International Journal of Web Based Communities

ISSN online: 1741-8216 - ISSN print: 1477-8394
https://www.inderscience.com/ijwbc

Feature extraction of news communication on Microblog platform based on multilevel sliding window model

Cheng Cui

DOI: 10.1504/IJWBC.2024.10061796

Article History:

Received:	17 May 2023
Last revised:	21 June 2023
Accepted:	10 October 2023
Published online:	04 November 2024

Feature extraction of news communication on Microblog platform based on multilevel sliding window model

Cheng Cui

Finance and Asset Management Office
of Shijiazhuang University of Applied Technology,
Shijiazhuang, 050081, China
Email: jiahewanshixing28@126.com

Abstract: In order to fully understand the characteristics of news dissemination on Microblog platforms, this article proposes a method for extracting news dissemination features on Microblog platforms based on a multi-level sliding window model. Firstly, identify the four major characteristics of Microblog platform news, including high timeliness, free writing style, rich personal emotional bias, and the ability to restore the truth of the news. Secondly, a Microblog platform news communication representation model is constructed using the news communication theme content as input and the Microblog platform news communication representation vector as output. Finally, determine the multi-level sliding window counter, correspond to the sub window positions, and segment the feature data to complete the feature extraction of news dissemination on Microblog platform. The experimental results show that the proposed method has a high recall rate for feature extraction and good feature data balance.

Keywords: multilevel sliding window model; Microblog platform; news dissemination; feature extraction.

Reference to this paper should be made as follows: Cui, C. (2024) 'Feature extraction of news communication on Microblog platform based on multilevel sliding window model', *Int. J. Web Based Communities*, Vol. 20, Nos. 3/4, pp.212–227.

Biographical notes: Cheng Cui received his Master's in Software Engineering from the Shandong University in 2012. He is currently a Lecturer at the Shijiazhuang University of Applied Technology. Her research interests include computer applications and software engineering.

1 Introduction

The rapid development of the electronic information age has become a driving force for the growth of multimedia social software. People have achieved the convenience of quickly accessing the internet, obtaining external news and news through social media software (Baek et al., 2021). As one of the most influential news media software, Microblog has also become an important platform for information exchange and news acquisition due to its fast, interactive, and wide source of information. The emergence of

Microblog platform has brought a new transformation to the way news is disseminated (Reynosa-Guerrero et al., 2021). The mode of news dissemination has changed from passive to active, making the news audience the main individual who obtains news. Microblog users can selectively select news of their interest on this platform, and can independently edit news content, making their desire to participate in news releases even stronger (Zhang et al., 2021). It can be said that the emergence of Microblog has become a new way of news dissemination, making the audience more willing to participate in news acquisition and creation. Microblog, as a new force in new media, has become a crucial means of news dissemination and provides new vitality for the dissemination and development of mass news (Zheng et al., 2021). But with the continuous emergence of social media software such as Microblog platform, the characteristics of news dissemination on Microblog platform have also undergone certain changes. The characteristics of news dissemination on Microblog platform refer to the way news is disseminated on Microblog platform, in order to gain more news audiences. But with the continuous changes in the characteristics of news dissemination on Microblog platform, the news power of the platform has also undergone certain changes (Sun, 2021a). If you want to always stay at the forefront of news communication, the extraction of news communication features on Microblog platform is essential. For this reason, media professionals have conducted a lot of research on feature extraction methods for news dissemination on Microblog platform and designed many feature extraction methods.

Tajanpure and Muddana (2021) designed a classification feature extraction algorithm based on cyclic convolution for high-dimensional news dissemination datasets. The research method points out that high-dimensional data analysis of news has become the most challenging task today. Dimension reduction plays an important role in this regard. Therefore, a dimensionality reduction method based on input feature convolution is proposed. This method mainly uses decision trees, support vector machines, and K-nearest neighbour classifiers to extract news dissemination features. The most significant feature of the proposed model is that it reduces attributes, thereby reducing computational time without compromising the accuracy of feature extraction. However, this method ignores other features of news data in feature extraction, and only performing dimensionality reduction processing cannot meet the feature extraction requirements of all data. Li and Liu (2021) designed a news dissemination feature extraction algorithm based on deep learning. This method designs shared convolutional layers and RPN networks in the detection network, and uses VGGNet instead of ZFNet to improve the depth of shared convolutional layers. A classification method combining semantic enhancement was designed for fine-grained topic news classification. Combining the idea of modular fusion mechanism, a semantic enhanced multiple fusion is proposed. The Bert module replaces the traditional Word2Vec for semantic vector representation and introduces Bi LSTM to adaptively extract news context features. Through self attention mechanism, feature expression is strengthened and network weights are adjusted, ultimately enabling the model to achieve accurate classification. A feature extraction method based on deep learning was proposed. The algorithm framework includes particle filtering, pre trained features, convolutional neural networks, discriminant classifiers, and online parameter updates. This method combines deep learning models with traditional object extraction methods and framework depth. This method has relatively outstanding performance and can adapt to many interferences encountered during the extraction

