

Research Article

Neural Network: An Improved FCM for Multimodal Cultural Data Analysis

Yan Wang , Di Liu , Lingling Tian , and Aiping Tan 

College of Information, Liaoning University, Shenyang 110036, China

Correspondence should be addressed to Aiping Tan; aipingtan@lnu.edu.cn

Received 14 July 2021; Revised 10 September 2021; Accepted 25 November 2021; Published 15 December 2021

Academic Editor: Mamoun Alazab

Copyright © 2021 Yan Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the development of cloud computing, big data, and artificial intelligence (AI) technology, there is a growing interest in “cultural analysis.” Cultural analysis requires different types of data such as texts, pictures, and videos. The richness and differences of resources in the cultural field lead to diverse modalities of cultural data. Traditional text analysis methods can no longer meet the data analysis needs of current multimedia cultural resources. This article starts from cultural data’s feature information to solve the heterogeneity problem faced by massive multimodal cultural data analysis. It analyzes it from geography, time, art, and thematic character, classified and aggregated to form a multimodal cultural feature information matrix. The corresponding correlation measurement methods for different matrices from the above dimensions are proposed, solved in turn, and substituted into the optimized training back propagation (BP) neural network to obtain the final correlation degree. The improved fuzzy C-means (FCM) clustering algorithm is used to aggregate the high correlation cultural data based on the degree. The algorithm proposed in this study is compared with the existing algorithm. The experimental results show that the optimized BP neural network is at least 58% more accurate than the current method for calculating different matrices’ correlation degrees. In terms of accuracy, the improved fuzzy C-means algorithm effectively reduces the random interference in the selection of the initial clustering center, which is significantly higher than other clustering algorithms.

1. Introduction

The term “cultural analysis” was proposed by Lev Manovich in 2007 and referred to the analysis of contemporary and historical cultural data using computing, visualization, and extensive data methods [1]. Pierre Bourdieu’s historical survey of the cultural consumption of Parisians in the mid-twentieth century in “Differentiation” opened up a way of studying culture and aesthetics with a large number of data sets, and then, Franco Moretti’s “Charts, Maps, Trees: An Abstract Model of Literary History” demonstrates the benefits of large-scale analysis of cultural data [2]. In recent years, researchers in various fields have gradually used the term “cultural analysis.” The Journal of Cultural Analysis was published in 2016. Many universities have gradually developed research plans in cultural analysis, such as the MIT monograph “Cultural Analysis,” published in the fall of 2020. Traditional humanity analysis focuses on studying text

data. Still, cultural research also focuses on massive multimedia data sets, including digital visual products and contemporary visual and interactive media [3]. So far, cultural analysis techniques have been applied to movies, animations, video games, comics, magazines, books, and other printed publications, works of art, photographs, and various other media contents. In the past few years, the number of digital and publicly available art collections has been increasing. With the emergence of such a large number of digital artworks, facing the massive cultural digital resources, how to explore its rich content to the greatest extent has become an enormous challenge in the field of artistic research.

When researching a work of art, we make some inferences about the job. For example, in addition to understanding the subject, we may also classify it according to period, style, and artist. So, can computer algorithms “understand” a work of art as profoundly as humans and quickly

type it? Babak Saleh and Ahmed Elgarmar of the Art and Artificial Intelligence Laboratory of Rutgers University have compiled a database of thousands of paintings in the past six centuries and proposed knowledge based on the field of art-historical interpretation. The optimized method of measuring the similarity between paintings enables the algorithm's accuracy to classify images by style to 60% [4, 5]. However, this algorithm's application scenarios are more limited and can only be classified for a single type of data such as painting and are not suitable for multimodal cultural data. Because only three indicators of style, genre, and artist are considered in the classification, the algorithm's classification effect is not satisfactory.

Due to the fuzzy relationship between the similarities of cultural data, the effect of using a hard clustering algorithm is not ideal. In view of the current situation that the existing multimodal clustering and fusion algorithms cannot make full use of cultural data information, starting from the cultural feature information in the data, this study proposes a fuzzy fusion strategy multimodal cultural data based on BP neural network and FCM. The main contributions are as follows:

- (1) The extracted multimodal cultural data's feature information is classified and summarized from the dimensions of geography, time, art, and thematic characters to form a multimodal cultural feature information matrix (MCM).
- (2) The correlation measurement standards between multimodal cultural data from multiple dimensions are defined, and the data set is substituted into the optimized BP neural network for training to obtain the final correlation degree.
- (3) The fuzzy C-means clustering strategy, which aggregates the strongly correlated cultural data, solves the problem of data disorder and provides a better fusion strategy for the fusion of multimodal cultural data.

The organization of the article is as follows. Section 2 briefly introduces related work. The third and fourth parts present the multimodal cultural data fusion strategy. Section 3 introduces the establishment of multimodal cultural feature information matrix data set, the quantification method of cultural relevance between MCMs, and the optimization and adjustment of the BP neural network. Section 4 introduces the improved fuzzy C-means clustering algorithm. The experiment in Section 5 proves the effectiveness of the fusion strategy. Section 6 summarizes the full text.

2. Related Work

To understand the world, humans must distinguish different things and recognize the similarities between things, while for computers, they need to use unsupervised learning methods to determine things. From the Code of Hammurabi to the Mona Lisa and from the Beatles to the Forrest Gump, cultural data sources are vast and diverse. They are carried in different modalities such as text, pictures, audio, and video.

The integration of data between other modalities is a multimodal data fusion problem, which integrates the information extracted from different modalities into a stable multimodal representation. In recent years, scholars at home and abroad have been widely concerned about multimodal fusion. A variety of multimodal fusion methods based on model-independent, graph models, and neural networks have been proposed. Scholars such as Xia proposed a hash algorithm to integrate multimodal visual features into binary codes and calculate image kernel features [6]. However, it is difficult for this algorithm to fuse other modal data's feature information except for image and video. Jiang et al. assumed that data of different modalities share the potential structure and learn this potential through the connection between multimodal data sharing structure and mining of the interaction between the structure and category information, to be applied to classification tasks [7]. However, prior knowledge is required as a reference in classification, and the computational complexity is relatively high. Onan et al. proposed a term weighted neural language model to process social media data and achieved good results [8, 9]. The application of the above results fully considers the relationship between social media data, but does not consider the cultural feature. However, it still provides a great reference for the characteristic analysis of cultural data. Jin et al. proposed a recurrent neural network with an attention mechanism, which uses the LSTM network to fuse the text and social context features. Then, it uses the attention mechanism to connect it with image features to perform end-to-end rumor prediction [10]. The fusion mechanism relies on the correspondence between text and image, which is difficult to handle in the face of disordered multimodal data. Traditional clustering has successfully solved the clustering problem of low-dimensional data. However, due to data complexity in practical applications, existing algorithms often fail when dealing with many issues, especially for high-dimensional data and large-scale data. Because traditional clustering methods mainly encounter two problems when clustering high-dimensional data sets, one is that there are a large number of irrelevant attributes in the high-dimensional data set, making the possibility of clusters in all dimensions almost zero [11, 12]; the second is that the data in the high-dimensional space should be sparsely distributed in the lower-dimensional area, and the distance between the data is virtually equal [13]. However, the traditional clustering method is based on the distance to cluster [14, 15], so it is impossible to construct groups based on distance in high-dimensional space.

