

ADAPTIVE FUZZY KERNEL CLUSTERING ALGORITHM

Weijun Xu¹

¹The Department of Electrical and Information Engineering, Northeast Petroleum University at Qinhuangdao, Qinhuangdao, P.R. China

ABSTRACT

Fuzzy clustering algorithm can not obtain good clustering effect when the sample characteristic is not obvious and need to determine the number of clusters firstly. For this reason, this paper proposes an adaptive fuzzy kernel clustering algorithm. The algorithm firstly use the adaptive function of clustering number to calculate the optimal clustering number, then the samples of input space is mapped to high-dimensional feature space using gaussian kernel and clustering in the feature space. The Matlab simulation results confirmed that the algorithm's performance has greatly improvement than classical clustering algorithm and has faster convergence speed and more accurate clustering results.

KEYWORDS

Fuzzy clustering; Gaussian kernel; Adaptive clustering number; fuzzy kernel clustering

1. INTRODUCTION

Clustering is an unsupervised learning process. C-means method (HCM) and fuzzy C-means method (FCM)^[1, 2] directly cluster on the sample's characteristics and the effect of clustering largely depends on the distribution of the sample. These methods are not suitable for found the nonconvex shape of clusters or cluster size difference is very big, and the noise data points would greatly affect the end result of clustering. This paper introduces gaussian kernel^[3] into FCM, for this the samples of input space is mapped to high-dimensional feature space and clustering in the feature space in order to get better clustering results. In addition, the HCM and the FCM need to make sure of the number of clustering. This paper puts forward an adaptive fuzzy kernel clustering algorithm. The algorithm firstly use an adaptive function of clustering number to calculate the optimal clustering number, then the samples of input space are mapped to high-dimensional feature space and clustering in the feature space. The Matlab simulation results confirmed the algorithm has faster convergence speed and more accurate clustering results than classical clustering algorithm.

2. FUZZY C-MEANS ALGORITHM

Based on the distance measure is the basis of many clustering algorithms, like HCM and FCM. The FCM is generalized from HCM by J. C. Bezdek^[1, 2]. It has become one of the most commonly used and more discussed clustering algorithm. Its principle is described as below: $X = \{x, i=1, 2, \dots, n\}$ is a sample set, $X \subseteq R^p$, c is expected clustering number, $v_i (i = 1, 2, \dots, c)$ is the i th

cluster center, u_{ij} ($i = 1, 2, \dots, c; j = 1, 2, \dots, n$) is the membership functions of the j th sample to the i th class, and $0 \leq u_{ij} \leq 1; 0 < \sum_{j=1}^n u_{ij} < n$. The objective function of FCM is:

$$J_m(U, V) = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m \|x_j - v_i\|^2 \quad (1)$$

Where $U = \{u_{ij}\}$, $V = (v_1, v_2, \dots, v_c)$, $\|\bullet\|$ is the Euclidean distance generally, m ($m > 1$) is the fuzzy weighting exponent. The constraint condition is:

$$\sum_{i=1}^c u_{ij} = 1, \forall j = 1, 2, \dots, n \quad (2)$$

J. C. Bezdek gave an iterative algorithm to get the optimal solution of the above mathematical programming problem. The iterative formulae in the algorithm are:

$$u_{ij} = \frac{(1/\|x_j - v_i\|^2)^{1/(m-1)}}{\sum_{k=1}^c (1/\|x_j - v_k\|^2)^{1/(m-1)}}, \forall i = 1, 2, \dots, c; j = 1, 2, \dots, n \quad (3)$$

$$v_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m}, \forall i = 1, 2, \dots, c \quad (4)$$

FCM is suitable for the ball or elliptic globular cluster and is very sensitive to noise and outliers. By introducing the kernel method to fuzzy clustering can effectively solve those problems^[3,4]. In HCM and FCM clustering algorithm the number of clustering must be given in advance, in order to solve this problem, this paper proposes an adaptive function of clustering number to calculate the optimal clustering number, so as to realize the adaptive of clustering parameters.

