

## RESEARCH ARTICLE

# Enhancing Low Carbon Awareness in Social Media Discourse: A Fuzzy Clustering Approach

CHAO HAN<sup>1b</sup> AND XUEZHI SUN

Krikk University, Bangkok 101400, Thailand

Corresponding author: Chao Han (hanchao024@163.com)

**ABSTRACT** The frequent occurrence of extreme weather makes people pay more attention to environmental protection. To cope with the global climate problem, various countries re-plan social development through the concept of low-carbon. As greatly popularized by the Internet, the topic of low carbon concept is spread more through online social media, so it is urgent to understand the user's attention to low carbon topics in a more intelligent way for subsequent relevant publicity and policy guidance. This paper studies the low-carbon topic of attention in the context of social media. First, the BERT (Bidirectional Encoder Representation from Transformers) model is used to complete the word vector feature extraction of acquired data; Secondly, the FCM method was used to complete the clustering analysis of the main topics in the low-carbon concept, and the PSO method was used to optimize the model. After optimization, the accuracy of clustering for various topics was higher than 80%. For the *Esse* index of cluster center variance, the method proposed in this article is also close to 10% due to other classic methods; Finally, this paper carried out an application test of low-carbon topics in the region, achieved good results, and made a detailed analysis of the distribution of various topics. It can be predicted that this method will provide more public opinion references for low-carbon development paths in various countries and regions in the future, and provide technical support for information dissemination and analysis under social media.

**INDEX TERMS** Social media, FCM, low carbon, text clustering.

## I. INTRODUCTION

As a professional concept in environmental climate science, "low carbon" has been widely spread around the world since the Copenhagen Climate Conference in 2009 [1]. In recent years, its connotation has been magnified infinitely and has been integrated into the concepts of sociology, industrial ecology, economics, and other related fields. From the basic concept of environmental protection, it has been upgraded to a new socio-economic development model applicable to all countries. Each stage of human society development will have its own theme and focus, such as human rights, race, democracy, urbanization, etc. Each topic is continuous, but it will have a different order in different periods [2]. Therefore,

The associate editor coordinating the review of this manuscript and approving it for publication was Maria Chiara Caschera<sup>1b</sup>.

it is important for the spread and development of low-carbon ideas in the future to grasp the importance of different branches of such topics in the minds of the public and focus.

The propagation path of the low-carbon concept can be shown in Figure 1 [3], which is mainly divided into three categories: disseminator, information media, and audience. With the continuous development of technology, the media has also changed tremendously, and the means of communication shown in Figure 1 are more abundant. With the development of the mobile Internet, social media has gradually become the main way of public information communication. Users can pay continuous attention to a topic by commenting on relevant news and short videos and publishing relevant content. In these published contents, users' behavioral characteristics, psychological characteristics, etc. are included. The analysis of the relevance of their published topics will

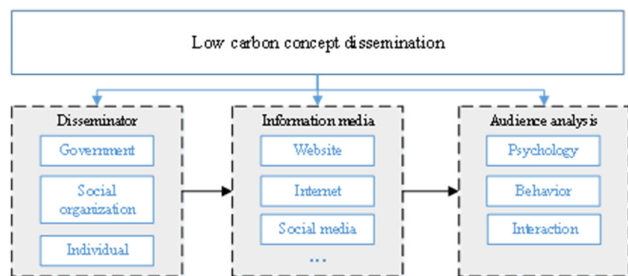


FIGURE 1. The low carbon concept dissemination process.

help to provide a more detailed reference for the further dissemination of low-carbon concepts and relevant policy formulation [4].

The text information analysis technology published by social media users can be seen from the text analysis problem [3]. The main method of text analysis problem is to use artificial intelligence methods to complete intelligent analysis of different users' text information [4]. Recently, with the development of machine learning, using intelligent clustering and classification methods to study such issues has become a hot spot [2]. The relevant content on social media has a rich variety and several items, which makes it difficult to use supervised learning to achieve classification. Therefore, semi-supervised clustering has become an important method of social media theme analysis [5]. Clustering is to group or abstract a set of objects into certain classes with similarity, while text clustering pays more attention to the text content. It is an important application of clustering analysis technology in text processing. At the same time, due to different user portraits and different expression methods, most of the content is not either or, and there is a certain degree of fuzziness. Therefore, combined with text analysis methods, character vector features are extracted, and fuzzy clustering methods are used to mine related topics, to complete further refinement of topics of concern, which will promote future policy guidance and public opinion development.

