



An Improved Clustering Method for Multi-view Images

Yang Dong^(✉), Dazhao Fan, Qiuhe Ma, and Song Ji

Information Engineering University, Zhengzhou, China
wenku34@163.com

Abstract. Existing algorithms do not meet the requirements of multi-view image clustering under big data conditions. Here, we design a multi-view clustering algorithm for massive and unstructured images. To meet the multi-view requirements, improve speed and accuracy, a response layer is introduced to the self-organizing map neural network. An online self-organizing map neural network with simple parameters and without prior training is proposed and used for the multi-view clustering process. Experiments are performed using multiple datasets. The results show that the proposed algorithm is capable of multi-view clustering of image data with high accuracy, low error rate and favorable stability.

Keywords: Image clustering · Multi-view · Self-organizing map neural network

1 Introduction

In recent years, the widespread application of information network technology has promoted changes in lifestyle. The Internet, the Internet of Things, knowledge services, and intelligent services have become indispensable parts of people's lives. These components form a huge sensor network collecting unsustainably massive amounts of image data that are complex in type, huge in volume, critical in time, and have prominent big data features; such data have become an important research object. The first step in processing these inaccurate, unstructured big image data is to conduct autonomous clustering of images to find a collection of images with similar content in the same target area. Image clustering can be classified into two steps: generating the global description of each image and clustering the image descriptors using a clustering method. The global descriptor of an image is typically obtained through the aggregation of local image descriptors. Many scholars have conducted related studies [1–6]. This part work has been done in our previous conference papers [7].

For image clustering methods, it can be divided into various types according to the clustering characteristics, including the following: clustering algorithms based on partitioning, such as the k-means algorithm [8–10]; hierarchical clustering algorithms, such as the clustering using representatives (CURE) algorithm [11]; density-based clustering algorithms, such as the density-based spatial clustering of applications with noise (DBSCAN) algorithm [12]; grid-based clustering algorithms, such as the statistical information grid (STING) algorithm [13]; and model-based clustering algorithms,

such as the self-organizing map (SOM) algorithm [14]. Current clustering algorithms mostly consider clustering between similar content, but there is little research on multi-view image clustering between the same target. Finding multi-view image of the same target has important applications in three-dimensional reconstruction, image registration, and data fusion, etc. Therefore, achieving better clustering of multi-view images has important research significance.

In this study, an improved clustering method is proposed, achieving an accurate multi-view clustering process. The traditional SOM neural network is extended, and a response layer is introduced to produce a three-layer online SOM neural network clustering algorithm to comprehensively cluster multi-view image data. We introduce the response layer, simplify the input parameters and eliminate the pre-training process, thereby improving the accuracy of the overall multi-view clustering results and achieving a suitable performance and stability of the multi-view clustering process.

2 Materials and Methods

Most current clustering algorithms do not consider multi-view images. Some of the image content between multi-view images may be completely different, which brings great difficulties to clustering. Therefore, we improved SOM neural network to achieve this task, in the hope of obtaining complete multi-view clustering results.

2.1 Online SOM Neural Network

The SOM neural network is a method used to numerically simulate human brain neural function. This network is an unsupervised competitive learning feedforward neural network that can achieve unsupervised self-organized learning in training [14]. The traditional SOM is a two-layer neural network: the first layer is the input layer, which accepts input eigenvectors, and the second layer is the competing layer (in which each node is a neuron) that outputs the classification result of the input samples. The basic working principle is that when the sample data enter the input layer, the distance between each input sample data value and the neuron's weight is calculated and that the neuron with the smallest distance to the input sample data becomes the competitive neuron. The weights of the competitive neuron and the adjacent neuron are adjusted to attain values similar to the input sample data values. After many rounds of training, the neurons are divided into various regions that can map the input sample and cluster the input data.

The SOM method is a model-based clustering algorithm whose strategy is to construct the data model based on prior knowledge and use the model to cluster the new data; this strategy has unique advantages for high-dimensional data processing. However, the traditional SOM neural network must be trained based on prior knowledge, making this network type unsuitable for clustering multi-view, massive and disordered image data. It is important to optimize the SOM network to achieve the characteristics of online learning and multi-view clustering.