3. THE ADAPTIVE FUNCTION OF THE OPTIMAL CLUSTERING NUMBER

In this section, we present a new adaptive function of the optimal clustering number and the corresponding algorithm. At last, we use four groups of synthetic data to test the algorithm.

3.1. The Adaptive Function

Evaluation of good clustering effect under the geometric meaning should satisfy two requirements:

- (a) divergence: the distance between the classes should be as large as possible;
 - (b) compactness: the distance between data points in the same class should be as small as possible.
- The proportion of divergence and compactness can be a clustering validity function, so we

constructed the following adaptive function of the optimal clustering number $C_{opt}(c)$:

$$C_{opt}(c) = \frac{\sum_{i=1}^c \sum_{j=1}^n u_{ij}^m \left\| v_i - \bar{x} \right\|^2 / (c-1)}{\sum_{i=1}^c \sum_{j=1}^n u_{ij}^m \left\| x_j - v_i \right\|^2 / (n-c)} \quad (5)$$

Where the definition of $u_{ij}, x, v, \|\bullet\|$ is as same as formula (1), \bar{x} is the central vector of the overall data:

$$\bar{x} = \frac{\sum_{i=1}^c \sum_{j=1}^n u_{ij}^m x_j}{n} \quad (6)$$

The optimal clustering number c is the one which $C_{opt}(c)$ reaches its maximum value.

3.2. Algorithm

Now, we outline the $C_{opt}(c)$ algorithm process. For simplicity, we assume $m = 2$ in the study below:

Step 1. Initialization: the termination condition $\varepsilon > 0$, cluster number $c = 2$, $L(1) = 0$, $V^{(0)}$, $k = 0$; (It can also use the partition matrix $U^{(0)}$ as the initial condition.)

Step 2. Calculate the partition matrix:

$$u_{ij}^{(k)} = 1 / \sum_{r=1}^c \left(\frac{d_{ij}^{(k)}}{d_{rj}^{(k)}} \right)^{\frac{2}{m-1}}$$

if there exist j and r , so that $d_{ij}^{(k)} = 0$, then $u_{ij}^{(k)} = 1$ and for $i \neq r$, $u_{ij}^{(k)} = 0$

Step3. Calculate the prototypes:

$$v_i^{(k+1)} = \sum_{j=1}^n (u_{ij}^{(k)})^m x_j / \sum_{j=1}^n (u_{ij}^{(k)})^m$$

Step4. Calculate the variation of partition matrix:

$$\|V^{(k+1)} - v^{(k)}\|$$

where $\|\bullet\|$ is some kind of matrix norm if $\|V^{(k+1)} - v^{(k)}\| < \varepsilon$ then goto step 4, else let $k = k + 1$ and go to Step2.

Step5. calculate $C_{opt}(c)$ under $c > 2$ and $c < n$:

If $C_{opt}(c-1) > C_{opt}(c-2)$ and $C_{opt}(c-1) > C_{opt}(c)$, then stop the iteration, else go to Step 2 with $c = c + 1$.

3.3. Examples

In this example, we use four groups of synthetic data to test the algorithm presented above. The four datasets are all consist of 2-dimension points, see Fig. 1. The optimal clustering number(OCN) of each data group evaluated by $C_{opt}^{(c)}$ are also shown there.