This paper studies the topic attention of the low carbon concept in the process of social media communication, and the overall framework is shown in Figure 2. The BERT model is utilized for extracting the word vector from obtained social media text information. Subsequently, the fuzzy clustering method is applied to perform the clustering of related low-carbon topics. This approach facilitates research on the attention devoted to low-carbon topics, ultimately offering a reference for future public opinion policy orientation and low-carbon development. The contributions of this paper are as follows:

(1) The crawler method is employed for acquiring social media text data, and the BERT model is utilized for high-precision extraction of word vector features.

(2) The clustering model is optimized using the PSO-FCM method, achieving recognition precision exceeding 80% for each topic in the model test.



FIGURE 2. Framework for the low carbon topic research based on FCM.

(3) In accordance with the proposed model framework, an actual test was conducted, resulting in the completion of topic subdivision within the low-carbon concept of social media in the region. The findings indicate that the user group expressed the highest attention to low-carbon technology and transportation, thereby offering public opinion references for subsequent low-carbon development.

The rest part of the paper is as follows: In section II, the related works for the social media analysis and FCM clustering are introduced. Section III introduces the methods used, and section IV describes the experiment and results in an analysis of model optimization and comparison, where the model application is also discussed. In section V, we discuss the result and the notice that should be paid for the low carbon development; conclusion is presented at last.

## II. RELATED RESEARCH WORKS

### A. SOCIAL MEDIA DATA MINING USING THE MACHINE LEARNING METHODS

In the research on text data of social media, the main object of current research is social media APPs with many users, such as Weibo, Twitter, and Facebook. Davidov et al. [7] analyzed the difficulties of microblog analysis, solved the problem of data imbalance, and built an emotion classification system through supervised learning; Lin et al. [8] used text statistics to track hot event topics on social networks Barbosa and Feng [10] detected the emotion of Twitter by exploring the text structure and keyword information; Jiang et al. [11] improved the accuracy of emotion classification through adding relevant texts. Zhang et al. [12] used an emotion dictionary and machine learning methods to classify emotional polarity. From the above research, we can see that in the previous social media information research, there were many types of research on user sentiment analysis through text data mining, dictionary modeling, or using deep learning-related learning methods, while there were few types of research on the topic of the text in a certain field. This is unfavorable for using social media information to grasp the trend of users' topics of concern. It is urgent to study their topics of concern through clustering to help analyze current network trends.

### B. THE FCM MODEL RESEARCH ON THE TEXT ANALYSIS

For topic analysis of social media content, it is not appropriate to conduct fine-grained research on it through affective computing and classification methods, while clustering methods can automatically analyze topics of low-carbon topics according to the set number of clusters to help relevant personnel complete the analysis. The earliest relatively mature cluster-

ing algorithm for text clustering is the K-Means algorithm [13], which is a segmentation clustering algorithm; After that, Ng and Han [14] proposed the CLARINS algorithm to improve the K-Means, which can efficiently cluster large-scale data sets. Traditional hard partition clustering strictly divides a sample point into a category, but in real life, the category attribute of many things is not so obvious. In order to correctly describe the real relationship between objects and classes, it is necessary to find a clustering method different from the traditional hard partitioning method. The fuzzy set theory proposed in the last century [15] provides a favorable theoretical basis for fuzzy clustering partitioning. The most successful one is the FCM algorithm proposed based on the K-means algorithm [16]. PCM (possibility c-means) algorithm, which solves the problem that the traditional FCM algorithm is sensitive to noise data points and was improved in the following years [17]; FCS (Fuzzy compactness and separation) algorithm, but the algorithm has a fixed core boundary, and all objects in a core have the same membership, thus reducing the classification accuracy [18]. At present, fuzzy clustering analysis has been widely used in parameter estimation, climate analysis, channel equalization, medical diagnosis, and water quality analysis.

Through the above research, it can be found that the FCM method can complete the information clustering of imprecise situations due to the existence of fuzziness, and the low carbon-related theme content of social media is similar. It needs to complete its theme analysis through continuous fuzzy iteration of related words and sentences and relationship mining, so the FCM method can be used to complete related tasks. In the practical applications listed above, FCM has also made a variety of improvements to solve the problems of initial value sensitivity and falling into a locally optimal solution. In this paper, according to the data type and application needs, we use the PSO method to optimize the model to form a low-carbon topic clustering research under social media text analysis.

### III. MODEL ESTABLISHMENT USING THE TEXT FROM THE SOCIAL MEDIA

#### A. BERT MODEL FOR THE TEXT ANALYSIS

Since computers can't understand human language and characters, when doing natural language processing (NLP) tasks, we need to first solve the problem of how to express text in computers. There are two commonly used representations: one hot representation and distribution representation [19]. The method based on a neural network language model predicts the next word according to the above in the window size by building a neural network and training the model to optimize it. The word vector is generated incidentally after the model optimization. Common methods using neural network language models to generate word vectors include Word2Vec, and BERT.