The disordered characteristics of massive multimodal cultural data have become a challenge for mining its common feature themes [16, 17]. To solve this problem, we start with the data's feature information and classify its cultural feature information from multiple dimensions to form a multimodal cultural feature information matrix containing various distinct details. The relevance of geography, time, art, and thematic character from each dimension is calculated using their respective relevance calculation formulas for MCM, and the final bearing is manually marked through experience to form the MCM

relevance data set and test set [18–20]. They are substituting the data set into the BP neural network for training. The correlation degree of the neural network output is compared with the correlation degree manually calibrated in the test set by adjusting and optimizing parameters such as the number of iterations and learning rate [21, 22]. An accuracy rate of more than 96% is achieved. Therefore, we can regard the BP neural network's output as the final cultural relevance between MCMs, forming a measure of relevance between multimodal cultural data.

Taking the degree of association between MCMs as the basis for clustering, the fuzzy C-means clustering algorithm, which is widely used in practical applications, especially in the field of multimodal learning [23], is used to cluster all MCM. The fuzzy C-means clustering algorithm is more sensitive to the selection of initial clustering centers [24]. To ensure that the clustering results meet the same category's tightness and ensure the separation of each class, the choice of initial clustering centers is restricted here. The Euclidean distance expresses the membership degree of the traditional FCM [25]. This article uses the cultural relevance output of the trained BP neural network as the membership degree of the FCM, restricts the degree of membership between the initial cluster centers, and requires each other to belong. The degree is greater than the constraint thresholds, and the reverse analysis is performed here through the experimental results. The experimental results prove that the improved C-means clustering strategy based on BP neural network can gather highly correlated multimodal cultural data and solve massive multimodal data disorder [26]. It is compared with the classic clustering algorithm, which proves the superiority of this strategy. Figure 1 clearly shows the proposed method.

3. Relevance Degree of the Multimodal Cultural Characteristic Information Matrix

3.1. Establish a Multimodal Cultural Characteristic Information Matrix. The four most representative features of multimodal cultural data, geographical features, time features, art features, and thematic character features, are used to describe the characteristics of cultural data and obtain the corresponding feature information to express the classified multimodal cultural data in a unified form, as shown in the following matrix, where $l_1, l_2, l_3, \dots, l_n$ represents the geographical feature information of the multimodal cultural data, $t_1, t_2, t_3, \dots, t_n$ represents time feature information, $a_1, a_2, a_3, \dots, a_n$ represents art feature information, and $s_1, s_2, s_3, \dots, s_n$ represents thematic character feature information.

$$\begin{bmatrix} l_1 & l_2 & l_3 & \cdots & l_n \\ t_1 & t_2 & t_3 & \cdots & t_n \\ a_1 & a_2 & a_3 & \cdots & a_n \\ s_1 & s_2 & s_3 & \cdots & s_n \end{bmatrix}. \quad (1)$$

The geographic feature information refers to the country, city, etc., related to the data, such as the location where the artwork was created and the site of the existing museum. The time feature information is the time related to the data, such as the sculpture's creation time, the creation of the text, and the historical period reflected. Art feature information includes macro-art forms and micro-art forms, among which macro-art forms include nine categories, such as painting and sculpture. The feature information of the thematic character consists of two parts: cultural subject and character. The concept of the cultural subject was first proposed by the American Scholar M.E-Opler [27]. He believes that for any ethnic culture, its behavior is guided by value. For multimodal cultural data, the scope of cultural subjects is more extensive, such as religion, ethnic customs, and cultural symbols; character feature includes the author and the people related to the cultural data. The above four types of feature information to form a multimodal cultural feature information matrix are summarized.

The example in Figure 2 shows the MCM of picture-type data. The geographical feature information is the current location of the work, the Louvre in France, and the author's country, Italy. The time feature information is the work created and completed in 1503 and 1507 AD. The art feature information is paintings and oil paintings, and the thematic character feature information is the Renaissance, Mona Lisa, and author Leonardo da Vinci. The MCM classifies cultural feature information and filters noise information that has nothing to do with the cultural field, which plays an important supporting role for subsequent correlation measurement and clustering fusion.

3.2. Association Quantification. The multimodal cultural feature information matrix contains the geographical feature information, time feature information, art feature information, and thematic character feature information. Starting from the four types of features, we get the geographical relevance r_l , time relevance r_t , art relevance r_a , and thematic character relevance r_s . The four relevance degrees are weighted and summarized to obtain the relevance r between the MCMs.

Geographical relevance: according to the first and second geography laws, the correlation between features is related to distance, and spatial isolation causes differences between components [28, 29]. Generally speaking, the closer the space, the more significant the correlation between features; the farther the distance, the greater the dissimilarity between features. Google Geocoding API [30] is used to convert the geographic feature information in MCM to geographic coordinates. For example, convert Louvre to (48.860611, 2.337644), for MCM₁ and MCM₂ use the latitude and longitude coordinates to find the average distance between the two geographic feature information, and convert the average distance into the geographic association degree r_l between [0,1]. Here, the distance between the longitude and

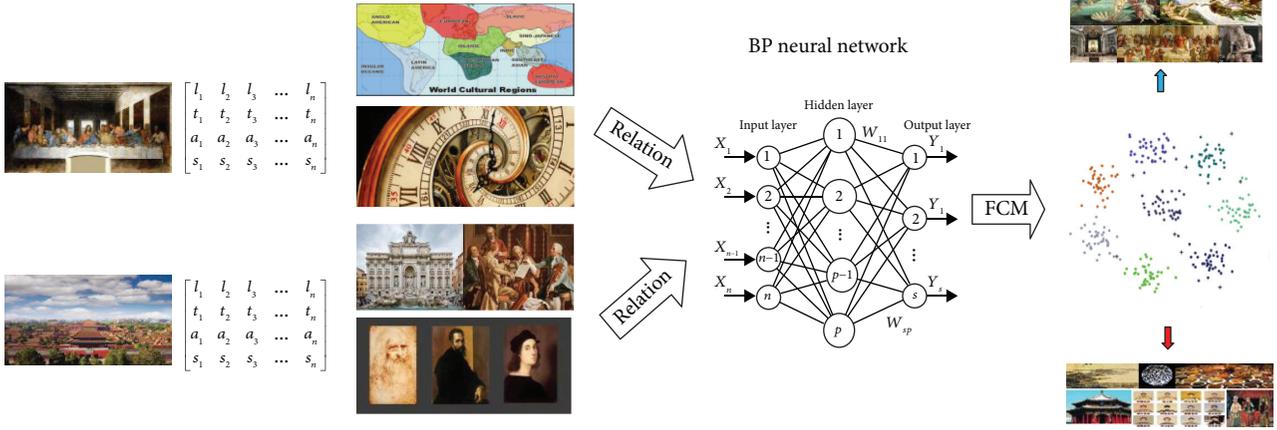


FIGURE 1: Multimodal cultural data clustering strategy based on BP neural network.