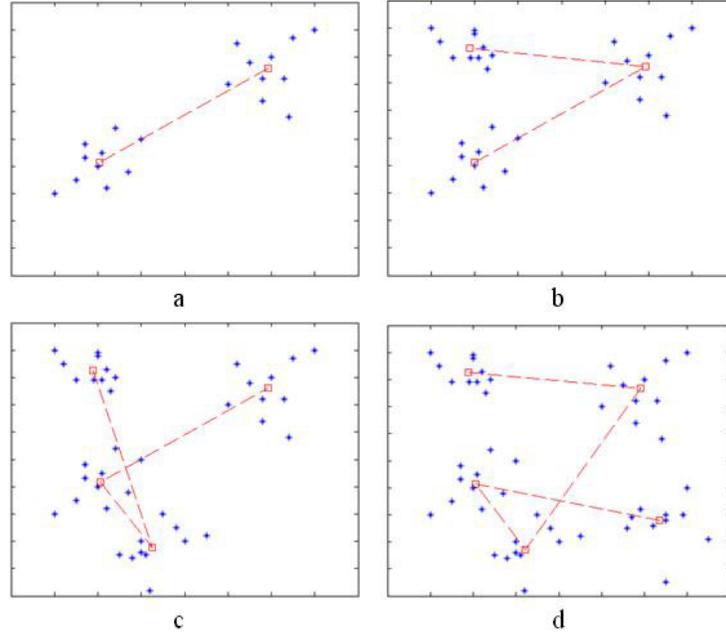


Figure 1. Four groups of synthetic data

Use the algorithm given in Section 3.2, we get the following OCNs corresponding to the four datasets(the bolded values indicate the OCNs):

a:BCN=2,(L(2),L(3))=(**116.1173**,98.0299)

b:BCN=3,(L(2),L(3), L(4))=(52.1941,**119.8551**,112.3083)

c:BCN=4,(L(2),L(3), L(4), L(5))=(74.0647, 114.8352, **155.9011**,145.8195)

d:BCN=5,(L(2),L(3),L(4),L(5),L(6))=(74.0877,81.1528,129.8848,**165.9403**,161.8719)

The four figures show that the adaptive function can figure out the optimal clustering number.

4. FUZZY KERNEL CLUSTERING ALGORITHM

Kernel clustering method can increase the optimization of sample characteristics and effectively improve the classification results of the algorithm. Based on the statistical learning theory, as long as the function meet the Mercer^[5] conditions, it can be regarded as a kind of kernel function. Through the using of Mercer kernel, the samples of input space are mapped to high-dimensional feature space and clustering in the feature space. Through the mapping of kernel function, make originally un conspicuous features stand out, so as to better clustering^[6].

We replace $\|x_j - v_i\|$ in the formula (1) with $\|\phi(x_j) - \phi(v_i)\|$, where $\phi(\bullet)$ is the nonlinear transformation function. The formula (1) can be changed as below:

$$J_m(U, V) = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m \|\phi(x_j) - \phi(v_i)\|^2 \quad (7)$$

Where $\phi(x_j)$ and $\phi(v_i)$ respectively represent the image of the sample and the clustering center in H feature space. The $\|\phi(x_j) - \phi(v_i)\|^2$ calculates as following:

$$\|\phi(x_j) - \phi(v_i)\|^2 = K(x_j, x_j) + K(v_i, v_i) - 2K(x_j, v_i) \quad (8)$$

In this paper, the kernel selects gaussian kernel:

$$K(x, y) = \exp\left(\frac{-\|x - y\|^2}{2\sigma^2}\right) \quad (9)$$

Where σ is the width of gaussian kernel. For the sake of convenience, we write the FCM algorithm which meeting the formula (7) as KFCM. In the constraint condition of formulae (2), the iterative formulae in KFCM are:

$$u_{ij} = \frac{(1 / (K(x_j, x_j) + K(v_i, v_i) - 2K(x_j, v_i)))^{1/(m-1)}}{\sum_{k=1}^c (1 / (K(x_j, x_j) + K(v_k, v_k) - 2K(x_j, v_k)))^{1/(m-1)}} \quad (10)$$

$$v_i = \frac{\sum_{j=1}^n u_{ij}^m K(x_j, v_i) x_j}{\sum_{j=1}^n u_{ij}^m K(x_j, v_i)}, \forall i = 1, 2, \dots, c \quad (11)$$

The formula (11) shows that v_i is still belongs to the input space, but because of the addition of the weighted coefficient of $K(x_j, v_i)$ (especially the gaussian kernel) , making it to noise and outliers with different weights, this greatly reduces the influence of noise and outliers on the clustering result.