Considering the existence of nerve endings in the actual brain model, an improved SOM neural network is proposed for obtaining an online self-organizing map neural

network (OSOM). The OSOM is designed as a three-layer neural network. The first two layers are similar to those of the traditional SOM and correspond to the data input and competition layers. The additional layer responds to the nerve endings and corresponds to the response layer. In Fig. 1, the first layer is illustrated as light yellow cells, which represent input data; the neurons of the second layer are illustrated as large circles, of which the blue circles represent inactive neurons and the yellow circles represent the neurons that have won the competition; and the third layer consists of small circles that represent nerve endings, where red circles correspond to inactivated nerve endings and yellow to neurons that have successfully entered the activated state. To meet the requirements of performance and accuracy, this approach uses the input of high-dimensional data based on stream mode and incremental learning using an improved SOM neural network to realize multi-view clustering of data and online learning of the neural network.

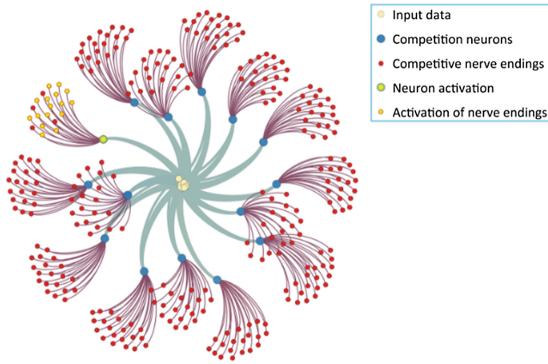


Fig. 1. OSOM neural network diagram. (Color figure online)

The OSOM can continuously generate neurons and nerve endings during data processing and does not need to perform prior training or initialize the network in advance. This is an important reason that why OSOM could get complete multi-view clustering results. The specific process is as follows:

1. Input the data stream from the input layer. The input data stream consists of two parts, namely, the global image descriptor $x = \{x_1, x_2, \dots, x_m\}$ and the local image descriptor set $Y = \{y_1, y_2, \dots, y_n\}$, where x is a single vector, m is its dimension, Y is a vector set, n is the number of local image feature points, and $y_i = \{\gamma_1, \gamma_2, \dots, \gamma_h\}$ is the i -th descriptor of the local feature point of the image, which is of dimension h ;
2. Enter the neuronal competition mode. Calculate the distance between the input global descriptor x and each competing neuron connection weight ω , and identify the nearest N neurons, that is, impose the following condition

$$N\|x - \omega\| = \min_i(N\|x - \omega_i\|) \quad (1)$$

where N represents the first N nearest neuron sets, and ω_i represents the connection weight of the i -th competing neuron;

3. The first N competitive neurons enter the response mode in turn, which is called the response to competitive neurons. Calculate the shortest distance l between each local descriptor y in set Y and nerve end ϖ under each competitive neuron, that is,

$$l_{ik} = \min_j \|y_j - \varpi_{ik}\| \quad (2)$$

where y_j denotes the j -th input local descriptor that traverses set Y , ϖ_{ik} denotes the connection weight of the k -th nerve terminal under the i -th responding competitive neuron, and l_{ik} denotes the shortest distance between the corresponding nerve endings and set Y .

If distance l_{ik} is less than a threshold value α , the corresponding nerve endings determine that the response is successful. If the number of successes of the nerve endings in responding to the i -th responsive neurons is greater than a threshold value β , then the overall response is judged to be successful, and the other neurons no longer respond.

$$\sum_{k=1}^K \xi(\alpha - l_{ik}) > \beta \quad (3)$$

where k is the total number of nerve endings under the i -th competitive neuron and $\xi(x)$ is the response function, which is expressed as

$$\xi(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases} \quad (4)$$

4. Enter the feedback learning mode. If there is a neuron and nerve ending overall response to the success of the corresponding neurons and nerve terminals using a certain learning efficiency to learn the data.

