation, Tian Shihai [16] et al. constructed a Bayesian network model for prediction. From the perspective of theme evolution, Zhang Jiyang [17] et al. used a gated recurrent neural network to tackle the theme evolution and constructed a prediction model.

Overall, the existing research on the prediction of public opinion reversal focuses on cause exploration, evolutionary mechanism analysis and governance strategy formulation, with few on prediction using multiple integrated machine learning models. Moreover, most indicators rely on manual assignment, which consumes time and energy with difficulty in guaranteeing the accuracy. Model comparison is mainly limited to the result comparison of machine learning models, with little attention paid to the impact on model results when different analysis methods are used in other parts of the model. Based on the research gaps and theoretical basis, this paper combines objective feature assignment with subjective feature assignment to construct a feature system with higher accuracy. During the research, the results of different machine learning algorithms and the text analysis methods are compared and evaluated to provide insights into the research on the prediction of public opinion reversal.

### 3. Study Design

#### 3.1 Research Methods

This study adopts a research combining qualitative and quantitative, with the specific process as follows: (1) Use web crawler to obtain data from Weibo. (2) Preprocess the text data such as word segmentation and stop words. (3) Use manual assignment, descriptive statistics, text mining, etc. to process feature attributes. (4) Use numerous machine learning algorithms to build models for prediction. (5) Evaluate the model results and construct a reasonable feature evaluation system.

#### 3.2 Selection of Public Opinion Events

Xinhuanet, People’s Daily Online, Phoenix.com, The Paper and other official news websites will select the annual reversal news. In addition, www.zhiweidata.com can monitor public opinions, which provides the entry names of hot events, and plays a crucial role in the event screening. Based on the website information and the event selection criteria in research literature, this paper follows the three selection criteria:

Consistency of event scale and popularity: Ensure the selected reversal and non-reversal events have similar popularity, discussion scale and public opinion cycle.

(2) Event controversy: The public opinion events should have discussion room.

(3) Adequacy of text data: The selected public opinion events should have a time duration and sufficient heat to ensure data volume and representativeness.

This paper selects 28 public opinion events from 2022 to 2024, including 14 reversal events and 14 non-reversal events, with the specific events shown in Table 1.

Table 1 28 Public Opinion Events from 2022 to 2024

No.	Reversal Event	No.	Reversal Event	No.	Non-reversal Event	No.	Non-reversal Event
1	Jiang Ping Incident	8	CLP Overtime Incident	1	Several Men Beat Women in Tangshan Barbecue Restaurant	8	Shanghai Four-year-old Girl Lost on the Beach
2	Zhengzhou Sky-high Price Dog Hunting Incident	9	Milk Tea Shop Entrapment Incident	2	Zhong Xuegao Left at Room Temperature at 31 Degrees for One Hour Without Melting	9	BMW Treats Chinese and Foreign Tourists Differently at Shanghai Auto Show
3	Wuxi Three-year-old Girl Was Sexually Assaulted	10	Cambodian Blood Slave Case	3	Audi Xiaoman Advertisement Exposed to Plagiarism	10	Mixed Use of Oil Tanks Transported by Grain Storage Tanker
4	“Thuman” Cat Pose Event	11	Harbin Man Burn Incident	4	CNKI Was Fined 87.6 Million for Suspected Monopoly	11	Mercedes-Benz Owner Jumps in Line and Punches Hood
5	Woman Died in Rental House	12	“Fat Cat” Incident	5	Essay of East Buy Holding Limited	12	Huang Lei’s Personality Was Questioned
6	Shanxi “Yin-Yang Daoxiao Noodles”	13	Chengdu No.49 Middle	6	Two-Year-Old Girl Bitten by Vicious Dog in Chengdu	13	Hangzhou Bentley Owner Stabbed BMW

7	Tianjin University Grant Incident	14	School Incident "Reader Girl" Exposes Herself to Brain Tumor	7	Hospitals Publicly Sell Birth Certificates and Babies	14	Owner on the Elevated Road Shanghai Female Teacher Cheating High School Students
---	-----------------------------------	----	--	---	---	----	--

### 3.3 Model Feature Construction

In this paper, the feature assignment of public opinion events is divided into two categories. One is objective feature assignment, and the other is subjective feature assignment. Objective features are information from web pages or through calculation, while subjective features greatly affect public opinion during the public opinion evolution, but cannot be obtained from web pages and should be labeled artificially.