process and changes in the target itself, with strong robustness. (Yin and Chen, 2022) studied a cross modal news communication text feature retrieval and extraction method based on efficient feature extraction and interactive learning. This method proposes a framework to estimate the displacement of news features from reference news to another news text. This is a random search process in a different metric space. This method is achieved by mapping the variable domain to be estimated into the symbol space using a set of operators in the discrete space, where the random search method is described through a unified sampling process and genetic operators. This method searches for the best suboptimal solution of locality under predefined metrics, avoiding divergence of the worst suboptimal solution. This proposal is based on the formalisation of Lucas and Kannard methods. In the study, the suggestions of Shi and Tomasi methods were considered as spatial solutions for news dissemination feature extraction, and experimental feature extraction was studied. This method has a wide research scope, but there is a problem of incomplete analysis of the news dissemination process in feature extraction.

In order to improve the effectiveness of feature extraction for news dissemination on Microblog platform, this paper designs a feature extraction of news communication on Microblog platform based on multilevel sliding window model. The main technical route studied in this article is:

1 Determine the four characteristics of Microblog platform news, which are respectively: timeliness characteristics of Microblog platform news, characteristics of news content writing, characteristics of news content emotional bias, and characteristics of fragmentation model to restore the truth of news. On this basis, summarise the language characteristics of Microblog platform news, and complete a comprehensive analysis of the characteristics of Microblog platform news.

2 Set up Microblog users and their historical click sequences in Microblog platform news dissemination, as well as other Microblog platform news dissemination vectors, set up sequence representation sets, use NLTLS model to design Microblog platform news dissemination representation model, and calculate news dissemination attention weights of fully connected networks through multiple hidden layers to complete Microblog platform news dissemination vector representation.

3 In the feature extraction of Microblog platform news dissemination, the expected cross entropy of Microblog platform news dissemination vector is calculated to determine the occurrence probability of vector features, and the Microblog platform news dissemination vector is placed in the vector space to determine the similarity of feature text. Use the data in the determined Microblog platform news dissemination vector feature matrix as input data for the multi-layer sliding window model to determine the multi-level sliding window counter. By defining a circular cursor for the global sub window, corresponding to the position of the sub window, and segmenting the Microblog platform news dissemination feature data, the Microblog platform news dissemination feature extraction is completed.

2 Feature extraction of news communication on Microblog platform based on multilevel sliding window model

2.1 Analysis of news characteristics on Microblog platform

News dissemination is mainly carried out through media such as language, images, and videos. Therefore, before extracting the characteristics of news dissemination on Microblog platform, it is necessary to analyse the characteristics of the news. Microblog platform news is relative to professional news. The authenticity of news on this platform needs to be constantly examined. News based on Microblog platform differs from traditional professional news in that the main difference lies in the authenticity, objectivity, and comprehensiveness of the news (Sun, 2021b). Microblog platform news has the following characteristics:

- Feature 1: Microblog platform has good timeliness of news and plays a role as a source of information during the information vacuum period.

  The timeliness of news dissemination is crucial for the freshness of news, which mainly refers to the time difference between the occurrence of news events and their dissemination. The timeliness of news release and dissemination is related to the value of news. As a platform for quickly releasing news in the first place, Microblog platform can compensate for the lack of professional news personnel present, so as to release statements in the first place. Especially when critical events occur, Microblog platform can use the information released by the on-site public as a buffer source for news timeliness (Zhou et al., 2021). In this platform, it is not necessary for the publisher to be a professional journalist. The publisher mainly serves as a news witness and provides a buffer time for subsequent news dissemination.

- Feature 2: The writing method of news content is free and does not restrict any templates.

  As is well known, professional news needs to be written based on a fixed set of templates before dissemination, and multiple elements of the news must be clearly explained, and key parts of the news need to be carefully edited without any omissions (Sun and Du, 2021). Unlike traditional news content writing models, the writing and publishing of news content on Microblog platform is no longer limited by traditional news formats and no longer requires word count restrictions. It can be directly published based on what you see and hear.

- Feature 3: News content is rich in personal emotional bias.

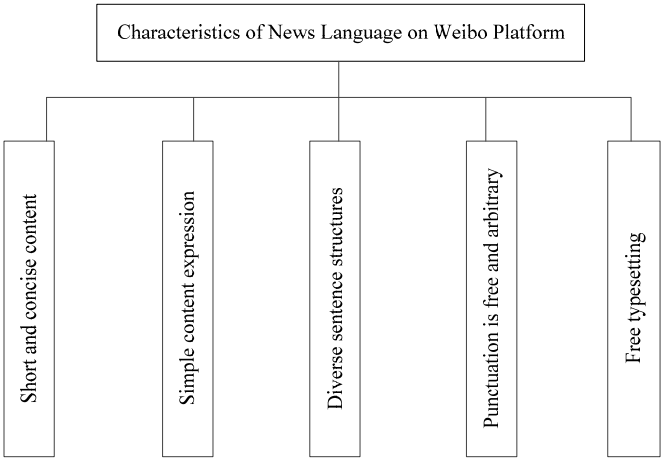
  There is a significant difference from traditional neutral explanations of news. The emotional bias of the publisher can be clearly felt in Microblog platform news, and the transparency of the news on this platform makes it easier for the public to accept. The platform's news release disseminators have not received professional training, and there is a significant difference in the level of mastery in language expression and news templates compared to traditional news media. The news release time on Microblog platform is more free and emotional expression is clearer.