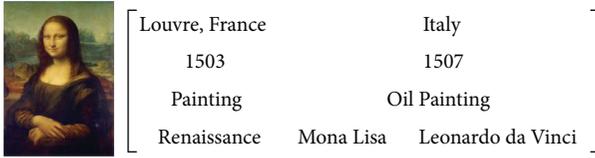


FIGURE 2: Multimodal cultural feature information matrix (MCM).

latitude between MCMs is calculated by the Haversine formula:

$$\text{hav}(\theta) = \sin^2\left(\frac{\theta}{2}\right) = \frac{(1 - \cos(\theta))}{2},$$

$$\text{haver}\left(\frac{d}{R}\right) = \text{hav}(\varphi_2 - \varphi_1) + \cos(\varphi_1)\cos(\varphi_2)\text{hav}(\Delta\lambda), \quad (2)$$

where R is the Earth's radius, where the average value is 6371 km, φ_1 and φ_2 represent the latitude, and $\Delta\lambda$ represents the difference in longitude. The circumference of the Earth's equator is 40076 kilometers, and the rim of the north and south poles is 39,900 kilometers. Here, $L=20000$ km. Therefore, the calculation formula for the degree of geographic association r_1 is as follows:

$$r_1 = \frac{L-d}{L}. \quad (3)$$

In the "World Culture Report on Cultural Diversity, Conflict, and Coexistence of Cultures (2000)" compiled by UNESCO, the world's countries are divided into eight regions [31]. According to Wikipedia, there are 197 countries in the world. To avoid factors such as the size of the country being too large, the degree of geographical relevance is revised: if it is in the same area, the bearing is $8 r_1$; if it is in the same country, the relevance is $16 r_1$.

Time relevance: in the multimodal cultural feature information matrix MCM, $T = [t_1 \ t_2 \ t_3 \ \dots \ t_n]$ is its time feature information matrix and $T' = t_1 \ t_2 \ t_3 \ \dots \ t_n$ is its time feature sequence. For MCM₁ and MCM₂, the Pearson correlation coefficient is introduced to determine

the degree of time correlation. The time correlation degree r_t between MCMs reflects the correlation degree of its time feature information, and its calculation formula is as follows:

$$r_t = \frac{\sum_{i=1}^n (T_{1_i} - \bar{T}_2)(T_{2_i} - \bar{T}_1)}{\sqrt{\sum_{i=1}^n (T_{1_i} - \bar{T}_2)^2} \sqrt{\sum_{i=1}^n (T_{2_i} - \bar{T}_1)^2}} \quad (4)$$

Art relevance: the art relevance r_a is determined by the macro-art category and the micro-art category. On the macro-level, it is divided into nine categories, including the eight traditional arts and the ninth art. The eight traditional arts mainly refer to social ideologies that use images to reflect reality but are more typical than reality, including eight categories: literature, music, dance, painting, sculpture, drama, architecture, and film. With the development of science and technology, art expression is also constantly developing, such as TV art, TV series, comic books, and video games. These art forms are difficult to be classified into the above eight states, so they are collectively referred to as the ninth art.

Take painting as an example at the micro-level: from the geographical point of view, painting can be divided into Eastern painting and Western painting; from the perspective of tools and materials, painting can be divided into ink painting, oil painting, mural painting, printmaking, watercolor painting, gouache painting, etc.; from the subject matter content, painting can be divided into figure paintings, landscape paintings, still life paintings, animal paintings, etc.; and from the perspective of the form of works, painting can be divided into murals, new year pictures, comic strips, comics, propaganda paintings, oil paintings, blow paintings, illustrations, etc. For MCM₁ and MCM₂, by judging the macro- and micro-art categories successively, the art relevance r_a is obtained. The calculation formula is as follows. For example, when the comparative art information is of the same macro, but different micro, the value of r_a is 0.7.

$$r_a = \begin{cases} 1, & \text{(Same macro, same micro),} \\ 0.7, & \text{(Same macro, different micro),} \\ 0, & \text{(Different macro, different micro).} \end{cases} \quad (5)$$

Thematic character relevance: for multimodal cultural data, the thematic character feature information includes two parts: the subject and the character, both of which are based on nouns. The relevance of themes and characters is solved by comparing the number of the same nouns that appear in the feature information. The calculation formula is as follows, where s represents the number of the same nouns and n_1 and n_2 are the number of thematic character feature information in MCM_1 and MCM_2 , respectively.

$$r_s = \sqrt{\frac{s + (1/n_1 + n_2)^2}{s + 1/n_1 + n_2}}. \quad (6)$$

After obtaining the geographical relevance r_l , time relevance r_t , art relevance r_a , and thematic character relevance r_s , it is necessary to perform a weighted calculation to find the final cultural relevance [32]. Traditional weighting methods such as the AHP analytic hierarchy process rely more on expert scoring; in contrast, the CRITIC weighting method is objective. Still, it is only applicable to the situation where the data sample is fixed, and the model needs to be updated continuously in the face of scalable data. Therefore, the time complexity is relatively large. BP neural network does not need to determine the mapping relationship's mathematical equations between input and output in advance. It only learns a specific rule through its training and obtains the result closest to the expected output value when the input value is given. Therefore, the final determination of the correlation degree is cross-cutting. The BP neural network is given.

3.3. BP Neural Network Training

3.3.1. Establishment of Data Set. Before performing BP neural network training, to quantitatively analyze and compare the performance of different clustering algorithms, it is necessary to experiment with real cultural data sets, including creating the training data set required for training and the test data set needed for verification. However, as the clustering research on cultural data is still in its infancy, the existing mainstream data sets for the cultural field are less, which cannot meet the requirements of this experiment. Based on this, we call the world's major museum website API interface to write Web crawler to construct the experimental data set. As shown in Figure 3, here, a large amount of representative multimodal cultural data is collected when the above two data sets are established, including about 28644 works of art from 30 museums in the world, 700 representative sets of data and their corresponding cultural feature information are manually selected as the source of experimental data for BP neural network, and each set of data includes two multimodal cultural data and their corresponding multimodal feature information matrix. 700 sets of data are divided into two parts, 500 sets are used as the training data set of the BP neural network and 200 sets are used as the test data set.

For all kinds of multimodal cultural data collected, we employ a third party to label the feature information of the data set manually. In the process of labeling, we asked four

taggers to conduct a kappa test on the consistency of labeling results. The test results are shown in Table 1. The principle of complementarity is used for the result of labeling, and the principle of high voting is used for the contradiction of labeling information.

The quantification methods of the above four correlation degrees have been given. The corresponding four correlation degrees can be calculated by computer programming, and the final correlation degrees of the training and test datasets are marked manually. Here, the absolute relevance is divided into 20 levels, where 1 represents the most significant relevance and 20 illustrates the lowest relevance. Taking "Mona Lisa" and "Genesis" as examples, their geographic relevance r_l , time relevance r_t , art relevance r_a , and thematic character relevance r_s are all calculated by the computer based on the calculation formula proposed in the previous article, respectively, 0.445032, 0.9325, 0.7, and 0.938616. The final degree of relevance is scored artificially. Because they are all European cultural works during the Renaissance, the art forms are all paintings, and Leonardo and Michelangelo are also called the Three Great Renaissance, so the final degree of relevance is determined as 3. "Rachmaninoff No. 2 Piano Concerto" is very similar to "Rachmaninoff No. 3 Piano Concerto," and its correlation is the highest value of 1; while the difference between the movie "The Godfather" and the Pyramid of Khufu is enormous, hence, its final relevance value is 20.