#### 3.3.1 Objective Feature Assignment

Objective feature assignment mainly includes communication features, agenda setting and emotional tendency. Communication features reflect the popularity and communication of public opinion events in cyberspace. Based on Jiang Changbin [18] et al.'s research on the division of public opinion popularity indicators, this paper divides communication features into the number of forward, likes and comments, which respectively correspond to the secondary communication ability of information, the depth of public participation in discussion and netizens' recognition of content.

Agenda setting can reflect the significance of public opinion events in the information environment, which is a vital criterion to measure the discussion and topic dominance of an event on the social media platform. Lai Jidan [19] et al. pointed out that the formation of We Media's agenda setting needs to be echoed by netizens, which has higher freedom and extension than traditional media. In the public opinion reversal, the trend of public opinion is often accompanied by the changing public discussion focus, and the agenda setting degree can quantify the intensity and direction of this evolution. On this basis, this paper uses the text similarity between the blog post and the comment text to express the agenda setting degree, which is calculated by TF-IDF and Doc2Vec respectively.

(1) TF-IDF means the document by calculating  $TF \times IDF$  to get the weight of each word. Its core principle is the more times a word appears in the text, the more vital it is to the text. But if it appears in many documents, its discriminative ability will be weaker. TF measures the occurrence frequency of word  $t$  in document  $d$ :

$$TF(w, d) = \frac{f_{w,d}}{\sum_{w' \in d} f_{w',d}} \quad (1)$$

Among them,  $f_{w,d}$  is the number of occurrences of word  $w$  in document  $d$ , and  $\sum_{w' \in d} f_{w',d}$  is the summed occurrences of all words in document. The IDF measures the importance of a word  $w$  in the entire document set  $D$ :

$$IDF(w, D) = \ln \frac{N}{n_w + 1} \quad (2)$$

Among them,  $N$  is the total number of document sets,  $n_w$  is the number of documents containing the word  $w$ , and plus 1 for smoothing processing. TF-IDF combines word frequency with inverse document frequency:

$$TF - IDF(w, d, D) = TF(w, d) \times IDF(w, D) \quad (3)$$

The final weight of word  $w$  in document  $d$  is calculated by Formula 3.

(2) Doc2Vec is a Word2Vec-based document vectorization model proposed by Mikolov T [20] et al., which can map variable-length text into dense vectors of fixed length. Doc2Vec has two main model architectures, namely PV-DM similar to CBOW in Word2Vec, and PV-DBOW similar to Skip-gram in Word2Vec. In PV-DM, each sentence and word are mapped into a vector space, and the columns of the matrix are used to represent them respectively. Then, the sentence vector and word are linked to order or averaged to obtain features to predict the next word in the sentence, which is mathematically expressed as Formula (4)

$$P = (w_t | d, w_{t-k}, \dots w_{t-k}) = \text{Softmax}(U \cdot h + b) \tag{4}$$

$$h = \text{concat}(d, w_{t-k}, \dots w_{t-k})$$

Where  $w_t$  is the current word,  $d$  is the document vector,  $w_{t-k}, \dots w_{t-k}$  is the word in the whole context window,  $U$  is the word vector matrix, and  $b$  is the bias term.