The BERT is a pre-training language model proposed in 2018, which can complete various NLP tasks including text

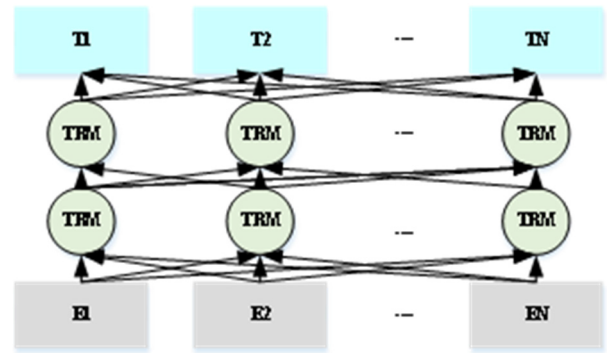


FIGURE 3. Framework of the BERT model.

question and answer discrimination, text emotion analysis, etc. The BERT method employed the encoder part of the transformer model to build the network structure, which improves the model in-text understanding. The structure of the BERT model is shown in Figure 3:

As shown in Figure 3, each row of TRM in the figure represents the Encoder part of a Transformer structure. The network is constructed through the connection of multi-layer TRM to realize the model function. The number of stacked layers corresponds to the cycles of the Encoder part of the Transformer. Parameters in the BERT model are trained through two tasks that are MLM and NSP. When training, two tasks are carried out at the same time, and the loss is the sum of two tasks. Although the training corpus of the BERT model is very large, and the constructed lexicon includes most emotional words, the model will still encounter some words or symbols that are not included in the lexicon. To ensure the input accuracy of any text data, the BERT model can smoothly convert it into a digital sequence. In addition to the words recorded from the corpus, several special characters are added to the thesaurus. The training of the BERT model can be carried out word by word without training word vectors, thus avoiding adverse effects on model results due to improper word segmentation and unreasonable word vector training results [20].

#### B. FUZZY CLUSTERING MODEL FOR THE TEXT ANALYSIS

In real life, most things have relative fuzziness, which is not an either-or accurate situation. Moreover, there are certain similarities between text words and sentences. If we only cluster them, we will not get good results, so we should use the advantages of the fuzzy clustering algorithm. Fuzzy clustering algorithms often have good adaptability and anti-noise ability. Through fuzzy segmentation of text, we can get a better clustering result [21].

The traditional FCM first needs to specify the number of clusters  $c$ , which also denotes centers, depending on the number of sources, and needs to be given in advance; then objective function is optimized to achieve the clustering, and it constraints that the sum of membership degrees of all observation data points belonging to a certain clustering

center is 1. The objective function of FCM is shown in Eq. (1) [22]:

$$\text{minimize } J_{FCM}(U, \hat{A}) = \sum_{k=1}^n \sum_{i=1}^c u_{ki}^m \|x_k - \hat{a}_i\|^2 \quad (1)$$

the restraint condition for the FCM can be expressed as Eq. (2):

$$s.t. \sum_{i=1}^c u_{ki} = 1 (k = 1, 2, \dots, n) \quad (2)$$

where  $U = (u_{ki})$  is the membership matrix, in which the element  $u_{ki} \in [0, 1]$  describes the degree to which a text observation signal sample point belongs to the cluster center corresponding to a blind source voice. The optimization problem of Eqs. (1) and (2) is solved by using the Lagrangian multiplier method to obtain Eqs. (3) and (4). After the initial value is randomly generated for the cluster center, the cluster center matrix  $\hat{A}$  and membership matrix  $U$  are updated iteratively through Eqs. (3) and (4).  $J^{(t)}$  is the objective function value at  $t$ ,  $\varepsilon$  is the precision requirement. If  $|J^{(t+1)} - J^{(t)}| < \varepsilon$ , stop iteration and get the final clustering center matrix, namely, the mixing matrix.

$$\hat{a}_i = \frac{\sum_{k=1}^n (u_{ki})^m x_k}{\sum_{k=1}^n (u_{ki})^m} \quad (3)$$

$$u_{ki} = \frac{\left(\frac{1}{d_{ki}}\right)^{\frac{1}{m-1}}}{\sum_{j=1}^c \left(\frac{1}{d_{kj}}\right)^{\frac{1}{m-1}}} \quad (4)$$

FCM algorithm is expressed as follows
FCM Algorithm
1. Specify the number of cluster centers $c$
2.Cluster center $a_i$ initialization
3. Update matrix $U$ according to the modified cluster center matrix $A$ through Eq. (4)
4. Update the cluster center matrix $A$ with the modified $U$ through Eq. (3)

### C. FCM OPTIMIZATION USING PSO

However, the FCM algorithm also has many shortcomings, which can be summarized as the following: (i) The clustering results are greatly affected by clustering center, and the initial center is randomly selected, lacking mathematical basis and(ii) The algorithm itself is sensitive to outlier data. Therefore, PSO method is selected for optimization to achieve better clustering effect. The algorithm is divided into two steps: the first step is to get the initial center point of the data set to be tested; Second, the FCM algorithm is used to re-cluster the clustering results to obtain more accurate results.