5. KFCMA ALGORITHM PROCESS

In order to describe convenient, the proposed adaptive fuzzy kernel clustering algorithm is abbreviated as KFCMA. Now, we outline the KFCMA algorithm process.

STEP 1. Use adaptive function of clustering number to calculate the optimal clustering number of the sample space data.

STEP 2. Initialization: clustering number c is the result of STEP 1, the termination condition $\varepsilon > 0, m = 2, V^{(0)}$; the iteration counter $l = 0$ and $l_{\max} = 50$; $\sigma = 150$.

STEP 3. According to the formula (10) used the current clustering center to update the membership matrix.

STEP 4. According to the formula (11) used the current clustering center and the membership matrix get from step 3 to update the clustering center.

STEP 5. Judgement and termination: If $|V^{(l)} - V^{(l-1)}| < \varepsilon$ or $l > L_{\max}$, then stop the algorithm and output the membership matrix and clustering center matrix, else go to step 3 with $l = l + 1$.

6. COMPARISON OF SIMULATION RESULTS

In order to verify the effective and feasible of KFCMA, we carried out experiments on artificial data sets and real data sets in Matlab respectively.

6.1. Artificial data

The first set of sample data contains 200 samples in two dimensional space, which belongs to 2 different categories (Each categories has 100 samples).The first category is the point on the circle, the radius of the circle is 10 and the center is (10, 10). The second category is the point inside the square, the center of the square is (10,10) and it's length is 6. There is no public crossing point in the two kind of data set.

The second set of sample data also contains 200 samples in two dimensional space, which also belongs to 2 different categories (Each categories has 100 samples).The first category is the point inside the circle, the radius of the circle is 8 and the center is (8, 8). The second category is the point inside the circle, the center of the circle is (16,16) and it's radius is 4. The two categories of points has 12 crossing points.

We cluster the two data sets respectively using FCM and KFCMA algorithm on the same computer. Comparison results are shown in table 2. From the table 2, we can seen that KFCMA has faster convergence speed, less number of iterations, and more accurate clustering results than FCM in the same termination conditions ϵ .

Table 1. Experimental results of the artificial data(100 randomized trial)

Data set	Algorithm	Average Percentage of Misclass(%)	Average Iterations	Average Time(S)
First set	FCM	0.7	8	0.06
	KFCMA	0.0	2	0.02
Sencond set	FCM	10.5	10	0.08
	KFCMA	0.0	4	0.03

6.2. Real data

The real data contains three data sets from the UCI machine learning database. The three sets of data is: Iris、Wine、Wisc(Wisconsin breast cancer data).

Iris is the best known database to be found in the pattern recognition literature. The data set contains 3 classes of 50 instances each, where each class refers to a type of iris plant. One class is linearly separable from the other 2; the latter are NOT linearly separable from each other.

The Wine data sets are the results of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. The analysis determined the quantities of 13 constituents found in each of the three types of wines.

This breast cancer databases (Wisc) was obtained from the University of Wisconsin Hospitals, Madison from Dr. William H. Wolberg.

Table 2 is the optimal clustering number of the three data sets which is calculated by $C_{opt}(c)$. The results in the table show that using this algorithm can correctly determine the real data sets' cluster number.

We cluster the three data sets respectively using FCM and KFCMA algorithm on the same computer. Comparison results are shown in table 3. From the table 3, we can see that KFCMA has faster convergence speed, less number of iterations, and more accurate clustering results than FCM in the same termination conditions \mathcal{E} .