In PV-DBOW architecture, instead of adopting PV-DM to model the display of word order, the document is regarded as a whole, the input context is ignored, and the random sampled words in the document are directly predicted through the document vector, so as to learn the global semantic representation of the document. Its mathematical expression is as follows:

$$P = (w_t | d) = \text{Softmax}(U \cdot d + b) \tag{5}$$

Where  $w_t$  is the randomly sampled word in the predicted document, and  $d$  is the document vector.

Netizens' emotion is the promoter of the development of public opinion. Using sentiment analysis in NLP to obtain emotion scores and analyzing netizens' emotional tendencies in blog posts and comments is important to predict public opinion reversal. The SnowNLP library in Python is a probabilistic model based on supervised learning, which is more used for Chinese text processing. Based on the naive Bayes algorithm, it calculates conditional probability through word frequency statistics and performs sentiment analysis on text data, with its emotional score based on Formula (6).

$$P = (c_1 | w_1, \dots w_n) = \frac{P(w_1, \dots w_n | c_1) \cdot P(c_1)}{P(w_1, \dots w_n)} \tag{6}$$

Where  $w_1, \dots w_n$  is the input feature and  $c_1$  is the category to which it belongs.

### 3.3.2 Subjective Feature Assignment

Subjective features should be labeled artificially. Compared with user features, event features are more important in prediction. Thus, some special event types will leave stereotypes in the public and affect the initial judgment of public opinion. Tian Junjing [21] et al. divided the features of events into eight types: disaster accidents, doctor-patient relationships, cultural education, etc. Since there is no unified definition of event types in academia, this paper divides event types according to the basic types of social events, including natural disasters, man-made accidents, public health, social ethics, law enforcement events, cultural education, etc. The social combustion-supporting theory holds that the accident instability and the contradiction of interests among the parties involved in the accident constitute a "combustion substance" and "combustion-supporting agent". On this basis, this paper divides the features of the event into event types, the identity of the exposor, the subject class of the event and the event outcome. According to Liao Mengxia's [22] division of people in public opinion reversal events, the events are divided into vulnerable groups, occupational groups, special groups and people that are not specifically targeted at certain groups. Based on An Lu [12] et al.'s division of incident results, this paper divides incident results into property loss, personal safety and health loss, reputation loss and rights loss. Given the identity authentication of different user groups in a public opinion event by Weibo, the identity of the event exposor is divided into parties, We Media, ordinary people, mainstream media and government departments. Overall, the characteristic assignment of public opinion reversal events is in Table 2.

Table 2 Feature Assignment of Prediction Model of Public Opinion Reversal Event

Primary Feature	Secondary Feature	Feature Value
Event Characteristics	Event Type (X1)	Natural Disasters/Man-made Accidents/Public Health/Social Ethics/Law Enforcement Incidents/Cultural Education
	Identity of Event Exposor (X2)	Parties/We Media/Ordinary People/Mainstream Media/ Government Departments
	Event-oriented Subject Hierarchy (X3)	Vulnerable Groups/Occupational Groups/Special Groups/Not Specifically Targeted
	Event Results (X4)	Damage to Property/Loss of Personal Safety and Health/Loss of Reputation/Loss of Rights
User Characteristics	User Authentication Identity (X5)	Top-class VIP/Second-class VIP/ Third-class VIP/Fourth-class VIP/Ordinary User

Propagation Characteristics	Number of Forwards (X7)	$[0, +\infty)$
	Number of Comments (X8)	$[0, +\infty)$
	Number of Likes (X9)	$[0, +\infty)$
Agenda Setting	Blog Agenda Setting Degree (X10)	$[0,1]$
	Comment Agenda Setting Degree (X11)	$[0,1]$
Emotional Tendency	Blog Post Content Emotion Score (X12)	$[0,1]$
	Review Content Emotion Score (X13)	$[0,1]$

### 3.4 Construction of Prediction Model of Public Opinion Reversal

The prediction model of public opinion reversal is a binary model, outputting “with reversal” and “without reversal” as the prediction results. This paper compares machine learning methods used by scholars, and selects five methods: random forest, XGBoost, Bayesian classification, logistic regression, and multi-layer perceptron, so as to build a model for public opinion reversal prediction.

