

2013 AASRI Conference on Intelligent Systems and Control

Comparison of Distance Measures on Fuzzyc-Means Algorithm for Image Classification Problem

Jiho Han^a, Dong-Chul Park^{a*}, Dong-Min Woo^a, and Soo-Young Min^b^a*Dept. of Electronics Eng., Myongji Univ., Gyeonggi-do, 449-728, Rep. of KOREA*^b*Korea Electronics Tech. Inst., SongNam, 463-816, Rep. of KOREA*

Abstract

A study on the use of two different distance measures, Euclidean distance and divergence distance, for FCM is conducted for an image classification problem in this paper. Conventional FCM algorithm which uses Euclidean distance measure utilizes only mean information from an image block for its feature while FCM algorithm with divergence utilizes both of variance and mean information. Evaluations on a set of Caltech databases show that Divergence-based FCM gives higher accuracy when compared with some conventional algorithms with Euclidean distance.

© 2013 The Authors. Published by Elsevier B.V. Open access under [CC BY-NC-ND license](#).
Selection and/or peer review under responsibility of American Applied Science Research Institute

Keywords: Fuzzy c-Means, Classification, Divergence Measure, SOM

1. Introduction

Recent increase use of compressed image data requires for an automatic tool that can retrieve image data efficiently. Several conventional algorithms adopted for this purpose include Fuzzy c-Means (FCM) algorithm and Self Organizing Map (SOM) [1][2]. FCM algorithm is the most widely used one and can be thought as an improved version of earlier clustering. The FCM algorithm shows more robustness when

* Corresponding author. Tel.: +82-31-330-6975; fax: +82-31-330-6977
E-mail address: parkd@dreamwiz.com

compared with SOM and k-Means algorithm. By combining the ideas of FCM and divergence measure for the problem of image classification, we expect a better feature extraction procedure for more accurate classification scheme. The method adopts the FCM algorithm with Divergence Measure [3]-[5] for acquiring texture information from image data.

Section 2 summarizes FCM algorithms with two different distance measures: Euclidean distance and divergence distance. Section 3 reports the evaluations and a comparison of different schemes through experiments. Section 4 presents conclusions.

2. Adopted Clustering Algorithms

2.1. Fuzzy c-Means Algorithm

The following equation is used as the objective function for FCM[2] :

$$J_m(U, \mathbf{v}) = \sum_{k=1}^n \sum_{i=1}^c (\mu_{ki})^m (d_i(\mathbf{x}_k))^2 \quad (1)$$

where $d_i(\mathbf{x}_k)$, μ_{ki} , m , n , and c follow the definitions in [2].

From Eq. (1), Bezdek finds the following equations [2]:

$$\mu_{ki} = \frac{1}{\sum_{j=1}^c \left(\frac{d_i(\mathbf{x}_k)}{d_j(\mathbf{x}_k)} \right)^{\frac{2}{m-1}}} \quad (2)$$

$$\mathbf{v}_i = \frac{\sum_{k=1}^n (\mu_{ki})^m \mathbf{x}_k}{\sum_{k=1}^n (\mu_{ki})^m} \quad (3)$$

2.2. FCM with Divergence Measure

The choice of different distance measure can affect the performance of clustering results [1]. This leads to the idea of divergence measure for image classification problem.

Given two Gaussian Probability Density Functions, $\mathbf{x} = (x_i^\mu, x_i^{\sigma^2})$ and $\mathbf{v} = (v_i^\mu, v_i^{\sigma^2})$, $i = 1, \dots, d$, the divergence distance can be defined as follows[6,8]:

$$\begin{aligned} D(\mathbf{x}, \mathbf{v}) &= \sum_{i=1}^d \left(\frac{x_i^{\sigma^2} + (x_i^\mu - v_i^\mu)^2}{v_i^{\sigma^2}} + \frac{v_i^{\sigma^2} + (x_i^\mu - v_i^\mu)^2}{x_i^{\sigma^2}} - 2 \right) \\ &= \sum_{i=1}^d \left(\frac{(x_i^{\sigma^2} - v_i^{\sigma^2})^2}{x_i^{\sigma^2} v_i^{\sigma^2}} + \frac{(x_i^\mu - v_i^\mu)^2}{x_i^{\sigma^2}} + \frac{(x_i^\mu - v_i^\mu)^2}{v_i^{\sigma^2}} \right) \end{aligned} \quad (4)$$

where $x_i^\mu, x_i^{\sigma^2}, v_i^\mu$ and $v_i^{\sigma^2}$ follow the definition in [6,8]. Note that the divergence measure is also called as *Kullback-Leibler Divergence*.