3.3.2. BP Neural Network Training. After the data set and test set are established, BP neural network is used for training and verification. BP neural network is a concept proposed by scientists led by Rumelhart and McClelland in 1986. It is a multilayer feedforward neural network trained according to the error back propagation algorithm and is the most widely used neural network. BP neural network's nonlinear mapping ability and strong self-learning and self-adaptive ability enable it to automatically extract the four cultural correlation features' inherent correlation feature through learning to obtain the final correlation degree.

When training a neural network, many parameters need to be adjusted. Here, learning rate, epoch (training times), and mini-batch size are used to adjust the object parameters. In the process of changing the parameters of the neural network, accuracy is the primary goal. For training neural networks, the initial learning rate significantly impacts accuracy, so it is selected as the primary parameter for adjustment. In the process of adjusting it, the training data set selects all the 500 sets of training data established above, and the rest of the parameters are in the default state. All 200 test data sets are chosen to verify the accuracy.

As shown in Figure 4, the abscissa is the learning rate, and the ordinate is the test accuracy. It can be seen that as the learning rate increases from 0.00001 to 0.1, the accuracy of the test also increases linearly, and the highest accuracy of 0.925 is achieved near 0.1. Then, as the learning rate increases, the accuracy rate drops slightly, so 0.1 is selected as the neural network's learning rate. After determining the learning rate, the learning rate is kept at 0.1 epoch is adjusted.



FIGURE 3: Multimodal cultural data set.

TABLE 1: Consistency examination of labeling cultural data information.

Data set	Response to (2,3,4)	Response to (1,3,4)	Response to (1,2,4)	Response to (1,2,3)
Training data set (500)	0.837	0.826	0.784	0.764
Test data set (200)	0.796	0.785	0.812	0.767

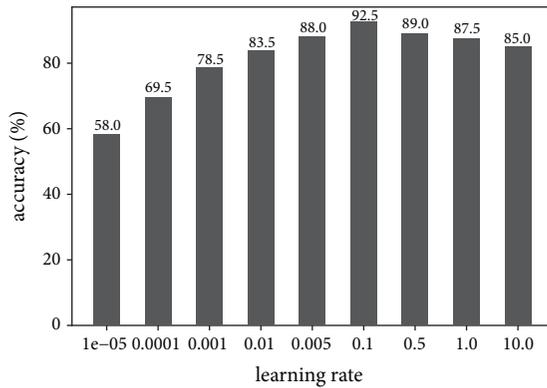


FIGURE 4: Effect of learning rate on accuracy.

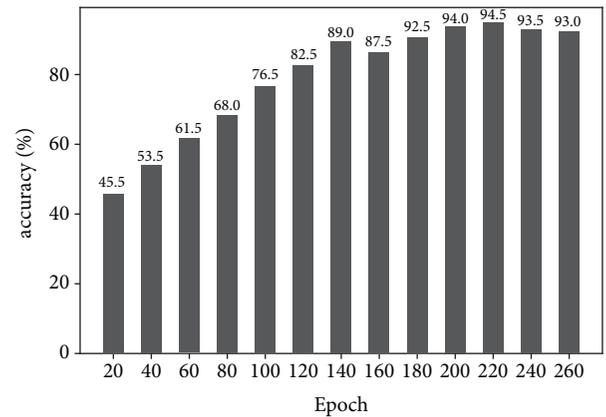


FIGURE 5: Effect of epoch on accuracy.

As shown in Figure 5, the abscissa is the epoch, and the ordinate is the test accuracy. As epoch starts to grow from 20, the accuracy rate has also improved significantly. When epoch is 220, an accuracy rate of 94.5% is achieved. When epoch reaches 220, the accuracy rate no longer improves or even begins to decrease, and overfitting occurs, so 220 is selected as the value of epoch.

As shown in Figure 6, the abscissa is the mini-batch size, and the ordinate is the test accuracy. It can be seen from the figure that the influence of batch size on test accuracy is not apparent. When the mini-batch size is between 50 and 125, the accuracy rate exceeds 95%, and when the mini-batch size is 100, the accuracy rate is 96.5%.

In summary, when the learning rate is set to 0.1, epoch and mini-batch size are set to 100, the highest accuracy rate is 96.5%, which is very close to the correlation degree given manually, and it will be slightly in the process of manually giving the correlation degree. With some subjective colors,

the BP neural network's correlation degree output can be used as a clustering standard to realize the correlation quantification between two cultural multimodal data.

4. Fuzzy C-Means Clustering

4.1. Initial Cluster Center Constraints. After obtaining the correlation degree of the two multimodal cultural feature information matrices through the BP neural network, it is necessary to cluster the entire matrix. The fuzzy C-means algorithm (FCM) is a fuzzy clustering algorithm based on the objective function. It is mainly used for data clustering analysis. It is widely used because of its efficiency and simplicity. Here, we introduce its clustering method, to cluster the multimodal cultural feature information matrix. The classic FCM clustering algorithm has two problems: the excessive dependence of the algorithm on the initial clustering center; the other is that the algorithm needs to know

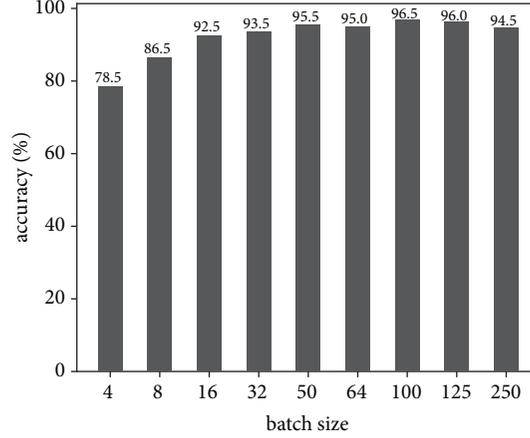


FIGURE 6: Effect of mini-batch size on accuracy.

the actual number of clusters in advance. Experts can give the number of groups here in the cultural field by analyzing specific application scenarios, so the initial cluster centers' randomness needs to be improved. Here, the fuzzy thresholds γ are set, and the correlation between the initial cluster centers is restricted.

The principle of selecting the initial clustering center is to distance the various initial clustering centers greater than the set thresholds γ . In the following clustering algorithm, the clustering center can be analyzed in multiple feasibility domains. The calculation avoids the situation that the algorithm is easy to converge to a local minimum when FCM randomly calculates the initial cluster centers. The fuzzy thresholds γ are difficult to obtain through estimation or formula, and it needs to be adjusted through a lot of experiments.