PSO is a stochastic optimization technology based on population, which is considered as a global search strategy. Each member of the particle swarm can represent the potential solution of the optimization problem. For this paper, the goal of optimal initial value can be achieved through example iteration [23].

The specific algorithm is as follows. First, define particles number and their corresponding characteristic parameters:

$$P_j = [C_j \varepsilon_j \sigma_j] j = 1, 2, \dots, Q \quad (5)$$

Performing random initialization of the particles after the definition is complete. Then let the particles be updated. Each particle in the  $k$ th iteration is defined by three characters: (i) the position in the search space  $P_j(k)$ ; (ii) the best position during iterations  $1 \sim k$ ,  $P_{jbest}(k)$ ; (iii) the flight speed  $V_j(k)$ .

Furthermore, the global optimal position of the whole particle swarm is defined as  $P_{jbest}(k)$ , then each particle is iteratively updated during the flight as a function of the velocity  $V_j$  and the position  $P_j$  is defined as

$$\alpha(k) = (\alpha_{max} - \alpha_{min})(k/K)^2 + \alpha_{min} \quad (6)$$

$$c_1(k) = (c_{1max} - c_{1min})(k/K)^2 + c_{1min} \quad (7)$$

$$c_2(k) = (c_{2min} - c_{2max})(k/K)^2 + c_{2max} \quad (8)$$

$$v(k+1) = \alpha(k)v_j(k) + c_1(k)r_1 [P_j(k) - P_{jbest}(k)] + c_2(k)r_2 [P_j(k) - P_{gbest}(k)] \quad (9)$$

$$P_j(k+1) = P_j(k) + v_j(k+1) \quad (10)$$

where  $\alpha(k)$  is the inertial variable, which has a great impact on the particle motion state. When the particle speed is too fast, there is a greater probability of forming a global spatial optimal solution, but when it is too small, the particle speed is too slow, and there is a greater probability of forming an optimal solution in local space. The flow of FCM clustering using PSO is shown in Figure 4.

By combining PSO with FCM algorithm, better performance can be achieved in clustering problems. The global search capability of PSO helps to prevent FCM from getting stuck in local optima, accelerate the convergence process of the algorithm, and improve the adaptability and robustness of the algorithm by optimizing parameters in FCM, such as ambiguity parameters. This comprehensive optimization method enables FCM to handle clustering tasks of different datasets more effectively, find suitable clustering centers more quickly, and achieve better performance in practical applications.

### IV. EXPERIMENT AND RESULT ANALYSIS

In experiment, we first collected the data of low-carbon related words as keywords through the crawler method, and referred to the existing low-carbon research (<https://github.com/bcgov/tfrs>)and relevant data division standards(<https://github.com/google-research/bert>)Based on the analysis, low carbon topics are divided into six categories, including the more important economic, technical, transportation and other contents in low carbon related research,

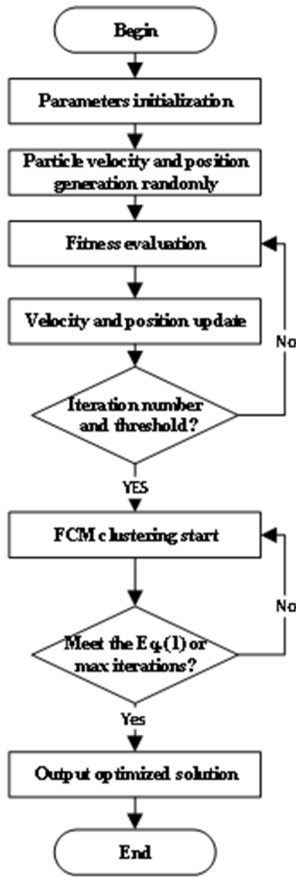


FIGURE 4. The flowchart of the PSO-FCM method.

and the accuracy of their clustering is analyzed. The evaluation indicators used are precision and  $E_{sse}$ , of which  $E_{sse}$  is defined as follows:

$$E_{sse} = \sum_{j=1}^k \sum_{i=1}^h d^2(x_i, c_j) \quad (11)$$

$E_{sse}$  represents the mean square error of all objects and their corresponding cluster centers, providing a quantitative measure of the error between each object and its corresponding cluster center. It is sensitive to the distance between objects and centers and comprehensively evaluates the performance of the algorithm on all clusters. The smaller the value, the better the clustering of the model.