$$\begin{cases} \Delta\omega_i = \chi_i(x - \omega_i) \\ \Delta\varpi_{ij} = \chi_{ij}(y_g - \varpi_{ij}) \end{cases} \quad (5)$$

$$\begin{cases} \omega_i(t+1) = \omega_i(t) + \Delta\omega_i(t) \\ \varpi_{ij}(t+1) = \varpi_{ij}(t) + \Delta\varpi_{ij}(t) \end{cases} \quad (6)$$

where t is the number of learning rounds, χ is the learning efficiency, x is the global descriptor for neuron response, and y_g is the local descriptor for obtaining the nerve ending response. Then, a new descriptor within each neuron is generated using a local descriptor that failed to successfully obtain the nerve ending response.

$$\varpi_{i(k+1)} = \chi_{i(k+1)} y_d \tag{7}$$

After learning, the learning efficiency χ of the neuron and the corresponding nerve endings decreases by a specified step size, which is designated δ .

$$\chi_i(t_i + 1) = \begin{cases} \chi_i(t_i) - \delta & t_i \leq m \\ E & t_i > m \end{cases} \tag{8}$$

where δ is the step-down rate of each learning rate; m is the upper limit on the number of reductions in the learning rate, which satisfies $m \leq \chi_i(0)/\delta$; and E is the learning termination rate. If all the nerve endings that correspond to the first N winning neurons fail to respond to the learning process, generate new neurons and nerve endings.

$$\begin{cases} \omega_r = \chi_r x \\ \varpi_{i(k+1)} = \chi_{i(k+1)} y_d \end{cases} \tag{9}$$

5. Continue to enter the data stream, and return to 2 until all the data have been processed.

Each neuron that corresponds to the image constitutes a category, thereby completing the multi-view clustering process for image data. In practice, to control the growth of the number of neurons under constant input, old neurons can be merged after a specified amount of time to yield more accurate clustering results.

2.2 Multi-view Clustering Algorithm

Using the OSOM neural network, a multi-view clustering algorithm for image data is proposed. The specific process is illustrated in Fig. 2.

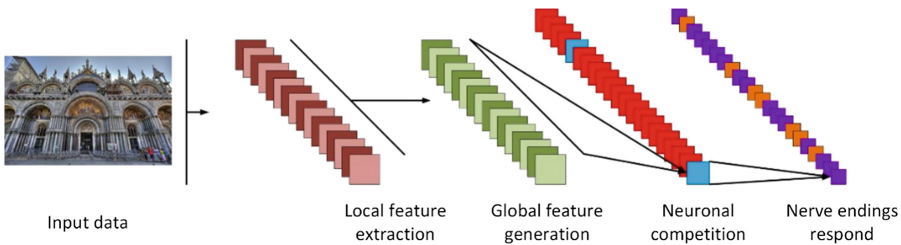


Fig. 2. Multi-view clustering algorithm diagram.

First, the global descriptor are extracted from the input image, which is designated u . Then, descriptor u is inputted into the OSOM neural network, and the N nearest neurons are identified. Finally, the corresponding nerve ending responses are calculated, and

learning and rules generation are performed to produce the image clustering results. Additionally, to improve the processing efficiency, parallel input of the images can be performed.

3 Results and Discussion

Clustering analysis of image data is conducted to identify images of the same target automatically. Therefore, there are two main indices for evaluating the clustering algorithm: the ratio of the number of correct clustering results to the total number of clustering results, that is, the accuracy rate (AR), and the number of correctly clustered images lost from the clustering result, that is, the negative true rate (NTR). The former index evaluates the locality of the clustering results and characterizes the in-class performance of the clustering algorithm. The latter index evaluates the integrity of the clustering results and characterizes the inter-class performance of the clustering algorithm. In the experiments, we use these two indicators as the main criteria. To evaluate the proposed OSOM neural network, it is designed for comparison with the classical k-means algorithm. To evaluate the performance and accuracy of the overall algorithm flow, 10,000 unordered and cluttered Internet images are used as inputs to conduct experiments, and the final clustering results are analyzed and evaluated.