4.2. Fuzzy C-Means Clustering. Compared with the traditional method, the improved fuzzy C-means algorithm proposed in this study increases the initial clustering center's restriction. The algorithm steps are as follows: first, k multimodal cultural feature information matrix $MCMC_i (i = 1, 2, \dots, k)$ is randomly selected as clusters

from the entire multimodal cultural feature information matrix $MCM' = \{MCM_1, MCM_2, \dots, MCM_n\}$. MCMC here is to distinguish the cluster center matrix from the rest of the matrix, and the cluster center matrix is recorded as the correlation degree of each cluster center should be greater than the fuzzy thresholds γ . r_{MCMC_i, MCM_j} is the membership function of the j th sample corresponding to the i th category, and the clustering loss function of the membership function is expressed as follows:

$$J_f = \sum_{i=1}^c \sum_{j=1}^n r_{MCMC_i, MCM_j}^b \|MCM_j - MCMC_i\|^2, \quad (7)$$

where $\sum_{i=1}^c r_{MCMC_i, MCM_j}^b$ refers to the weighted index, also known as the smoothing factor, which controls the degree of sharing between fuzzy classes. The best value of b is not yet the same standard, and the value is 2 under normal circumstances. $\|MCM_j - MCMC_i\|^2$ indicates the degree of relevance (distance) between the multimodal cultural feature information matrix to be clustered and the cluster center matrix. Let the partial derivative of J_f to r_{MCMC_i, MCM_j} and $\|MCM_j - MCMC_i\|^2$ be 0, and the necessary condition for the minimum formula value is obtained (7).

$$MCMC_i = \frac{\sum_{j=1}^n r_{MCMC_i, MCM_j}^b MCM_j}{\sum_{j=1}^n r_{MCMC_i, MCM_j}^b}, \quad (8)$$

$$r_{MCMC_i, MCM_j} = \frac{1}{\sum_{k=1}^c \|MCM_j - MCMC_i\|^2 / \|MCM_j - MCMC_k\|^{2^{1/b-1}}}. \quad (9)$$

The iterative method is used to solve (8) and (9), the membership degree r_{MCMC_i, MCM_j} and clustering center $MCMC_i$ are updated until it stops when the convergence

condition is met. $\max MCM_j, MCMC_i ((r_{MCMC_i, MCM_j}^{k+1} - (r_{MCMC_i, MCM_j}^k)) < \epsilon$, where ϵ is the termination criterion between 0 and 1, and k is the iterative step. Converged to the

- (1) Initialize MCM, input c , b , and ε .
- (2) Set fuzzy thresholds γ , and the MCMC is constrained by fuzzy thresholds.
- (3) At k -step: calculate the centers' vectors $MCMC(k)$ with $MCMC_i = \sum_{j=1}^n r_{MCMC_i, MCM_j}^b MCM_j / \sum_{j=1}^n r_{MCMC_i, MCM_j}^b$.
- (4) Update $MCMC(k)$, $MCMC(k+1) rr_{MCMC_i, MCM_j} = 1 / \sum_{k=1}^c \|MCM_j - MCMC_i\|^2 / \|MCM_j - MCMC_k\|^{2^{1/b-1}}$.
- (5) If $\max MCM_j, MCMC_i ((r_{MCMC_i, MCM_j}^{k+1}) - (r_{MCMC_i, MCM_j}^k)) < \varepsilon$ then stop; otherwise, return to step 3.

ALGORITHM 1: Improved fuzzy C-means.

local minimum of J_f , the disordered multimodal cultural feature information matrix completes the clustering through the mutual relationship.

The pseudo-code of the improved fuzzy C-means 1 that constrains the initial cluster centers is as follows:

5. Experiment

The purpose of the experiment is to test the effectiveness of the data clustering fusion technology proposed above under the environment of simulating a large number of multimodal cultural data. The experimental environment setting for the above BP neural network training and the following clustering effect is a single PC; the CPU configuration is Intel Core i7-8700 3.20 GHz, 32GB memory, and programming with MATLAB R2017b. The data used for the experiment use the multimodal cultural data processed above. To verify the effect of the clustering algorithm proposed in this study, three clustering algorithms are compared, namely the K-means algorithm [33], density-based spatial clustering of applications with noise (DBSCAN) [34] algorithm, and FCM algorithm [35].

FCM algorithm needs to determine the number of clusters in advance and has the problem of randomness of the initial cluster center. The experiment consists of three parts: Experiment 1 determines the optimal number of clusters, experiment 2 analyzes its influence on the clustering results by adjusting the fuzzy thresholds γ to determine the best value; and experiment 3 is carried out with traditional clustering algorithms based on determining the fuzzy thresholds γ in experiment 2.

5.1. The Influence of Clustering Number. To achieve a better clustering effect, it is necessary to determine the number of clusters. MSE value is used as an objective indicator to evaluate the effect of clustering [36]. In statistics, MSE calculates the squares of the errors between the fitted data and the corresponding points of the original data. MSE can evaluate the degree of change in the data. The smaller the value, the better the accuracy of the prediction model to describe the experimental data. The data used in the experiment are the multimodal cultural data, and its feature information matrix is used in training and testing of the BP neural network above. To ensure the reliability of the experimental results, five groups of experiments were carried out under the same test environment.

As shown in Figure 7, the abscissa is the clustering number, and the ordinate is the MSE. MSE value changes

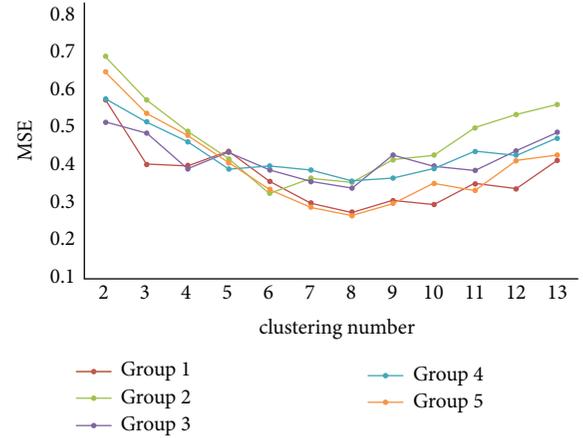


FIGURE 7: Effect of clustering number.

with the number of cluster centers. When the clustering number is small, the internal discrimination of each class is low. When the clustering number is too large, the randomness of the selection of cluster centers will increase the instability of the internal clusters of each class. The above two cases will reduce the accuracy of clustering results. To select the appropriate clustering number, we have carried out many experiments. We randomly divided the data needed for the experiment into five groups, and each group changed the clustering number and carried out five groups of comparative experiments. As shown in Figure 7, it can be seen that when it is taken 8, the clustering effect is the best, so the clustering number is 8.

5.2. The Influence of γ on Clustering Effect. Experiment 2 compares the influence of different fuzzy thresholds γ on the clustering results. As shown in Figure 8, (a), (b), and (c) are the experimental results of $\gamma = 7$, $\gamma = 8$, and $\gamma = 9$, respectively. Different γ bring different clustering effects. To minimize the impact of the randomness of the initial cluster center on the clustering effect, we determine the best value through experiments and take the MSE value as the basis of selection.