- Feature 4: The whole people are mobilised in the news dissemination, and the news truth is restored in the fragmentation mode.

The fragmentation news in the Microblog platform news is provided by the event witnesses. The news content released by each person varies greatly due to their different perspectives. Different news views provided by users of Microblog platform collect news events from more different perspectives, and comprehensively collect news truth in the form of fragmentation (Kozik et al., 2022).

Based on the above analysis, the language characteristics of Microblog platform news are also crucial for the effectiveness of news dissemination. Therefore, based on the analysis of the above characteristics, this article further summarises the language characteristics of Microblog platform news, as shown in Figure 1.

Figure 1 Schematic diagram of language characteristics of Microblog platform news



In the analysis of Microblog platform news characteristics, four major characteristics of Microblog platform news were identified, namely, the timeliness characteristics of Microblog platform news, the writing characteristics of news content, the emotional bias characteristics of news content, and the characteristics of fragmentation model to restore the truth of news. On this basis, the language characteristics of Microblog platform news were summarised, and the comprehensive analysis of Microblog platform news characteristics was completed, laying the foundation for the subsequent extraction of communication characteristics.

2.2 Research on news communication representation model on Microblog platform

Based on the above analysis of the characteristics of Microblog platform news, in order to achieve feature extraction of Microblog platform news dissemination, a Microblog platform news dissemination representation model is constructed. From the perspective of modelling news dissemination on Microblog platforms, the construction of the model fully reflects its dissemination characteristics, and the final feature extraction is carried out based on the model.

The advantages of the Microblog platform news dissemination representation model lie in real-time and diversity. As a real-time social media platform, Microblog can timely

publish and disseminate news information, allowing users to stay up-to-date with the latest updates. In addition, users on the Microblog platform come from different backgrounds and fields, and they can freely publish content in the form of text, images, videos, etc. making news presented to users in a more vivid and diverse way. These advantages make Microblog platforms a fast, diverse, and highly interactive news dissemination channel.

In the modelling of news communication on Microblog platform, it is necessary to first clarify the distribution of topics in the news communication (Ertek and Kailas, 2021), and then determine textual communication content such as the title, abstract, category, and subcategory of the news. These contents are unified into representation vectors in the news communication, which facilitates the data-driven representation of the characteristics of news communication on Microblog platform and is conducive to the effectiveness of this study. In the vectorisation processing of subject distribution in news communication, set the Microblog platform to represent Microblog users and their historical click sequences in news communication, namely:

$$D_i = \{d_1, d_2, ..., d_n\} \tag{1}$$

In formula (1), D_i represents the collection of historical click news topic sequences, d_1, d_2, …, d_n represents the composition data in the news topic sequence, and n represents the number of topic distributions.

For news dissemination on Microblog platform, it is also necessary to consider other Microblog platform news dissemination vectors, and set a sequence composed of several words to represent these features as:

$$W_i = \{w_1, w_2, ..., w_m\} \tag{2}$$

In formula (2), W_i represents the sequence of news dissemination vectors on the Microblog platform, and m represents the number of news dissemination vectors.

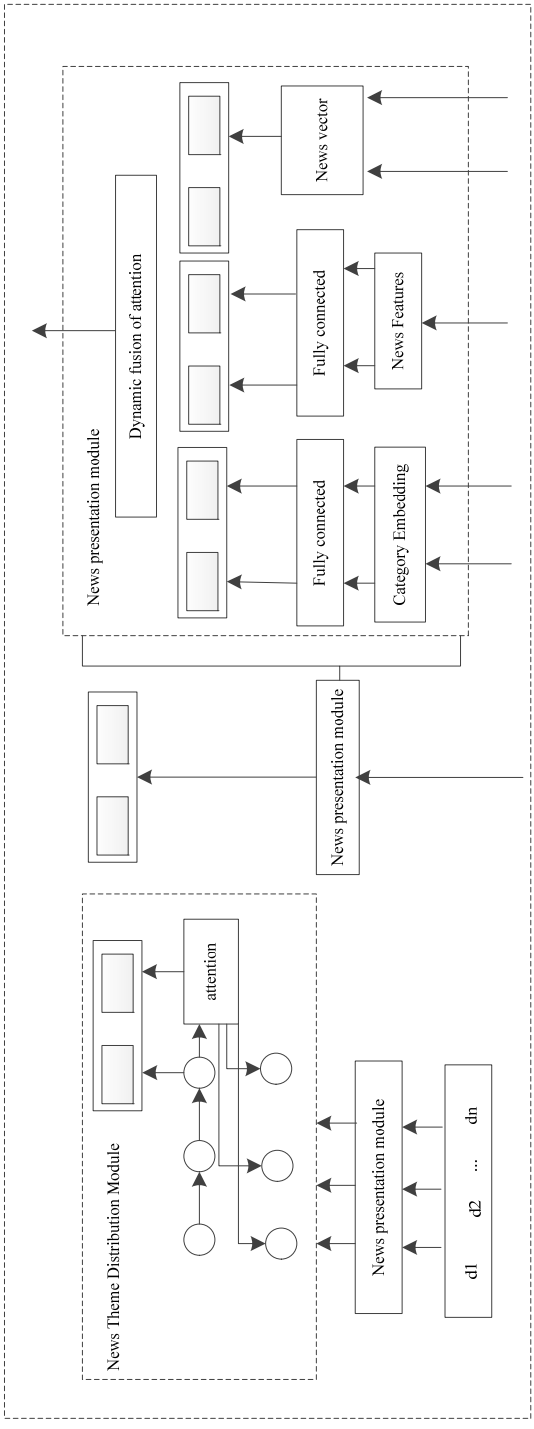
The formula for calculating the distribution vector of topics in news dissemination on Microblog platform is:

$$\theta_d = [z_d]i = 1...k, \sum_{i=1}^{k} z_d \tag{3}$$

In formula (3), θ_d represents the topic distribution vector in Microblog platform news dissemination, z_d represents the probability of the existence of potential news topics, and k represents the number of topic distribution vectors in news dissemination.

Based on the above representation of Microblog platform news dissemination vectors, in order to construct a Microblog platform news dissemination representation model, this model is constructed using the NLTLS model (Przybya and Soto, 2021) for the study of Microblog platform news dissemination representation model. By substituting the news dissemination vector in the model, the Microblog platform news dissemination is represented, taking into account the two core steps of the Microblog platform news dissemination audience, and conducting Microblog platform news dissemination vector training, sample input, etc. to complete the construction of the Microblog platform news dissemination representation model. The schematic diagram of the constructed model is shown in Figure 2.

Figure 2 Schematic diagram of news dissemination representation model on Microblog platform



According to the Microblog platform news dissemination representation model constructed in Figure 2, the core of this model is the news topic distribution module and news dissemination representation module. By combining these two modules, news is represented as a vector in the user representation module, and the main features of news dissemination are obtained through different representation sequences. Design an extractor for news dissemination content in this model, obtain key sequences such as news titles and abstracts, and output representation vectors about news dissemination, namely:

$$c_i = ReLu\left(v_i / d_n\right) + b_i \tag{4}$$

In formula (4), $ReLu(.)$ represents the activation function, v_i represents the splicing of the news representation vector, and c_i represents the output news propagation representation vector.

By setting the attention module in the model, the different importance degrees of news content on different Microblog platforms are represented as vector relations. A fixed news transmission vector is initialised and its attention weight is calculated through a fully connected network with multiple hidden layers (Costa et al., 2021). The results are as follows:

$$\beta_i = \frac{\exp\left(e^j \tanh\left(y_i c_i\right)\right) + b_i}{\sum_{i=1}^{n} \exp\left(e^j \tanh\left(y_i c_i\right)\right) + b_i\right)} \tag{5}$$

In formula (5), β_i represents the attention weight result, e^j represents the trainable parameters of the model, y_i represents the attention network, and h represents the dimension of the news dissemination vector.

Finally, based on the above analysis, the result of the Microblog platform news dissemination representation vector is:

$$\gamma_i = \sum_{i=1}^{n} \beta_i c_i \tag{6}$$

In formula (6), γ_i represents the vector for news dissemination on Microblog platform.

In the construction of the Microblog platform news dissemination representation model, Microblog users and their historical click sequences, as well as other Microblog platform news dissemination vectors, are set up with sequence representation sets. The NLTLS model is used to design the Microblog platform news dissemination representation model, and the news dissemination attention weights of the fully connected network are calculated through multiple hidden layers to complete the Microblog platform news dissemination vector representation.

2.3 Extraction of news communication features on Microblog platform

Based on the above representation of Microblog platform news dissemination vectors, in order to better achieve feature extraction of Microblog platform news dissemination; this article introduces a multi-level sliding window model for feature extraction.

The multi-level sliding window model is a method used for data analysis and feature extraction. This model is based on the concept of sliding windows, dividing data into windows at different scales and extracting corresponding features. In a multi-level sliding

window model, multiple window sizes are usually set, starting from the smaller window and gradually increasing the window size. This can capture information at different levels, from local details to global structure. By extracting features from the data within each window, feature representations at multiple scales can be obtained (Miao, 2021).

In the process of extracting news dissemination features on Microblog platforms using a multi-level sliding window model, multi-scale feature extraction can be performed. By setting sliding windows of different sizes to capture information at different levels, data can be more comprehensively understood. And this model takes into account contextual relevance, utilising the overlap and connection between adjacent sliding windows to better understand the overall meaning of the data and improve the expression ability of features. Finally, the multi-level sliding window model has strong robustness and can cope with noise and changes in the data, improving the stability of feature extraction. In summary, the multi-level sliding window model can more comprehensively and accurately capture the relevant features of news dissemination in the feature extraction of news dissemination on Microblog platforms.