**A. THE RESULT OF THE CLUSTERING PRECISION**

On this basis, the low-carbon topic was divided into six categories in the study, including important economic, technological, and transportation aspects in low-carbon related research, and the accuracy of their clustering was analyzed. The six indicators are Low carbon economy, Low carbon technology, Low carbon sport, Low carbon entertainment, Low carbon transportation, and others. The clustering recognition accuracy is given in Figure 5. The results in Figure 5 show that the clustering accuracy of the proposed method for

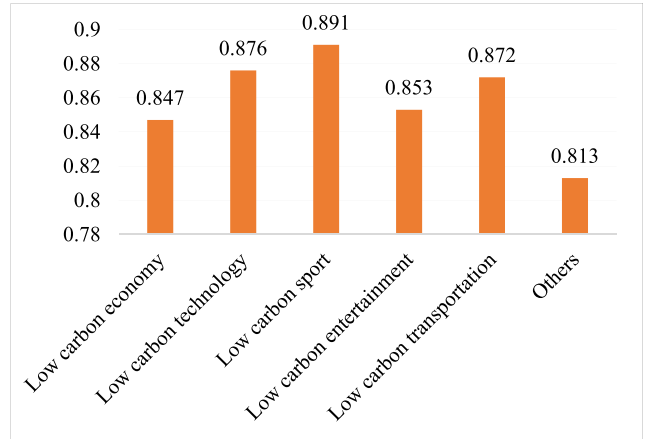


FIGURE 5. The result of the clustering precision of different topics.

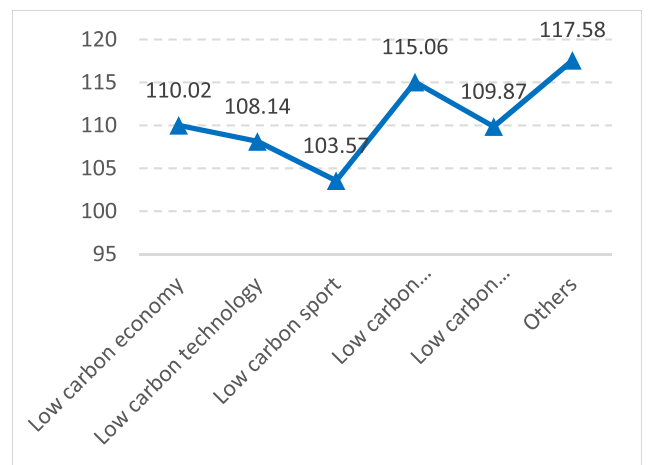


FIGURE 6. The result of the Esse of different topics.

each topic exceeds 80%, indicating that the proposed method can better complete clustering in subsequent analysis.

This paper calculates the mean square error of the clustering center shown in equation (11) when evaluating the clustering effect for each low-carbon topic, and the results are shown in Figure 6.

According to the central clustering variance shown in Figure 6, the model proposed is relatively balanced and its error distribution is relatively uniform. It can prove that the robustness and generalization capacity of the model are improved after correcting the number of iterations and the initial optimization of PSO.

**B. THE COMPARISON AMONG DIFFERENT CLUSTERING METHOD**

In order to better illustrate the FCM effect after PSO optimization in this paper, conventional FCM, K-means, and GA-FCM are selected for comparison in the comparison experiment. The results are shown in Figure 7.

In Figure 7, the left coordinate axis represents precision, while the right coordinate axis is  $E_{sse}$ . Through the compar-

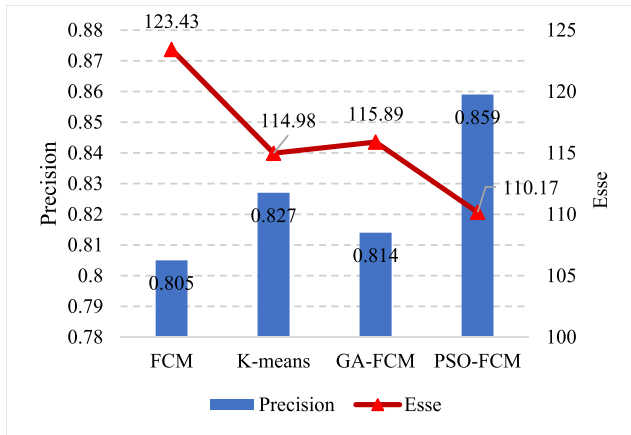


FIGURE 7. Comparison results of methods.