As shown in Figure 9, the abscissa is the fuzzy thresholds, and the ordinate is the MSE. If the fuzzy thresholds are too low, it will lead to low data discrimination in different clusters. If the fuzzy thresholds are too high, it will produce clustering imbalance, affecting the clustering effect. Like the experiment above, to select the best clustering number, we randomly divided the data needed for the experiment into five groups. Each group changed the fuzzy thresholds γ and

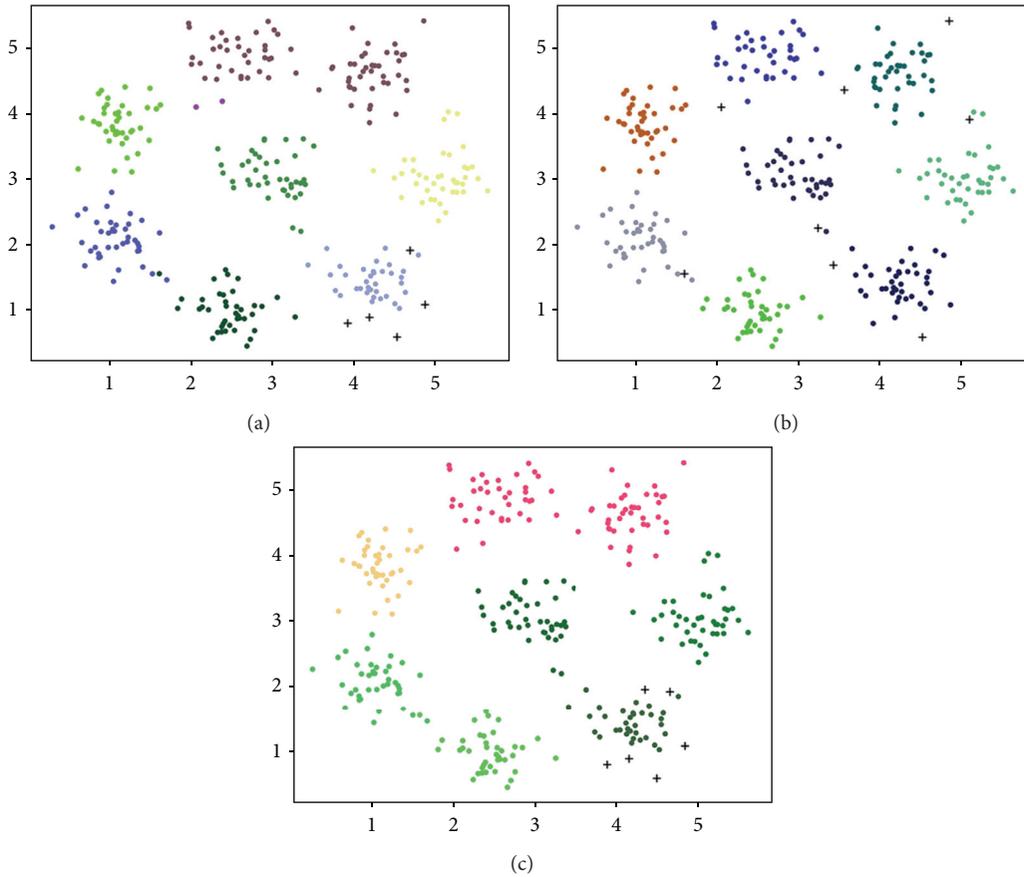


FIGURE 8: Influence of different fuzzy thresholds γ on the clustering results. (a) $\gamma = 7$. (b) $\gamma = 8$. (c) $\gamma = 9$.

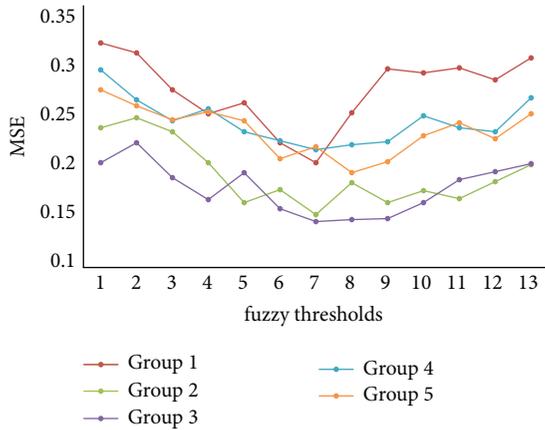


FIGURE 9: Effect of γ .

carried out five groups of comparative experiments. As shown in Figure 9, the selection of fuzzy thresholds γ has the same impact on the accuracy of clustering as the clustering number. It can be seen that when it is taken 7, the clustering effect is the best, so the value of the fuzzy thresholds γ is 7.

5.3. Comparison of Clustering Algorithms. Experiment 3 compared the fuzzy C-means clustering strategy with initial clustering center constraints proposed in this study with

K-means clustering strategy, density-based clustering strategy, and traditional fuzzy C-means clustering strategy. Here, K-means and fuzzy C-means are partitioned clustering algorithms, in which K-means is hard clustering and FCM suggests is soft clustering. The density-based clustering strategy is also widely used in the field of multidimensional data processing, and its classic DBSCAN algorithm is used here. The experimental results of the four clustering algorithms are shown in Figure 10.

As shown in Figure 10, (a), (b), (c), and (d) are the experimental results of K-means, DBSCAN, traditional FCM, and improved FCM, respectively. The K-means clustering algorithm is one of the most classic iterative clustering algorithms. It is widely used because of its simplicity and efficiency [37]. However, the clustering of multimodal cultural data is not sensitive to real-time performance, and the clustering effect is mainly used as the criterion. However, as a hard clustering algorithm, K-means has significant differences in clustering effects and poor robustness. DBSCAN is a density-based spatial clustering algorithm. The algorithm divides the area with sufficient density into clusters (i.e., the number of objects contained in a specific site in the clustering space is not less than a given threshold) and finds clusters of arbitrary shape in a noisy spatial database. It defines clusters as the most extensive collection of points connected by density [38]. It can be seen