3.1 OSOM Neural Network Experiments

The OSOM neural network is tested and compared using the classical k-means clustering algorithm. The specific experimental design is as follows: Clustering experiments are conducted using 10 categories of data from the BMW dataset [15]. A total of 100 images of the BMW dataset are used in the experiments, each of which contains 10 images of the same scene from different angles. VELAD [7] is used to obtain a global description of the images. Clustering experiments are performed using OSOM and k-means, respectively.

To describe the clustering results more intuitively, the AR, the NTR and false positive rate (FPR) of each method for each type of data clustering result are calculated. The cluster type defined as the main component in the clustering results. The AR of the calculation is $\eta = m/M$, where m is the number of correctly clustered images in the class, and M is the total number of clustered images in the class. The NTR $\gamma \gamma = w/I$ is the true-negative rate, where w is the number of images that cannot be divided into the class, and I is the total number of images that should be assigned to the class. The FPR ε is expressed as $\varepsilon = r/M$, where r is the number of correctly clustered images in the class. According to the above definition, the higher AR, the lower FPR and NTR, and the better the overall performance of the method.

The results are shown in Fig. 3, Tables 1 and 2. The OSOM threshold is set to 25, the learning efficiency is set to 1.0, the learning step is set to 0.3, the number of clusters in the k-means clustering process is set to 10. The image type that is the main component of the specified clustering result is the final representative result of the cluster type.

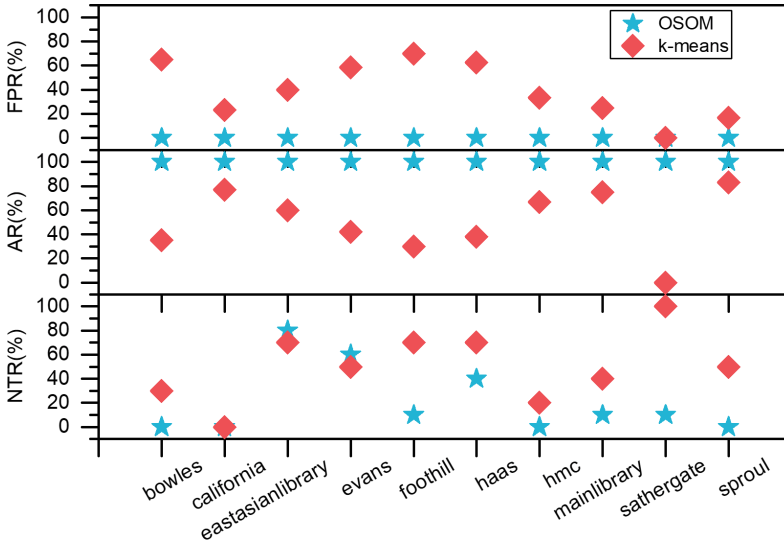


Fig. 3. Clustering results of comparative analysis.

Table 1. OSOM clustering statistics table.

Class name	Total number of clusters	Correct number	AR	FPR	NTR
Bowles	10	10	100.00%	0.00%	0.00%
California	10	10	100.00%	0.00%	0.00%
East asian library	2	2	100.00%	0.00%	80.00%
Evans	4	4	100.00%	0.00%	60.00%
Foothill	9	9	100.00%	0.00%	10.00%
Haas	6	6	100.00%	0.00%	40.00%
HMC	10	10	100.00%	0.00%	0.00%
Main library	9	9	100.00%	0.00%	10.00%
Sathergate	9	9	100.00%	0.00%	10.00%
Sproul	10	10	100.00%	0.00%	0.00%