FCM calculates the parameters of center and membership values by applying all the data at once in batch mode. However, the D-FCM used in the proposed classifier calculates and updates their parameters at each application of each data vector as the Gradient-based FCM [3]. The advantage of this iterative application and updating the center parameters was reported in [3]. When each data vector \mathbf{x} is presented to the network, the following can be also found :

$$\mu_i(\mathbf{x}) = \frac{1}{\sum_{j=1}^c \left(\frac{D(\mathbf{x}, v_i)}{D(\mathbf{x}, v_j)} \right)^2} \quad (5)$$

Subsequently, the following update equation for the mean and variance was derived[6]:

$$v_i^\mu(n+1) = v_i^\mu(n) - \eta \mu_i^2(\mathbf{x})(v_i^\mu(n) - \mathbf{x}^\mu) \quad (6)$$

$$v_i^{\sigma^2}(n+1) = \frac{\sum_{k=1}^{N_i} (x_{k,i}^{\sigma^2}(n) + (x_{k,i}^\mu(n) - v_i^\mu(n))^2)}{N_i} \quad (7)$$

where $v_i^\mu(n), v_i^{\sigma^2}(n), x_{k,i}^\mu(n), x_{k,i}^{\sigma^2}(n), \eta$ and N_i follow the definitions in [6].

3. Experiments and Results

For evaluation and comparison purpose for different classification schemes, we performed some experiments on a set of the Caltech image database[7]. The Caltech image database has been widely used as a benchmark data for various image classification problems. Examples of the four-category data used in experiments are given in Fig. 1. 200 data are used for each category. Preparation for experiments follows the procedure used in [4]: 150 randomly selected training data and 50 remaining data for each class are used for the evaluation of different classifier schemes. Note that all the image data are converted to grey images with identical resolution.

The first step for designing a data classifier is the feature extraction process. The local feature extraction adopted in our experiments allows us to find localized feature information from entire image space. By collecting the feature information from local points, each image data can be expressed in terms of features. In our experiments, 8×8 blocks are used for extracting localized information of an image data. The local information in each image window can be extracted by using conventional feature extraction methods [9]-[11]. Among them, Gabor filters and wavelet filters show reasonable performance except requiring extensive computational burden. This computational burden keeps these methods from being used for feature extraction tools for several applications. On the other hand, the Discrete Cosine Transform (DCT) is more suitable to acquire frequency information in images and DCT is used for our experiments because DCT does not require so much computational effort. Even though DCT coefficients out of each block image produces 64-dimensional one, the 32 lower frequency part of the 64 coefficients are used [4] and this yields a total 32-dimensional vector value for each local block image.

The model for a data category is formed by deciding the code vectors using clustering algorithms on its feature vectors. The modeling or training process consists of finding representative code vector for

each category with its feature vectors. Bayesian classifier is utilized for the performance evaluation on various classification schemes. It allocates a class for a given test data according to the following probability calculation [4]:

$$\text{Genre}(x) = \arg \max_i P(x|v_i) \quad (8)$$

$$P(x|v_i) = \sum_{i=1}^M c_i \mathfrak{N}(x, \mu_i, \Sigma_i) \quad (9)$$

$$\mathfrak{N}(x, \mu_i, \Sigma_i) = \frac{1}{\sqrt{(2\pi)^d |\Sigma_i|}} e^{-0.5(x-\mu_i)^T \Sigma_i^{-1} (x-\mu_i)} \quad (10)$$

where $M, c_i, d (=32)$, μ_i and Σ_i follow the definitions in [4].



(a) Car



(b) Motorbike



(c) Bike



(d) Airplane

Fig. 1 Examples of data

The same training and test procedures for the same data sets are performed on two conventional clustering algorithms, SOM and FCM, and compared with the accuracy of the proposed algorithm. During experiments, several numbers of code vectors up to 19 are used for each algorithm. Fig. 2 summarizes their classification accuracies. The average accuracies over different numbers of code vectors (3 to 19) for SOM, FCM, and D-FCM are observed as 42.22%, 44.61%, and 73.33%, respectively. The classification accuracy is increased with code vector numbers. Note that their classification accuracies are somewhat saturated when code vector numbers reach 15. Note also that SOM and FCM don't use the variance while D-FCM does. Table 1 summarizes the classification performance for each category of image data when 15 code vectors are used. From Fig. 2 and Table 1, we can notice that the covariance value utilized in D-FCM along with mean information plays an important role in improving the classification accuracy for image data. The divergence distance measure allows the image classifier to have an advantage over the Euclidean distance and similar results are reported in [4]. Table 2 shows a confusion matrix of D-FCM based image classifier for the case of 15 code vectors. Car images are classified well enough while motorbike data are confused with airplane data. Some other feature extraction methods that can discriminate car data from airplane data need to be done in future research. Future research also include the comparison of FCM and Gradient-based FCM under divergence measure environment.