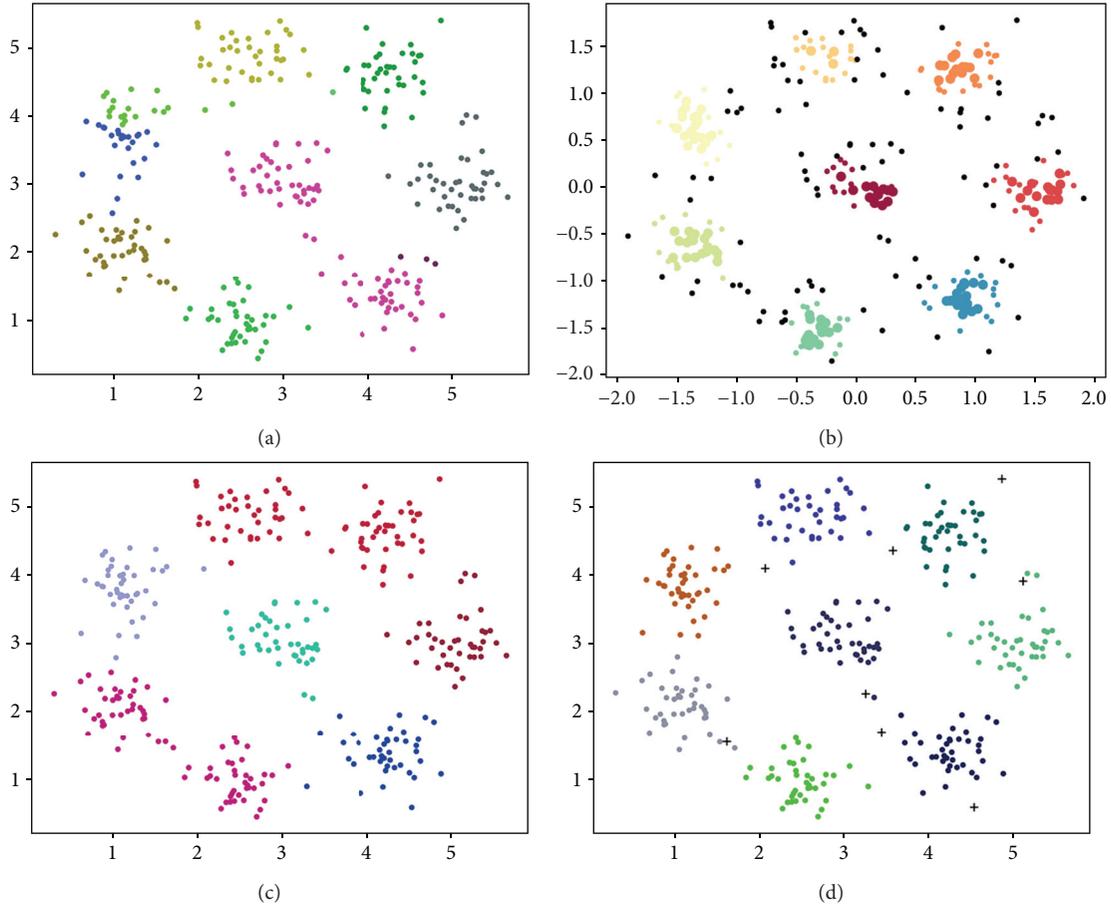


FIGURE 10: K-means, DBSCAN, traditional FCM, and improved FCM. (a) K-means, (b) DBSCAN, (c) traditional FCM, and (d) improved FCM.

TABLE 2: Comparison of clustering experiment results.

Experimental sequence	Algorithm	Number of iterations	MSE
1	K-means	11	0.398678
	DBSCAN	23	0.174943
	Traditional FCM	25	0.165567
	Improved FCM	19	0.096729
2	K-means	8	0.225329
	DBSCAN	21	0.147179
	Traditional FCM	18	0.195696
	Improved FCM	16	0.119679
3	K-means	13	0.262847
	DBSCAN	22	0.145397
	Traditional FCM	19	0.185628
	Improved FCM	22	0.156727
4	K-means	16	0.375326
	DBSCAN	41	0.189642
	Traditional FCM	33	0.235626
	Improved FCM	28	0.204647
5	K-means	6	0.228643
	DBSCAN	20	0.174254
	Traditional FCM	19	0.167687
	Improved FCM	16	0.128036

TABLE 2: Continued.

Experimental sequence	Algorithm	Number of iterations	MSE
6	K-means	11	0.275984
	DBSCAN	29	0.228587
	Traditional FCM	24	0.203187
	Improved FCM	26	0.180997
7	K-means	14	0.435079
	DBSCAN	37	0.330578
	Traditional FCM	32	0.372068
	Improved FCM	29	0.239304
8	K-means	12	0.267865
	DBSCAN	28	0.246841
	Traditional FCM	30	0.253245
	Improved FCM	25	0.192422
9	K-means	16	0.296781
	DBSCAN	27	0.234778
	Traditional FCM	25	0.241987
	Improved FCM	29	0.237179
10	K-means	12	0.443921
	DBSCAN	25	0.246895
	Traditional FCM	23	0.298679
	Improved FCM	25	0.224349

that DBSCAN has achieved a better clustering effect than K-means. The traditional FCM here uses the FCM algorithm that comes with MATLAB, and it can be seen that it has a better clustering effect and robustness than K-means. We can see that the improved FCM has achieved a better clustering effect from the clustering effect graph than the other three clustering algorithms. However, the evaluation from the experimental effect graph is still subjective. Here, the number of iterations and MSE value are used as an objective indicator to evaluate the effect of clustering.

The experimental results obtained by the four clustering algorithms are shown in Table 2. It can be seen that the improved fuzzy C-means algorithm that constrains the initial clustering centers has achieved better clustering results than the traditional fuzzy C-means algorithm. Because K-means is a hard clustering algorithm, the clustering effect is quite different, and the robustness is poor. Density-based DBSCAN is more sensitive to the sparseness of clustering. Because the data set used in the experiment is used for BP neural network training, its sparseness is relatively uniform. Therefore, it has also obtained experimental results close to the improved fuzzy C-means algorithm, but it is challenging to apply when the density is not constant. In summary, the improved fuzzy C-means algorithm proposed in this study, which constrains the initial clustering centers, has achieved the best clustering effect and has good robustness.

6. Conclusions

Aiming at the problem of disorder infusion caused by the richness and diversity of multimodal cultural data, a method for quantifying the correlation between multimodal cultural data based on BP neural network and an improved fuzzy C-means clustering fusion strategy are proposed. In this strategy, the extracted cultural feature information of multimodal data is classified and integrated from multiple

dimensions to form a multimodal cultural feature information matrix. Based on the matrix's feature information, a quantitative method of cultural relevance is proposed for the multimodal cultural feature information matrix. Part of the multimodal cultural data of world-renowned museums is used as a training sample, put into the adjusted BP neural network for training, and output the final degree of relevance. Taking the degree of correlation between the MCMs as the basis for clustering, the fuzzy C-means clustering method, which improves the randomization of the initial cluster centers, is used to cluster and fuse multimodal cultural data. Experiments show that the data fusion algorithm proposed in this study can aggregate highly correlated data and has achieved better clustering results than traditional multidimensional data clustering methods.

Data Availability

The multimodal cultural dataset used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this study.

Acknowledgments

This work was supported in part by National Key R&D Program of China (2019YFB1406002).

References

- [1] T. Blanke, *Cultural Analytics*, Springer International Publishing, New York, NY, USA, 2017.