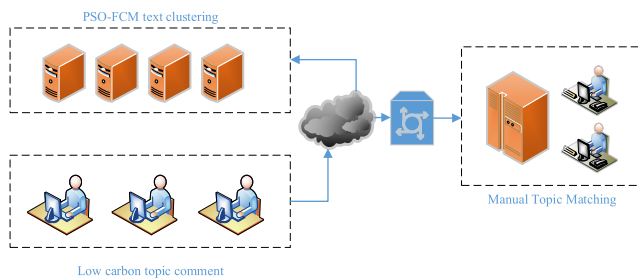


FIGURE 8. The framework for the practical test.

ison of different methods, it can be found that the precision and  $E_{sse}$  are better than the comparison method. At the same time, it is easy to see by comparing the curve trend in the figure that with the improvement of precision, the  $E_{sse}$  of each method also decreases, which is consistent with the expectation of the model during design. The Precision of the proposed PSO-FCM reaches 0.859, which is 6.7% higher than that of the single FCM model.

### C. MODEL TEST AND APPLICATION

After the verification and comparison of the models in 4.1 and 4.2, the actual application is conducted in this section, and the communication status of current public low-carbon topics in social media is analyzed and studied. During the actual testing process, the data from three quarters of this year are selected for research. First, according to the requirements of the model algorithm framework, social media text data is obtained through the crawler method, then data is pre-processed and sent to the model for training, and data analysis and statistics are conducted by specially assigned personnel. The overall process framework is shown in Figure 8.

For data collection, we push news topics through local news apps and collect comments from relevant users. In order to ensure the accuracy and availability of data, we also conduct targeted push. Then, we clear the collected nearly 2000 survey data, remove invalid data, and perform data segmentation processing. Then, we send it to the proposed

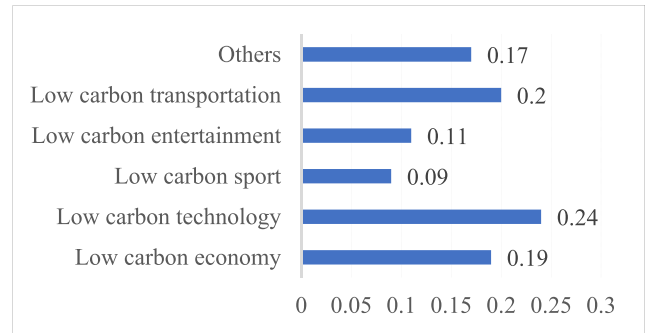


FIGURE 9. The low carbon topic distribution.

PSO-FCM model to complete sentiment clustering. At the same time, We compared the manual labels to complete the actual testing evaluation.

In the test process, users commented on relevant content, automatically clustered such content according to the proposed method framework, and compared it with manual tags, thus completing the practical application test of the method. The results obtained after completing the corresponding content comparison analysis are shown in Figure 9.

Figure 9 shows the proportion of relevant comments from volunteers recruited in the actual test and low carbon-related topics expressed in words. It can be found that volunteers are more concerned about low-carbon technology, transportation, and the economy, which is closely related to the policy orientation of relevant departments, news reporting preferences, and public opinion guidance, and points out the future development direction of low carbon issues. It will also play a good role in promoting economic development and environmental development in the future.

### V. DISCUSSION

Information discussion based on social media is a way of expressing users' emotions. Effective analysis of its theme is helpful to understand users' trends and establish user profiles. Low carbon topic information analysis based on social media is essentially a text clustering analysis. In this paper, clustering analysis is an important means and method in data mining, especially in text mining. The key point to improve the success rate and clustering efficiency in cluster analysis is to select an appropriate algorithm in cluster analysis [24]. In this paper, a text clustering algorithm based on the combination of PSO and FCM is proposed. The algorithm overcomes the shortcomings of the FCM method, reduces sensitivity to the initial point and data input order, and shortens the high time and space complexity associated with using the PSO algorithm alone. This combination makes the algorithm more stable and effective. At the same time, compared with the K-means algorithm, FCM introduces ambiguity factor to describe each data object, so it is more robust; [25] However, the complexity of the FCM is higher due to the introduction of membership factor. Therefore, according to the characteristics of data in social media, this paper selects the FCM

method for topic clustering analysis to avoid either or, so as to improve the clustering effect of low-carbon topics for subsequent analysis.

Considering historical research, low-carbon topics are in line with the social trend in terms of the current global social development [26], which is urgent to be solved in the current global environment. While issues such as human rights, and the gap between the rich and the poor are issues left over from history, it is impossible to change people's values or adjust the country's macro management model in a short time, but low carbon can be easier to do in these aspects. Secondly, the urgency of the low carbon concept itself and the sense of crisis it brings to people also make it easy to ignore the cognitive process of the topic itself. Therefore, the network communication of the low-carbon concept has a vital impact on our production and lifestyle. Developed countries have mastered energy through their technological and financial advantages, and have begun to choose to develop clean industries. However, developing countries are different. They are difficult to complete the low-carbon economic transformation in a short time due to the impact of technology and supply chain industry level, which is why people in these regions pay more attention to low carbon technology. By seeking high-technology and constantly upgrading the industry, low-carbon development can be achieved.