Table 2. The optimal clustering number of data sets

Data sets	$C_{opt}(c)$	optimal clustering number	actual clustering number
Iris	$C_{opt}(1) = 0, C_{opt}(2) = 10.63$ $C_{opt}(3) = 17.31$	3	3
Wine	$C_{opt}(1) = 0, C_{opt}(2) = 11.81$ $C_{opt}(3) = 10.93, C_{opt}(4 \sim 12) = 0$	2	2
Wisc	$C_{opt}(1) = 0, C_{opt}(2) = 7.91$ $C_{opt}(3) = 6.02, C_{opt}(4 \sim 8) = 0$	2	2

Table 3. The results of KFCMA and FCM algorithm on Real data sets(100 randomized trial)

Algorithm	Data set	Average Percentage of misclass(%)	Average Iterations	Average Time(S)
FCM	Iris	10.667	17	0.1683
	Wine	5.056	18	0.2114
	Wisc	3.367	13	0.5058
KFCM	Iris	10.0	2	0.0877
	Wine	3.371	2	0.1087
	Wisc	2.343	11	0.4520

7. CONCLUSION

Fuzzy clustering algorithm cannot obtain good clustering effect when the sample characteristic is not obvious, very sensitive to noise and need to determine the number of clusters firstly. For those reason, this paper proposes an adaptive fuzzy kernel clustering algorithm. The algorithm firstly use the adaptive function of clustering number to calculate the optimal clustering number, then the samples of input space is mapped to high-dimensional feature space and clustering in the feature space. The Matlab simulation results confirmed that the KFCMA having faster convergence speed, less number of iterations and more accurate clustering results than classical clustering algorithm.

REFERENCES

- [1] MacQueen J. Some methods for classification and analysis of multivariate observations[A]. Proc5th Berkeley Symposium in Mathematics, Statistics, Probability[C]. California,1967. 281-297.
- [2] Bezdek JC.Pattern Recognition with Fuzzy Objective Function Algorithms[M]. New York: Plenum Press, 1981.
- [3] Zhang, D.Q., Chen, S.C. Clustering incomplete data using kernelbased fuzzy C-means algorithm[J]. Neural Process. Lett. 18(3), 155–162 (2003)
- [4] Scholkopf B, Mika S, Burges C. Input space versus feature space in kernelbased methods[J]. IEEE Trans on Neural Networks, 1999, 10(5): 1000-1017.
- [5] Liguozheng, Wangmeng, Zenghuajun. An introduction to support vector machine[M].Beijing: China Machine PRESS, 2004:1-123
- [6] Kamel S M ohamed. New algorithms for solving the fuzzy c-means clustering problem [J].Pattern Recognition, 1994, 27(3): 421-428.
- [7] Blake C, Merz C J. UCI repository of machine learning databases, University of California Irvine. <http://www.ics.uci.edu/~mllearn>
- [8] S.R. Kannan, S. Ramathilagam and P.C. Chung, Effective fuzzy c-means clustering algorithms for data clustering problems, Expert Systems with Applications (2012), 6292–6300.
- [9] P.Y. Mok, H.Q. Huang, Y.L. Kwok, et al., A robust adaptive clustering analysis method for automatic identification of clusters, Pattern Recognition (2012), 3017–3033.
- [10] S. Ramathilagam, R. Devi and S.R. Kannan, Extended fuzzy c-means: an analyzing data clustering problems, Cluster Computing (2013), 389–406.
- [11] S. Ghosha, S. Mitraa and R. Dattagupta, Fuzzy clustering with biological knowledge for gene selection, Applied Soft Computing (2014), 102–111.
- [12] Bijalwan, Vishwanath, et al. "Machine learning approach for text and document mining." arXiv preprint arXiv:1406.1580 (2014).

Author

Weijun Xu(1981 –), male, native place:Fuping- Shaanxi Province. Lecturer. E-mail:xwjsm@163.com.I obtained my B.S. Degree at Yanshan University in China in 2004, and completed my M.S.at Beijing Information and Technology University in China in 2011. Now I'm a teacher in Northeast Petroleum University at Qinhuangdao,China. My research area specializes in fuzzy control and pattern recognition.