Firstly, calculate the probability of news dissemination vectors on Microblog platform becoming their features (Zhao et al., 2021). It is determined by calculating the expected cross entropy of the microblog platform news propagation vector, and its calculation formula is as follows:

$$u(x) = p_i \sum_{i=1}^{n} p(c_i|w) \log \frac{p(c_i|w)}{p(c_i)} \tag{7}$$

In formula (7), $u(x)$ represents the expected cross entropy result of the microblog platform news dissemination vector, p_i represents the vector feature probability, and w represents the vector weight value.

Then, based on the probability of the occurrence of Microblog platform news dissemination vector features determined above, the similarity between feature texts is further determined, which facilitates subsequent feature extraction. Put the news dissemination vector of microblog platform into vector space, and express it as word sequence (Hsueh et al., 2022) through screening, namely:

$$t = \{t_1, t_2, ..., t_m\} \tag{8}$$

In formula (8), m represents the length value of the vector feature.

According to the determined feature length value, the corresponding weight value is represented as 0, and the remaining vector data is represented as 0. The matrix composed of all Microblog platform news dissemination feature vectors is shown in Figure 3.

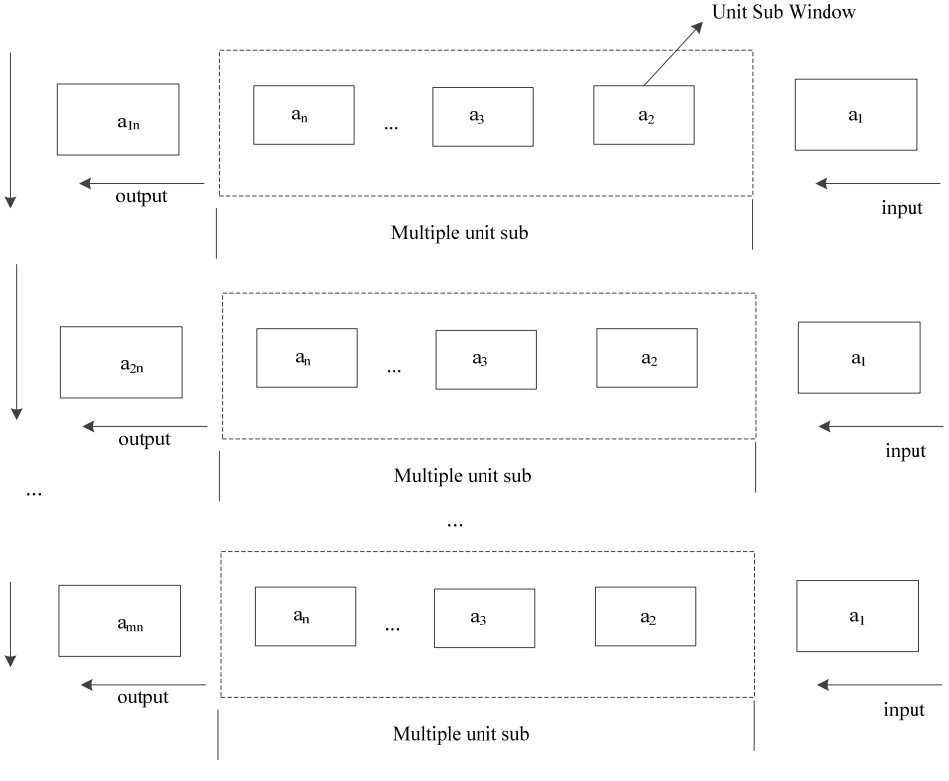
Finally, the data in the determined Microblog platform news dissemination vector feature matrix will be used as input data for the multilevel sliding window model, and the final extraction of Microblog platform news dissemination features will be completed through the processing of this model.

The constructed multilevel sliding window model (Wei et al., 2021) is shown in Figure 4.

Figure 3 Schematic diagram of Microblog platform news dissemination vector feature matrix

	w_1	w_2	⋯	w_m
D_1	t_{11}	t_{12}	⋯	t_{1m}
D_2	t_{21}	t_{22}	⋯	t_{m2}
⋯	⋯	⋯	⋯	⋯
D_N	t_{n1}	t_{n2}	⋯	t_{mn}

Figure 4 Schematic diagram of a multilevel sliding window model for extracting news dissemination features on Microblog platform



Based on the constructed multilevel sliding window model for news dissemination feature extraction on Microblog platform, a multilevel sliding window counter is calculated. The length of this counter is a one-dimensional fixed array used to record the support of feature data extraction, the support count of each window, etc. namely:

$$ItemsWindow = [num_{all}, num_{w1}, num_{w2}, ..., , num_{wn}] \tag{9}$$

In formula (9), *ItemsWindow* represents the multilevel sliding window counter, num_{all} represents the support data counter, and num_{wn} represents the support counters for each window.