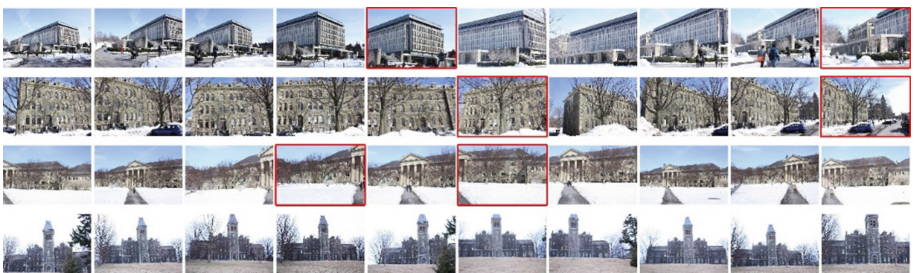
According to the data in Table 1 and Fig. 3, for the BMW dataset, the OSOM clustering AR remains at 100%, which is much higher than that of the k-means algorithm. However, the correctness of k-means clustering is generally low, and the clustering of the ‘sathergate’ class is fully incorrect. According to the data in Table 2 and Fig. 3, the ER of OSOM is consistently 0%, which is lower than that of k-means clustering. The OSOM FPR and NTR values are also lower than those of k-means clustering. Overall, the experiments on the BMW dataset show that OSOM achieves a higher overall clustering performance than traditional k-means clustering and exhibits a higher clustering AR and better stability.

Table 2. K-means clustering statistics table.

Class name	Total number of clusters	Correct number	AR	FPR	NTR
Bowles	20	7	35.00%	65.00%	30.00%
California	13	10	76.92%	23.08%	0.00%
East asian library	5	3	60.00%	40.00%	70.00%
Evans	12	5	41.67%	58.33%	50.00%
Foothill	10	3	30.00%	70.00%	70.00%
Haas	8	3	37.50%	62.50%	70.00%
HMC	12	8	66.67%	33.33%	20.00%
Main library	8	6	75.00%	25.00%	40.00%
Sather gate	0	0	0.00%	0.00%	100.00%
Sproul	6	5	83.33%	16.67%	50.00%

3.2 Multi-view Clustering Algorithm Experiments

To evaluate the multi-view clustering algorithm for practical applications, the following experiments are designed: Using a dataset of 10,000 cluttered Internet images, the multi-view clustering experiments of this method are conducted. The images are downloaded from Flickr [16]. The VELAD parameters and the OSOM parameters are set to the same values as in the previous experiments. The final clustering result contains 81 image classes, each containing 20 or more images. Partial clustering results are shown in Fig. 4. The clustering result in Fig. 4 contains multi-view images of the same scene. As shown in the red box in the Fig. 4, it can be seen that the content in the red frame is different side images of the same building, and the content is completely different. This shows that our method can achieve the clustering of multi-view images well.

**Fig. 4.** Experimental clustering results for ten thousand images. (Color figure online)

In the experiments, the image type that constitutes the main component of the specified clustering result is the final representative of the clustering type, and the AR of the clustering results is calculated, as shown in Fig. 5.

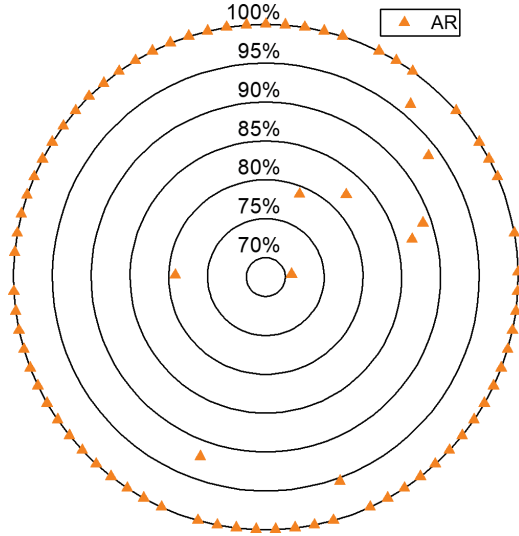


Fig. 5. AR chart for the results of clustering experiments on ten thousand images.

The ARs of the 81 clustering results obtained in the clustering experiment on 10,000 images are relatively high. The ARs of 71 clustering results are 100%, and the ARs of 3 clustering results are as low as 80%. The lowest AR of the clustering results is 70.8%, and the average correct classification rate is 98.33%. According to the results of the clustering experiments on ten thousand images, the clustering results obtained by the method proposed here exhibit high correctness and favorable stability and can satisfy the clustering requirements of images under multi-view data conditions.