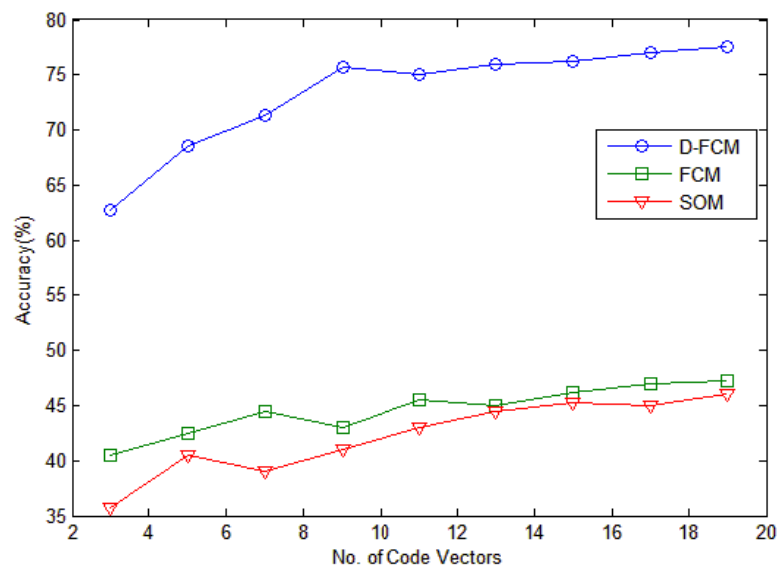


Fig. 2. Comparison of classification accuracies of various classifier schemes over numbers of code vectors

Table 1. Classification performance of SOM, FCM, and D-FCM for the case of 15 code vectors

| | Car | Bike | Motorbike | Airplane | Avg. Accuracy(%) |
|-------|-----|------|-----------|----------|---------------------|
| SOM | 98 | 63 | 7 | 13 | 45.25 |
| FCM | 97 | 66 | 5 | 17 | 46.25 |
| D-FCM | 100 | 78 | 51 | 76 | 76.25 |

Table 2. Detailed classification results of D-FCM

| | | <i>Classified</i> | | | |
|---------------|-----------|-------------------|------|-----------|----------|
| | | Car | Bike | Motorbike | Airplane |
| <i>Inputs</i> | Car | 0 | 100 | 0 | 0 |
| | Bike | 2 | 78 | 12 | 8 |
| | Motorbike | 0 | 6 | 51 | 43 |
| | Airplane | 7 | 2 | 15 | 76 |

4. Conclusions

A comparative study Euclidean distance and divergence distance for FCM is conducted in this paper. When used for an image classification problem, conventional FCM algorithm which uses Euclidean distance measure utilizes only mean information from an image block for its feature. On the other hand, FCM algorithm with divergence utilizes both of variance and mean information from an image block for its features. Self-Organizing Map is also used in our experiments as a baseline method for the performance comparison purpose. Experiments on Caltech database are performed for a four category problem. Classification accuracies from different classification schemes imply that divergence-based FCM is a better fit for image classification problem than the conventional FCM or SOM that use only the mean information as their features for image data. We can conclude that the divergence information of image data plays an important role in image classification problem and the resultant classification scheme that uses the divergence information as well as the mean information from image data is a better choice for image classification problem. However, a further effort should be given to find a feature extraction method for overcoming the confusion between Motorbike data and Airplane data.

Acknowledgements

This work was supported by IT R&D program of The MKE/KEIT (10040191) and by National Research Foundation of Korea Grant funded by the Korean Government (2010-0009655).

References

- [1]Bezdek, J.C. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum;1981.
- [2]Kohonen,T.The Self-Organizing Map. Proceedings of IEEE, 1990;78 :1464-1480.
- [3]Park D.C.et. al. Classification of Audio Signals Using Gradient-Based Fuzzy c-Means Algorithm with Divergence Measure. Proc. Pacific Rim Conference on Multimedia.2005; 698-708.
- [4]Park D.C., Woo D.M. Image classification using Gradient-Based Fuzzy c-Means with Divergence Measure. Proc. Int. Joint Conference on Neural Networks.2008; 2520-2524.
- [5]Park D.C. Satellite Image Classification Using a Divergence-Based Fuzzy c-Means Algorithm.Proc.Int. Conf. on Image and Signal Processing. 2012;555-561.
- [6]Park D.C. et. al.Centroid Neural Network With a Divergence Measure for GPDF Data Clustering.IEEE

Transactions on Neural Networks. 2008;19(6): 948-957.

[7]<http://www.vision.caltech.edu/html-files/archive.html>

[8]FukunagaK. Introduction to Statistical Pattern Recognition, Academic Press Inc.;1990.

[9]Daugman J.G. Complete Discrete 2D Gabor Transform by Neural Networks forImage Analysis and Compression. IEEE Transactions on Acoustics, Speech, and SignalProcessing.1988;36:1169-11179.

[10]Pun C.M., LeeM.C. Extraction of Shift Invariant Wavelet Features for Classification of Images with Different Sizes. IEEE Transactions on Pattern Analysis andMachine Intelligence. 2004;26(9) :1228-1233.

[11]Huang Y.L., Chang R.F. Texture Features for DCT-Coded Image Retrieval andClassification. Proc. Int. Conf. on Acoustics, Speech, and SignalProcessing.1999;6:3013-3016.