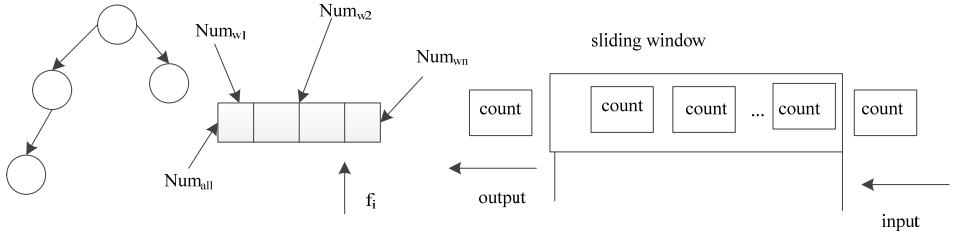
Define a global sub window loop cursor and correspond the sub window positions corresponding to the above counters one by one (Wang et al., 2021). The update method is as follows:

$$F_i = (f_i + 1)\%m \tag{10}$$

In formula (10), m represents the size of the multilevel sliding window, F_i represents the count update result, and f_i represents the circular cursor of the window.

The corresponding sub window positions are shown in Figure 5.

Figure 5 Schematic diagram of corresponding sub window positions



Based on the corresponding relationship mentioned above, feature extraction is performed on the Microblog platform news dissemination vector. The dissemination vector is first segmented in a multilevel sliding window model for subsequent feature extraction. The result of news dissemination vector segmentation on Microblog platform is represented as:

$$\sigma_i = \frac{(L+R)}{2E} \tag{11}$$

In formula (11), σ_i represents the window width of the initial segmentation of the current Microblog platform news dissemination vector, L represents the left boundary value in the undivided vector, R represents the right boundary value in the undivided vector, and E represents the remaining undivided Microblog platform news dissemination vector.

The feature extraction of the segmented Microblog platform news dissemination vector features is as follows:

$$\delta_i = \sum \frac{(L+R)}{2E} + (k+1)H \tag{12}$$

In formula (12), δ_i represents the result of feature extraction for news dissemination on Microblog platform, k represents the number of multilevel sliding windows, and H represents the step size of feature extraction for window increase.

In the feature extraction of Microblog platform news transmission, the expected cross entropy of Microblog platform news transmission vector is calculated to determine the probability of vector features. The Microblog platform news transmission vector is placed in the vector space to determine the similarity of feature text. The data in the determined Microblog platform news transmission vector feature matrix is taken as the input data of the multi-layer sliding window model. To determine the multi-level sliding window counter, by defining the circular cursor of the global sub-window, corresponding to the position of the sub-window, and segmentation of the Microblog platform news transmission feature data, complete the Microblog platform news transmission feature extraction.

3 Experimental testing research

3.1 Experimental plan design

To demonstrate that the method proposed in this article can effectively extract news dissemination features from Microblog platform, an experimental analysis was conducted. In the experiment, news communication content from the past three months was selected as the research object on Microblog platform. In this testing study, two datasets were selected from the Microblog platform news corpus as the research subjects, with a total of 500,000 news dissemination content. From these datasets, news data was randomly selected for the extraction of news dissemination features. The types and quantities of news dissemination content on Microblog platform in this study are shown in Table 1.

Table 1 Details of the types and quantities of news dissemination content on Microblog platform

Theme	*Number*
Financial news	300
Domestic sports news	260
Technology news	170
Military news	100
Political news	380
Cultural and educational news	290
Entertainment news	210

3.2 Experimental plan and indicators

This experiment compares and verifies the method proposed in this paper with method of Li and Liu (2021) and method of Yin and Chen (2022), using the recall rate of news dissemination feature extraction on Microblog platform and the imbalanced coefficient of feature extraction data as indicators.

1 The recall rate of feature extraction for news dissemination on Microblog platform: this indicator measures the correct proportion of feature extraction data and can reflect the completeness of feature extraction. It is the proportion of actual positive samples extracted as positive samples. The calculation formula for this indicator is expressed as:

$$U_i = \frac{1}{n}\sum_{i=1}^{n} p_i / p_j \times 100\% \tag{13}$$

In formula (13), U_i represents the recall rate of news dissemination feature extraction on Microblog platform, p_i represents macroscopic feature data, and p_j represents data recalled by feature extraction.

2 The imbalanced coefficient of feature extraction data for news dissemination on Microblog platform: this indicator reflects whether the feature data obtained in feature extraction conforms to the characteristics of news dissemination. The higher the imbalanced coefficient, the poorer the quality of feature extraction, and the lower the imbalanced coefficient, the better the quality of extracted features.

3.3 Experimental results

3.3.1 Analysis of the recall rate of news dissemination feature extraction on Microblog platform

The experiment first used method in this paper, method of Li and Liu (2021) and method of Yin and Chen (2022) to extract any 1,000 Microblog platform news dissemination feature data. The recall of the three methods in news dissemination feature extraction was compared, and the results are shown in Table 2.