## VI. CONCLUSION

This study investigates public attention to low-carbon topics on social media and proposes a BERT model-based method for word vector extraction. Building upon this, considering the fuzzy nature of topic text content, the FCM method is employed for low-carbon topic clustering analysis, and the PSO method is utilized for model optimization. The results demonstrate that the proposed method achieves clustering accuracy exceeding 80% for each topic. The PSO-FCM method notably enhances clustering accuracy compared to the single FCM method and GA optimization in algorithm verification. In the actual test, the proportions of low-carbon economy, science and technology, transportation, and other contents of interest to local network users on social media were successfully clustered, yielding favorable outcomes. A detailed analysis of each topic's proportion provides valuable insights for future low-carbon analysis. Nevertheless, this study has limitations in the selected samples. With the increasing popularity of short videos, future research will focus on video content clustering. Simultaneously, low-carbon topics will be further refined and analyzed in detail, covering aspects such as low-carbon science and technology, low-carbon lifestyle, and low-carbon transportation. This approach aims to gain a better understanding of public opinion hotspots and formulate precise low-carbon development strategies.

For future research directions, the study could expand its scope by delving into the clustering of video content, considering the growing prevalence of short videos on social media platforms. Additionally, further refinement and in-depth anal-

ysis of low-carbon topics, specifically in areas such as low-carbon science and technology, low-carbon lifestyle, and low-carbon transportation, would contribute to a more comprehensive understanding of public opinion trends. Exploring the evolving landscape of social media and its impact on low-carbon discourse, as well as devising strategies for accurate low-carbon development policies, should be key areas of focus. Integrating emerging technologies and methodologies to enhance the accuracy and efficiency of clustering algorithms could also be explored for advancing the capabilities of future research in the field.

## FUNDING STATEMENT

The study received no funding.

## ACKNOWLEDGMENT

The authors thank the anonymous reviewers whose comments and suggestions helped to improve the manuscript.

## REFERENCES

- [1] A. Markandya, B. G. Armstrong, S. Hales, A. Chiabai, P. Criqui, S. Mima, C. Tonne, and P. Wilkinson, "Public health benefits of strategies to reduce greenhouse-gas emissions: Low-carbon electricity generation," *Lancet*, vol. 374, no. 9706, pp. 2006–2015, Dec. 2009.
- [2] J. Qiao, Y. Liu, F. Hong, and J. Zhang, "A review of catalysts for the electroreduction of carbon dioxide to produce low-carbon fuels," *Chem. Soc. Rev.*, vol. 43, no. 2, pp. 631–675, 2014.
- [3] W. G. Mangold and D. J. Faulds, "Social media: The new hybrid element of the promotion mix," *Bus. Horizons*, vol. 52, no. 4, pp. 357–365, Jul. 2009.
- [4] D. Tang, F. Wei, N. Yang, M. Zhou, T. Liu, and B. Qin, "Learning sentiment-specific word embedding for Twitter sentiment classification," in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics (Long Papers)*, vol. 1, 2014, pp. 1555–1565.
- [5] Y. Li, C. Luo, and S. M. Chung, "Text clustering with feature selection by using statistical data," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 5, pp. 641–652, May 2008.
- [6] A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," in *Proc. 7th Int. Conf. Lang. Resour. Eval. (LREC)*, 2010, pp. 1320–1326.
- [7] D. Davidov, O. Tsur, and A. Rappoport, "Enhanced sentiment learning using Twitter hashtags and smileys," in *Proc. Coling Posters*, 2010, pp. 241–249.
- [8] C. X. Lin, B. Zhao, Q. Mei, and J. Han, "PET: A statistical model for popular events tracking in social communities," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2010, pp. 929–938.
- [9] A. Celikyilmaz, D. Hakkani-Tür, and J. Feng, "Probabilistic model-based sentiment analysis of Twitter messages," in *Proc. IEEE Spoken Lang. Technol. Workshop*, Dec. 2010, pp. 79–84.
- [10] L. Barbosa and J. Feng, "Robust sentiment detection on Twitter from biased and noisy data," in *Proc. Coling, Posters*, 2010, pp. 36–44.
- [11] M. Dinsoreanu and A. Bacu, "Unsupervised Twitter sentiment classification," in *Proc. Int. Conf. Knowl. Manage. Inf. Sharing*, 2014, pp. 151–160.
- [12] L. Zhang, R. Ghosh, and M. Dekhil, "Combining lexicon-based and learning-based methods for Twitter sentiment analysis," HP Laboratories, Palo Alto, CA, USA, Tech. Rep. HPL-2011, 2011, pp. 1–8, vol. 89.
- [13] J. Macqueen, "Some methods for classification and analysis of MultiVariate observations," in *Proc. 5th Berkeley Symp. Math. Statist. Probab.*, 2015, pp. 281–297.
- [14] R. T. Ng and J. Han, "Efficient and effective clustering methods for spatial data mining," in *Proc. 20th Int. Conf. Very Large Data Bases*. San Mateo, CA, USA: Morgan Kaufmann, 1994, pp. 144–155.
- [15] S.-S. Lin, S.-L. Shen, A. Zhou, and Y.-S. Xu, "Risk assessment and management of excavation system based on fuzzy set theory and machine learning methods," *Autom. Construct.*, vol. 122, Feb. 2021, Art. no. 103490.