Given the characteristic variables of the model built in 3.3 and its features, the features of the input data should have the above 13 dimensions, and the input feature matrix  $X = [x_1, x_2, x_3, \dots, x_{13}]$ . When using Doc2Vec to calculate the agenda setting degree, an additional 100-dimensional dense vector will be generated. This part of the feature can help us understand the semantic information of the text more completely. This paper incorporates it into the input feature matrix of the model. Besides, the data set in the final modeling is a structured supervised learning sample set, each row represents an event sample, and the input feature matrix processed by TF-IDF contains 13 input features and 1 output variable, expressed as  $(x_1, x_2, x_3, \dots, x_{13}) \rightarrow Y$ . The input feature matrix processed by Doc2Vec has 113 input features and 1 output variable, expressed as  $(x_1, x_2, x_3, \dots, x_{113}) \rightarrow Y$ .  $Y$  is the output feature and  $Y \in \{0,1\}$ , where 0 means no inversion and 1 means inversion has occurred.

Since the data set in the final modeling must be a structured supervised learning sample set, the non-numerical data should be encoded as numerical, and the `get_dummies()` function of Pandas library in Python should be used to perform One-hot encoding on event features. Moreover, the category variables are converted into unordered and equidistant numerical vectors, which prevents the model from misunderstanding the category relationship due to the encoding size. The coding result returns a sparse matrix, with categories denoted as 1 and the absence of categories denoted as 0. After coding processing comes the final input matrix.

The accuracy rate, precision rate, recall rate, F1 index and AUC value are used as the evaluation indexes of the model, with the accuracy rate as the proportion of samples predicted correctly by the model to the total samples. The recall rate is the proportion of samples that are positive examples correctly predicted by the model. The F1 index is the harmonic average of the accuracy rate and recall rate, which synthesize their performance. The AUC is the area under the ROC curve, which reflects the ability of the model to distinguish between positive and negative examples. It is believed that when the AUC is more than 0.8, the model performs well.

## 4. Empirical Research

### 4.1 Experimental Data Acquisition

This paper uses Python crawlers to obtain comment data and blog post data. For comment data, requests and csv modules in Python are used to simulate browser access behavior by specifying request header information to realize batch capture and structured storage of related Weibo comment

data. Blog post data is based on the weibo-search project of the author of dataabc on GitHub for code writing. The comment text data is added to the data table as an emotion feature attribute of the event after emotional analysis through the SnowNLP library. After preliminary cleaning, a total of 60,215 pieces of uninverted data and 56,560 pieces of inverted data were obtained, totaling 116,775 pieces of data.

## 4.2 Construction and Comparative Analysis of Prediction Model of Public Opinion Reversal

### 4.2.1 Model Selection

#### (1) Random Forest

Random forest is an ensemble learning based on a decision tree, which is a crucial representative of the Bagging algorithm. The output class of the classification problem is determined by the mode of each tree output class.

#### (2) XGBoost (eXtreme Gradient Boosting)

XGBoost is a Boosting algorithm based on gradient lifting. It introduces second-order Taylor expansion when optimizing the loss function, performs pre-pruning when building the tree, and adds regular terms to prevent overfitting.

#### (3) Naive Bayes Classifier

Based on the independent assumption of Bayes theorem and characteristic conditions, Naive Bayes calculates the posterior probability of samples belonging to each category, and selects the category with the largest posterior probability as the prediction result.

#### (4) Logistic Regression

As a statistical method widely used in binary classification, logistic regression is a classification model that maps linear combinations to  $[0, 1]$  intervals based on logarithmic probability (logit) functions for classification.

#### (5) Multilayer Perceptron (MLP)

The multi-layer perceptron is a forward-structured artificial neural network, which is a deep learning model. A high-order feature abstraction of the input is performed by a nonlinear activation function, mapping a set of input vectors to a set of output vectors.