3.3 Discussion and Analysis

With the continuous formation of big image data, clustering between multi-view images has gradually become a hot research field. However, the existing algorithms mostly consider clustering of the same content, there is little discussion about clustering between different angle images of the same target. Although some content may vary greatly among multi-view images, there may be some content crossover between images. With this assumption, we have improved SOM. OSOM does not need to specify the number of clusters and does not require prior training. Through the online learning in the clustering, the multi-view feature of the same target is obtained. Thus, a better multi-view clustering result is obtained. From the 10,000 Internet images experiments, we can see that OSOM can complete multi-view clustering tasks, obtain a good result. However, this paper only proposes a multi-view processing method, the application of the method in actual engineering needs to continue to be optimized (processing time, hardware consumption, etc.).

4 Conclusions

We have designed an OSOM neural network to achieve an accurate multi-view clustering processing. Considering the multi-view clustering process, a three-layer OSOM neural network algorithm without prior training is proposed. This algorithm simplifies the clustering process and satisfies the accuracy and speed requirements. The experimental results of the BMW dataset and a dataset of 10,000 Internet images show that the proposed algorithm offers high accuracy and favorable stability and can accomplish clustering tasks for multi-view image data.

References

1. Csurka, G., Dance, C., Fan, L., Willamowski, J., Bray, C.: Visual categorization with bags of keypoints. In: Proceedings of the European Conference on Computer Vision, Prague, pp. 1–2 (2004)
2. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New York, pp. 2169–2178 (2006)
3. Yang, J., Yu, K., Gong, Y., Huang, T.: Linear spatial pyramid matching using sparse coding for image classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, pp. 1794–1801 (2009)
4. Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y.: Locality-constrained linear coding for image classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, pp. 3360–3367 (2010)
5. Perronnin, F., Sánchez, J., Mensink, T.: Improving the fisher kernel for large-scale image classification. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6314, pp. 143–156. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15561-1_11
6. Russakovsky, O., Lin, Y., Yu, K., Fei-Fei, L.: Object-centric spatial pooling for image classification. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012. LNCS, pp. 1–15. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33709-3_1
7. Dong, Y., Fan, D., Ma, Q., Ji, S., Lei, R.: Edge-based locally aggregated descriptors for image clustering. *Int. Arch. Photogram. Remote Sens. Spat. Inf. Sci.* **XLII-3**, 303–308 (2018)
8. MacQueen, J.: Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, pp. 281–297 (1967)
9. Wang, H.X., Jin, H.J., Wang, J.L., Jiang, W.S.: Optimization approach for multi-scale segmentation of remotely sensed imagery under k-means clustering guidance. *Acta Geodaetica Cartogr. Sin.* **44**, 526–532 (2015)
10. Jiayao, W., Mingxia, X., Jianzhong, G.: Improved high dimensional data clustering algorithm based on similarity preserving and feature transformation. *Acta Geodaetica Cartogr. Sin.* **40**, 269–275 (2011)
11. Guha, S., Rastogi, R., Shim, K.: CURE: an efficient clustering algorithm for large databases. In: Proceedings of the ACM SIGMOD International Conference on Management of Data, Washington, pp. 73–84 (1998)

12. Ester, M., Kriegel, H.P., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the International Conference on Knowledge Discovery and Data Mining, Portland, pp. 226–231 (1996)
13. Wang, W., Yang, J., Muntz, R.: STING: a statistical information grid approach to spatial data mining. In: Proceedings of the International Conference on Very Large Data Bases, Athens, pp. 186–195 (1997)
14. Kohonen, T.: Self-organization and associative memory. *Appl. Opt.* **8**, 3406–3409 (1989)
15. BMW Images Dataset. <http://download.csdn.net/detail/wmgd85/9700927>
16. Flickr. <https://www.flickr.com>