- [16] A. R. Camero, A. Lestido-Cardama, P. V. Loureiro, L. Barbosa-Pereira, A. R. B. de Quirós, and R. Sendón, "Presence of perfluoroalkyl and polyfluoroalkyl substances (PFAS) in food contact materials (FCM) and its migration to food," *Foods*, vol. 10, no. 7, p. 1443, Jun. 2021.
- [17] J. Chen, H. Zhang, D. Pi, M. Kantardzic, Q. Yin, and X. Liu, "A weight possibilistic fuzzy C-means clustering algorithm," *Sci. Program.*, vol. 2021, pp. 1–10, Jun. 2021.
- [18] M. Naderipour, M. H. F. Zarandi, and S. Bastani, "A fuzzy cluster-validity index based on the topology structure and node attribute in complex networks," *Expert Syst. Appl.*, vol. 187, Jan. 2022, Art. no. 115913.
- [19] Y. Xiong, S. Chen, H. Qin, H. Cao, Y. Shen, X. Wang, Q. Chen, J. Yan, and B. Tang, "Distributed representation and one-hot representation fusion with gated network for clinical semantic textual similarity," *BMC Med. Informat. Decis. Making*, vol. 20, no. S1, pp. 1–7, Apr. 2020.
- [20] A. Nath, D. Kandar, and R. Gupta, "An efficient cross-lingual BERT model for text classification and named entity extraction in multilingual dataset," *Int. J. Sci. Res. Comput. Sci., Eng. Inf. Technol.*, vol. 2021, pp. 280–286, May 2021.
- [21] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy C-means clustering algorithm," *Comput. Geosci.*, vol. 10, nos. 2–3, pp. 191–203, 1984.
- [22] T. Lei, X. Jia, Y. Zhang, L. He, H. Meng, and A. K. Nandi, "Significantly fast and robust fuzzy C-Means clustering algorithm based on morphological reconstruction and membership filtering," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 5, pp. 3027–3041, Oct. 2018.
- [23] M. Jain, V. Saihpal, N. Singh, and S. B. Singh, "An overview of variants and advancements of PSO algorithm," *Appl. Sci.*, vol. 12, no. 17, p. 8392, Aug. 2022.
- [24] D. C. Luckham, "The power of events: An introduction to complex event processing in distributed enterprise systems," in *Proc. IEEE IPCCC*, Sep. 2003, pp. 335–342.
- [25] A. K. Maddirala and R. A. Shaik, "Separation of sources from single-channel EEG signals using independent component analysis," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 2, pp. 382–393, Feb. 2018.
- [26] D. D. Houghton, *Global Climate Change, Impact on the Upper Midwest*. Madison, WI, USA: Atmospheric&Oceanic Sciences UW-Madison, Mar. 2007.



**CHAO HAN** was born in October 1981. He received the bachelor's degree in law from Yanshan University, the M.A. degree from the Communication University of China, in July 2009, and the master's degree in education from the Dhurakij Pundit University of Thailand, in January 2020. He is currently pursuing the Ph.D. degree in art with Krirk University, Thailand.

From June 2007 to September 2010, he was a Newscaster, the Host, and the Reporter at Southeast Satellite TV, Fujian, China. In September 2010, he was with the Drama School, Yunnan Arts University, and worked as a Professional Teacher of broadcasting and hosting. During the period of work, he has published dozens of articles and conducted in-depth research on news media, drama, and film.



**XUEZHI SUN** is currently pursuing the Ph.D. degree.

He was an Associate Professor. He is the Dean of the Beijing Zhongke Saibo Standardization Institute and a Ph.D. Supervisor of Krirk University, Thailand. He has published dozens of articles and more than ten books. His research interests include education, art, and business administration.

• • •