### 4.2.2 Comparative Analysis of Results

Based on the feature assignment in Table 2, this paper constructs agenda setting indexes based on TF-IDF and Doc2Vec respectively, and uses them as model inputs with other variables. The above five classification models are combined for training and testing, with the results shown in Table 3.

Table 3 The Results of Five Classification Models

Models	Accuracy	Precision	Recall	F1-score	AUC
LR-TFIDF	0.9327	0.9296	0.9302	0.9299	0.9832
LR-Doc2Vec	0.9863	0.9832	0.9892	0.9858	0.9985
NB-TFIDF	0.6793	0.5997	0.9992	0.7495	0.9301
NB-Doc2Vec	0.7762	0.6842	0.9915	0.8097	0.9784
RF-TFIDF	0.9326	0.9362	0.9225	0.9293	0.9574
RF-Doc2Vec	0.9551	0.9592	0.9467	0.9529	0.9930
XG-TFIDF	0.9218	0.9094	0.9304	0.9198	0.9598
XG-Doc2Vec	0.9685	0.9681	0.9666	0.9673	0.9962
MLP-TFIDF	0.9001	0.9467	0.8392	0.8897	0.9702
MLP-Doc2Vec	0.9954	0.9914	0.9991	0.9952	0.9999

According to the experimental results, the Doc2Vec vectorization is better than TF-IDF in all models, and performs best in the deep learning model MLP. The evaluation indicators all reach above 0.99, and the AUC (0.9999) is close to the full score, which is the best among all model combinations. Compared with TF-IDF, Doc2Vec can better capture contextual semantic information. Among all models, random forest and XGBoost perform well in this task. After using Doc2Vec to extract semantic information, the AUC of the model is significantly improved. After the logistic regression combined with Doc2Vec, the F1-score improves from 0.92 to 0.98, and the performance improves significantly, indicating that even linear models can benefit from more expressive text vectors. Among all models, Naive Bayes performs the worst. The F1-score under the TF-IDF condition is only 0.7495. The accuracy and precision under both methods are lower than 0.8 with misjudgments in the model.

As a traditional text representation method, TF-IDF has good adaptability to linear models such as logistic regression. However, due to its sparsity and loss of contextual semantics during the training, TF-IDF is not effective in deep models and complex classifiers with declining performance. Doc2Vec can retain semantic structural features by embedding the entire text into a dense vector space, and is more suitable for modeling combined with deep learning and complex nonlinear models. Combined with the model training results, Doc2Vec+ deep learning or a complex classifier can be used to model the public opinion reversal prediction.

## 5. Conclusions

Using deep learning and machine learning to build a prediction model of public opinion reversal is a vital technical development for public opinion monitoring and prediction in the future. Focusing on public opinion reversal, this paper studies public opinion events in recent years, adds text mining to construct a feature system for the prediction of public opinion reversal, and combines machine learning and deep learning models for modeling. Empirical research shows that Doc2Vec-based semantic vectors can better capture contextual semantic features, and the prediction effect is better than that of traditional TF-IDF feature vectors. Deep learning and complex classifiers are more effective in model construction with higher model accuracy. In future research, multi-modal data such as images and videos can be further integrated to improve model perception capabilities. For public opinion events with multiple reversals, time series analysis and reinforcement modeling are introduced to construct a reversal recognition model of sustainable learning.