Table 2 The recall rate of news dissemination feature extraction on Microblog platform (%)

News dissemination data/piece	*Method in this paper*	*Method of Li and Liu (2021)*	*Method of Yin and Chen (2022)*
200	99	96	97
400	99	95	95
600	98	93	93
800	98	90	91
1,000	98	89	90

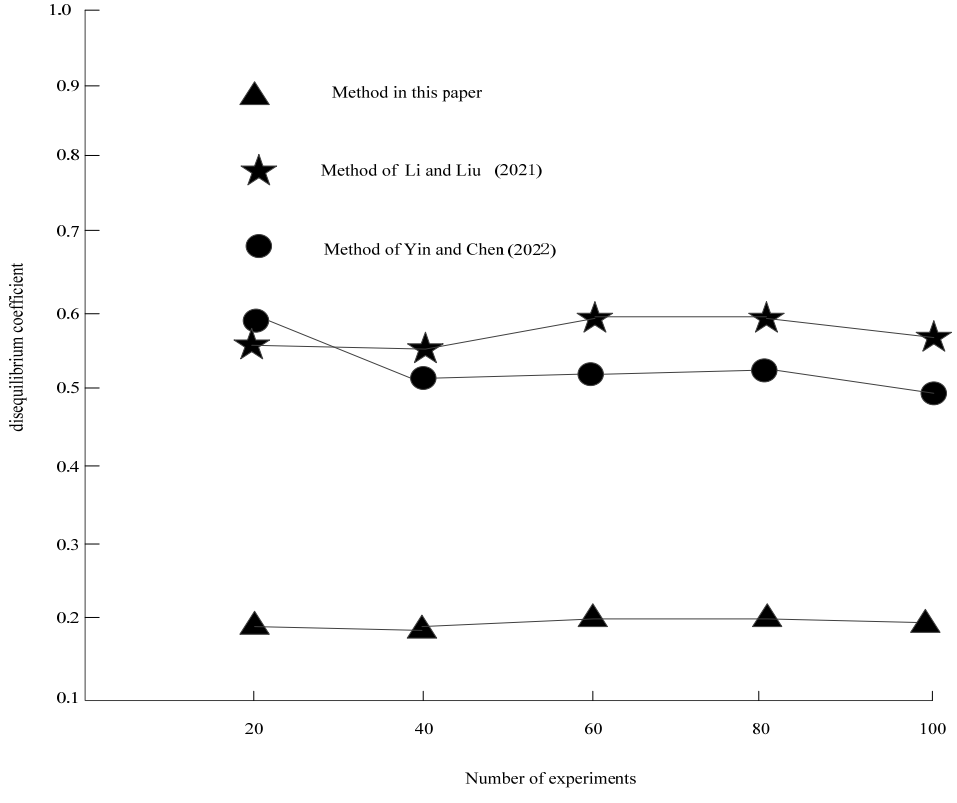
Analysing the experimental results in Table 2, it can be seen that as the amount of data on news dissemination features on Microblog platform continues to increase, the recall rate of feature extraction using method in this paper, method of Li and Liu (2021) and method of Yin and Chen (2022) has undergone certain changes. Among them, the recall rate of feature extraction by method of Li and Liu (2021) and method of Yin and Chen (2022) showed a decreasing trend, while the recall rate of feature extraction by method in this paper remained around 98%. Compared to method in this paper, the recall rate of feature extraction was higher, verifying the effectiveness of the proposed method. The reason why this method has a high recall rate is that it analyses the characteristics of Microblog platform news and constructs a Microblog platform news dissemination representation

model. Using a multi-level sliding window model to extract news dissemination features from Microblog platforms.

3.3.2 Analysis of data imbalance coefficient results for extracting news communication features on Microblog platform

The experiment further analysed the imbalanced coefficients of feature data after extracting news dissemination features using method in this paper, method of Li and Liu (2021) and method of Yin and Chen (2022). The results are shown in Figure 6.

Figure 6 Result of imbalanced coefficient of data extracted from news dissemination features on Microblog platform



Analysing the test results in Figure 6, it can be seen that there are certain differences in the imbalanced coefficients of feature data after extracting news dissemination features between method in this paper, method of Li and Liu (2021) and method of Yin and Chen (2022). From the curve in the figure, it can be seen that the imbalance coefficient of method in this paper remains below 0.2, while the imbalance coefficients of method of Li and Liu (2021) and method of Yin and Chen (2022) are always higher than those of method in this paper, which verifies the feasibility of this method. The reason why this method can reduce the imbalanced coefficient of extracted data is that it takes news communication theme content as input and Microblog platform news communication

representation vector as output to construct a Microblog platform news communication representation model, thereby reducing the imbalanced coefficient.