- [2] F. Moretti and FRANCO Graphs, *Maps, Trees: Abstract Models for a Literary History*, Verso, Brooklyn, NY, USA, 2005.
- [3] L. Manovich, *Cultural Analytics: Visualizing Cultural Patterns in the Era of 'More Media*, DOMUS, Rozzano, Italy, 2009.
- [4] B. Saleh and E. Ahmed, "A unified framework for painting classification," in *Proceedings of the 2015 IEEE International Conference on Data Mining Workshop (ICDMW) (ICDMW '15)*, pp. 1254–1261, IEEE Computer Society, Atlantic, NJ, USA, November 2015.
- [5] B. Saleh, K. Abe, R. S. Arora, and A. Elgammal, "Toward automated discovery of artistic influence," *Multimedia Tools and Applications*, vol. 75, no. 7, pp. 3565–3591, 2016.
- [6] Y. Xia, L. Zhang, Z. Liu, L. Nie, and X. Li, "Weakly supervised multimodal kernel for categorizing aerial photographs," *IEEE Transactions on Image Processing*, vol. 26, no. 8, pp. 3748–3758, 2017.
- [7] X. Jiang, F. Wu, Y. Zhang, S. Tang, W. Lu, and Y. Zhuang, "The classification of multi-modal data with hidden conditional random field," *Pattern Recognition Letters*, vol. 51, pp. 63–69, 2015.
- [8] A. Onan, S. Korukoğlu, and H. Bulut, "Ensemble of keyword extraction methods and classifiers in text classification," *Expert Systems with Applications*, vol. 57, pp. 232–247, 2016.
- [9] A. Onan and M. A. Tocoglu, "A term weighted neural language model and stacked bidirectional LSTM based framework for sarcasm identification," *IEEE Access*, vol. 99, p. 1, 2021.
- [10] Z. Jin, J. Cao, H. Guo, Y. Zhang, and J. Luo, "Multimodal fusion with recurrent neural networks for rumor detection on microblogs," *ACM Multimedia*, pp. 795–816, 2017.
- [11] S. Qian, T. Zhang, C. Xu, and J. Shao, "Multi-modal event topic model for social event analysis," *IEEE Transactions on Multimedia*, vol. 18, no. 2, pp. 233–246, 2016.
- [12] H. Ben-Younes, R. Cadene, M. Cord, and N. Thome, "Mutan: multimodal tucker fusion for visual question answering," in *Proceedings of the IEEE international conference on computer vision*, pp. 2612–2620, Venice, Italy, October 2017.
- [13] L. Peng, Z. Chen, L. T. Yang, L. Zhao, and Q. Zhang, "A privacy-preserving high-order neuro-fuzzy c-means algorithm with cloud computing," *Neurocomputing*, vol. 256, pp. 82–89, 2017.
- [14] Q. Zhang, L. T. Yang, Z. Chen, and F. Xia, "A high-order possibilistic c-means algorithm for clustering incomplete multimedia data," *IEEE Systems Journal*, vol. 11, no. 4, pp. 2160–2169, 2017.
- [15] Q. Zhang, L. T. Yang, Z. Chen, and P. Li, "High-order possibilistic c-means algorithms based on tensor decompositions for big data in IoT[]," *Information Fusion*, vol. 39, pp. 72–80, 2017.
- [16] C. Kang, S. Xiang, S. Liao, C. Xu, and C. Pan, "Learning consistent feature representation for cross-modal multimedia retrieval," *IEEE Transactions on Multimedia*, vol. 17, no. 3, pp. 370–381, 2015.
- [17] F. Nian, B.-K. Bao, T. Li, and C. Xu, "Multi-modal knowledge representation learning via Webly-supervised relationships mining," *ACM Multimedia*, pp. 411–419, 2017.
- [18] S. Ding, C. Su, and J. Yu, "An optimizing BP neural network algorithm based on genetic algorithm," *Artificial Intelligence Review*, vol. 36, no. 2, pp. 153–162, 2011.
- [19] H. Li, J. Sun, Z. Xu, and L. Chen, "Multimodal 2D+3D facial expression recognition with deep fusion convolutional neural network," *IEEE Transactions on Multimedia*, vol. 19, no. 12, pp. 2816–2831, 2017.
- [20] W. Wu, J. Wang, M. Cheng, and Z. Li, "Convergence analysis of online gradient method for BP neural networks," *Neural Networks*, vol. 24, no. 1, pp. 91–98, 2011.
- [21] N. S. Keskar, D. Mudigere, J. Nocedal, M. Smelyanskiy, and P. T. Tang Peter, "On large-batch training for deep learning: generalization gap and sharp minima," 2016, <https://arxiv.org/abs/1609.04836>.
- [22] P. Hurtik, V. Molek, and J. Hula, "Data preprocessing technique for neural networks based on image represented by a fuzzy function," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 7, pp. 1195–1204, 2020.
- [23] F. Eustáquio and T. Nogueira, "Evaluating the numerical instability in fuzzy clustering validation of high-dimensional data[]," *Theoretical Computer Science*, vol. 805, pp. 19–36, 2020.
- [24] R. J. Hathaway and J. C. Bezdek, "Fuzzy c-means clustering of incomplete data," *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, vol. 31, no. 5, pp. 735–744, 2001.
- [25] Y. Zhang, X. Bai, R. Fan, and Z. Wang, "Deviation-sparse fuzzy C-means with neighbor information constraint," *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 1, pp. 185–199, 2019.
- [26] R. M. Aron, D. Carando, T. W. Gamelin, S. Lassalle, and M. Maestre, "Cluster values of analytic functions on a Banach space," *Mathematische Annalen*, vol. 353, no. 2, pp. 293–303, 2012.
- [27] M. E. Opler, "Themes as dynamic forces in culture," *American Journal of Sociology*, vol. 51, no. 3, pp. 198–206, 1945.
- [28] W. R. Tobler, "Jstor: a computer movie simulating urban growth in the detroit region," *Economic Geography*, vol. 46, no. jun, pp. 234–240, 1970.
- [29] M. F. Goodchild, "The fundamental laws of gScience," *Keynote Address Annual Assembly university Consortium for Gscience*, 2003, http://csiss.ncgia.ucsb.edu/aboutus/presentations/files/goodchild_ucgis_jun03.pdf.
- [30] M. Konarski and W. Zabierowski, "Using Google Maps API along with technology .NET," in *Proceedings of the 2010 International Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science (TCSET)*, pp. 180–182, Lviv, Ukraine, 2010.
- [31] U. Pub, *World Culture Report, 2000: Cultural Diversity, Conflict and Pluralism/*, 2000, <https://www.unescap.org/cgi-bin/koha/opac-detail.pl?biblionumber=10742>.
- [32] A. Onan, S. Korukoğlu, and H. Bulut, "Korukoglu, & Serdar. A multiobjective weighted voting ensemble classifier based on differential evolution algorithm for text sentiment classification," *Expert Systems with Applications*, vol. 62, pp. 1–16, 2016.
- [33] H. H. Nguyen, "Privacy-preserving mechanisms for k-modes clustering," *Computers & Security*, vol. 78, no. sep, pp. 60–75, 2018.
- [34] R. Trisminingsih and S. S. Shaztika, "St-dbscan clustering module in spagobi for hotspots distribution in Indonesia," in *Proceedings of the 2016 3rd International Conference on Information Technology, Computer, and Electrical Engineering (ICI-TACEE)*, pp. 327–330, IEEE, Semarang, Indonesia, October 2016.

- [35] X. Yang, F. Yu, and W. Pedrycz, "Typical characteristics-based type-2 fuzzy C-means algorithm," *IEEE Transactions on Fuzzy Systems*, no. 99, p. 1, 2020.
- [36] W. Zhou Wang and A. C. Bovik, "Mean squared error: l," *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 98–117, 2009.
- [37] K. Krishna and M. Narasimha Murty, "Genetic K-means algorithm," *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, vol. 29, no. 3, pp. 433–439, 1999.
- [38] M. Ester, "A density-based algorithm for discovering clusters in large spatial databases with noise," *Proc.int.conf.knowledg Discovery & Data Mining*, vol. 96, pp. 226–231, 1996.