## References

- [1] The Release of the 54th “Statistical Report on China’s Internet Development”. (2024). Media Forum, 7(17):121.
- [2] 2024 Weibo User Consumption Trend Report. Retrieved from <https://data.weibo.com/report/reportDetail?id=473>
- [3] Luo, Y. T. & Xuan, X. (2024). On the reversal of online public opinion and its guidance. Thinking, 50(04):18-25.
- [4] Jiang, J. X. (2018). Research on the causes and regulation of reverse news from the perspective of news media—Taking the pregnant woman falling from a building in Yulin as an example. Journal of News Research, 9(13):113.
- [5] Huang, H. Y. (2015). The causes and regulation of “public opinion reversal”. News Youth Journalist, (09):22-23.
- [6] Neumann, E.N. (2013). The Spiral of Silence: Public Opinion-Our Social Skin. Trans: Dong Lu. Beijing: Peking University Press.
- [7] Li, Y. (2022). “Anti-silence spiral” in the reversal of public opinion—Analysis based on the perspective of audience psychology. Youth Journalist, (05):50-52.

- [8] He, J. J. & Li, Y. (2022). Research on the emotional evolution characteristics of public opinion reversal events based on time and space perspective. *Journal of Information Resources Management*, 12(02):88-100.
- [9] Zhang, X. Y. (2024). Research on the generation mechanism of network public opinion inversion under the information ecology theory. China Management Modernization Research Association. Proceedings of the 19th (2024) China Management Annual Conference-Public Administration Forum. Wuhan University of Technology, School of Safety Science and Emergency Management, 2024: 2-16.
- [10] Niu, W. Y. (2001). The social physics and the warning system of China's social stability. *Bulletin of Chinese Academy of Sciences*, (01):15-20.
- [11] Jiang, Y. S. (2019). How is the truth close to the truth in the post-truth era—Based on the public opinion analysis of food safety incidents in Chengdu No.7 Experimental Middle School. *Southeast Communication*, (10):91-93.
- [12] An, L. & Hui, Q. Y. (2022). Prediction of microblogging public opinion reversal in the context of hot events. *Journal of Information Resources Management*, 12(03):21-34.
- [13] Bi, H. Y. & Tian, H. (2019). Analysis and reflection on the phenomenon of “quasi-reversal” in public opinion—Taking the “Wanzhou bus falling into the river incident” as an example. *Journal of Intelligence*, 38(07):103-110.
- [14] Zhang, L., Yi, S. Y., Wang, H. et al. (2024). The formation path of network public opinion reversal from the perspective of symbiosis theory—Based on qualitative comparative analysis of fsQCA. *Information Science*, 42(06):74-82.
- [15] Wang, N., Du, H., Tan, S. R. et al. (2025). Feature extraction and reversal prediction of public opinion events based on deep learning. *Journal of Intelligence*, 44(03):107-118.
- [16] Tian, S. H., Sun, M. Q. & Zhang, J. Y. (2019). Prediction of We-media public opinion reversion based on Bayesian network. *Information Studies: Theory & Application*, 42(02):127-133.
- [17] Zhang, J. Y., Zhang, P., Lan, Y. X. et al. (2023). Construction and empirical study of online public opinion inversion identification model based on dynamic topic clustering. *Information Studies: Theory & Application*, 46(10):174-181+129.
- [18] Jiang, C. B., Zou, Y. Q., Wang, H. et al. (2021). Research on We-media public opinion reversal prediction based on SVM. *Information Science*, 39(04):47-53+61.
- [19] Lai, J. D. & Li, D. (2020). The shortcomings of the agenda setting of the We media in the reversal news and their countermeasures. *Journal of Hainan University (Humanities & Social Sciences)*, 38(02):135-142.
- [20] Le Q, Mikolov T. Distributed representations of sentences and documents[C]// International conference on machine learning. PMLR, 2014: 1188-1196.
- [21] Tian, J. J., Lan, Y. X., Xia, Y. X. et al. (2019). Recognition and empirical study of network public opinion reversal based on decision tree method. *Information of Intelligence*, 38(08):121-125+171.
- [22] Liao, M. X. (2020). Conflict narrative, corrective intervention and public opinion reversal results—A qualitative comparative analysis of 43 media events with fuzzy sets. *Modern Communication (Journal of Communication University of China)*, 42 (05):51-56.