4 Conclusions

This article proposes a feature extraction method for news dissemination on Microblog platforms based on a multi-level sliding window model. By determining the characteristics of the Microblog platform news, the NLTLS model is used to design the Microblog platform news dissemination representation model, and the news dissemination attention weight of the full connected network is calculated through multiple hidden layers, the expected cross entropy of the Microblog platform news dissemination vector is calculated, the occurrence probability of vector features is determined, the data in the determined vector feature matrix is used as the input data of the multi-layer sliding window model, and the multi-level sliding window counter is determined. By defining a circular cursor for the global sub window, corresponding to the position of the sub window, and segmenting the Microblog platform news dissemination feature data, the Microblog platform news dissemination feature extraction is completed. The experimental results show that the recall rate of feature extraction for news dissemination on Microblog platforms using this method can reach around 98%, and the imbalanced reduction coefficient remains below 0.2.

References

Baek, S., Jung, W. and Han, S.H. (2021) 'A critical review of text-based research in construction: data source, analysis method, and implications', *Automation in Construction*, Vol. 132, No. 10, pp.1–16.

Costa, J.P., Rei, L., Stopar, L., Fuart, F. and Wallace, J. (2021) 'NewsMeSH: a new classifier designed to annotate health news with MeSH headings', *Artificial Intelligence in Medicine*, Vol. 114, No. 1, p.102053.

Ertek, G. and Kailas, L. (2021) 'Analyzing a decade of wind turbine accident news with topic modeling', *Sustainability*, Vol. 13, No. 14, pp.1–9.

Hsueh, H.F., Guthke, A., Woehling, T. and Nowak, W. (2022) 'Diagnosis of model errors with a sliding time-window Bayesian analysis', *Water Resources Research*, Vol. 58, No. 2, pp.1–48.

Kozik, R., Kula, S., Choras, M. and Wozniak, M. (2022) 'Technical solution to counter potential crime: text analysis to detect fake news and disinformation', *Journal of Computational Science*, Vol. 60, No. 4, pp.1–8.

Li, S. and Liu, Y. (2021) 'News video title extraction algorithm based on deep learning', *IEEE Access*, Vol. 2021, No. 9, pp.12143–12157.

Miao, T. (2021) 'Using the ship-gram model for Japanese keyword extraction based on news reports', *Complexity*, Vol. 2021, No. 13, pp.1–9.

Przybya, P. and Soto, A.J. (2021) 'When classification accuracy is not enough: explaining news credibility assessment', *Information Processing & Management*, Vol. 58, No. 5, pp.1–20.

Reynosa-Guerrero, J., Garcia-Huerta, J.M., Vazquez-Cervantes, A., Reyes-Santos, E. and Jimenez-Hernandez, H. (2021) 'Estimation of disparity maps through an evolutionary algorithm and global image features as descriptors', *Expert Systems with Applications*, Vol. 165, No. 11, p.113900.

Sun, B. (2021a) 'Attention feature network extraction combined with the generation algorithm of multimedia image description', *Advances in Multimedia*, Vol. 2021, No. 1, pp.1–8.

Sun, M.X. (2021b) 'Feature extraction method of face image texture spectrum based on a deep learning algorithm', *International Journal of Biometrics*, Vol. 13, No. 23, pp.1–12.

Sun, N. and Du, C. (2021) 'News text classification method and simulation based on the hybrid deep learning model', *Complexity*, Vol. 2021, No. 21, pp.1–11.

Tajanpure, R. and Muddana, A. (2021) 'Circular convolution-based feature extraction algorithm for classification of high-dimensional datasets', *Journal of Intelligent Systems*, Vol. 30, No. 1, pp.1026–1039.

Wang, W., Fang, C., Liu, S. and Xiang, Y. (2021) 'Reliability analysis and optimization of multi-state sliding window system with sequential demands and time constraints', *Reliability Engineering & System Safety*, Vol. 208, No. 4, pp.1–12.

Wei, Y.Q., Liu, D.Y., Boutat, D., Liu, H.R. and Wu, Z.H. (2021) 'Modulating functions based model-free fractional order differentiators using a sliding integration window', *Automatica*, Vol. 130, No. 9, pp.1098–1105.

Yin, X. and Chen, L. (2022) 'A cross-modal image and text retrieval method based on efficient feature extraction and interactive learning CAE', *Scientific Programming*, Vol. 2022, No. 4, pp.1–12.

Zhang, C., Zhang, Y. and Li, F. (2021) 'Feature extraction of sequence of keystrokes in fixed text using the multivariate Hawkes process', *Mathematical Problems in Engineering*, Vol. 2021, No. 23, pp.1–16.

Zhao, Y., Yun, J., Gao, S. and Liu, L. (2021) 'News image-text matching with news knowledge graph', *IEEE Access*, Vol. 27, No. 11, pp.108017–108027.

Zheng, X., Tao, Y., Zhang, R., Yang, W. and Liao, Q. (2021) 'TimNet: a text-image matching network integrating multi-stage feature extraction with multi-scale metrics', *Neurocomputing*, Vol. 465, No. 20, pp.540–548.

Zhou, Y., Jin, H., Liu, B. and Chen, Y. (2021) 'An improved feature parameter extraction algorithm of composite detection method based on the fusion theory', *Journal of Sensors*, Vol. 2021, No. 2, pp.1–10.