

Kernel FCM Based on Simulated Annealing Algorithm

Chengwei Jiang¹⁺, Xiaoyan Cui², Kaijian Zhang³, Tongliang Li⁴

¹ Department of Automation, Beijing University of Posts and Telecommunications, Beijing, China

² Department of Automation, Beijing University of Posts and Telecommunications, Beijing, China

³ School of Mechanical and Electronic Control Engineering, Beijing Jiaotong University, China

⁴ Department of Automation, Beijing University of Posts and Telecommunications, Beijing, China

Abstract. As everyone knows, birds of a feather flock together, which reveals the essence of clustering. Fuzzy c-means clustering algorithm (FCM) in many algorithms of clustering play a significant role in unsupervised learning. However, FCM still has the following three defects—unknown number of clusters, unstable process of clustering and discoveries of outliers. Aimed at these defects existed in the FCM, a kernel based fuzzy c-means clustering algorithm is proposed to optimize conventional FCM, based on simulated annealing algorithm which takes advantage of simulated annealing algorithm (SA) and kernel technology. The proposed algorithm can improve these shortcomings, but it is not ideal. Our explanation of the results is that the dataset IRIS is relatively simple and the proposed algorithm is an attempt to the original algorithm when realizing global optimization. Our suggestion is that if the dataset contains amount of duplicate data, please apply the proposed algorithm. Anyway, the proposed algorithm is an attempt when traditional ones can't find the global optimization, so it will spend more time.

Keywords: clustering, FCM; kernel methods, simulated annealing, SAKFCM

1. Introduction

Based on similarity principle, the data object into a plurality of classes or clusters, so that cluster similarity in the data object as high as possible similarity between clusters of data objects as low as possible. Currently clustering algorithm can be divided into four categories [1].

1.1. FCM

As the comparison of Kmeans [2], FCM [3] is also a clustering algorithm based on partition. Namely, FCM will similarly be affected by these defects.

FCM was first introduced by Dunn in 1973 [4] and optimized by Bezdek in 1981 [5]. Mahalanobish Distance (MD) maybe the best replacement of Euclidean Distance. Gustafson-Kessel (GK) [6] and Gath-Geva (GG) [7] clustering algorithms realize it. Regrettably, Krishnapuram and Kim [8] proved that MD is not suitable in a fuzzy clustering algorithm. Liu et al. [9] proposed a modification to MD by imposing a restriction on the covariance matrix. In addition to the two linear distance metrics, recently, there has been an increased interest in replacing linear metrics with non-linear measure [10,11,12]. Kernel methods, a non-linear measure, is the best way to optimize the conventional FCM. Kernel based fuzzy c-means (KFCM) [13,14], a proposed algorithm, is widely applied in many occasions, and performs better than the conventional FCM.

1.2. Simulated annealing algorithm

⁺ Corresponding author. Tel.: +86-13691030434
E-mail address: 1406124224@qq.com

As a probabilistic method, Simulated annealing (SA) is a better way to approximate the global optimum of a given function in a large search space. First proposed by Armen G. Khachaturyan, Svetlana V. Semenovskaya, Boris K. Vainshtein in 1979 [15] and then optimized by Armen G. Khachaturyan, Svetlana V. Semenovskaya, Boris K. Vainshtein in 1981 [16], simulated annealing is widely used in many fields. The paper written by them claims that computer simulation mimicking annealing and cooling of such a system could reach its global minimum. The paper written by A. G. Khachaturyan, S.V. Semenovskaya, B.K.Vainstein in 1979 and A. Khachaturyan, S. Semenovskaya, B. Vainshtein in 1981 shows that the simulation can be performed by a solution of kinetic equations for density functions. Independently described by Scott Kirkpatrick, C. Daniel Gelatt and Mario P. Vecchi in 1983 [17], by Vlado Černý in 1985 [18] and by Svetlana V. Semenovskaya, Karen A. Khachaturyan and Armen G. Khachaturyan in 1985 [19], the stochastic sampling method can also perform the simulation. Invented by M.N. Rosenbluth and published by N. Metropolis et al. in 1953[20], the method is an adaptation of the Metropolis–Hastings algorithm, a Monte Carlo method to generate sample states of a thermodynamic system.

Hybridized Tabu Search-Simulated Annealing (HTSSA) [21] algorithm was proposed to solve reactive power problem. The paper written by M. Almarashi, R. John, A. Hopgood, S. Ahmadi [22] reports the use of simulated annealing to design more efficient fuzzy logic systems to model problems with associated uncertainties. Simulated annealing is used within this work as a method for learning the best configurations of interval and general type-2 fuzzy logic systems to maximize their modeling ability. S. Kirkpatrick, C. D. Gelatt, Jr. and M. P. Vecchi [23] use simulated annealing algorithm to solve the NP-complete problem and then to reach the global optimization. Marc Arnaudon and Laurent Miclo [24] propose a stochastic algorithm and its principle is based on simulated annealing and homogenization. The paper written by Harsha Perera, Jack Davis and Tim B. Swartz [25] is use the simulated annealing algorithm to reach the global optimization of the lineup over a vast combinatorial space. The paper written by M. Almarashi, R. John, A. Hopgood, S. Ahmadi [26] use simulated annealing to design more efficient fuzzy logic systems to model problems with associated uncertainties.

2. FCM algorithm and its derivatives

2.1. Fuzzy clustering algorithm(FCM)

In D-dimensional Euclidean space, there is a sample of N observations, which is named as $X=\{x_1, x_2, \dots, x_N\}$, $x_i \in \mathbb{R}^d$. The purpose of clustering algorithm is to separate the X into C subsets, whose cluster centers are defined as $V=\{v_1, v_2, \dots, v_c\}$. The FCM algorithm can be summarized as follow with several steps:

Step 1: set values for the member of cluster C, the maximum time of iterations t_{max} , the degree of fuzziness $m=2$ and the error $\epsilon > 0$.

Step 2: initialize the membership matrix U_0 ;

Step 3: for $t=1, 2, \dots, t_{max}$ do:

Update all centers v_i with Eq. (1):

$$v_i = \frac{\sum_{j=1}^N \mu_{ij}^m x_j}{\sum_{j=1}^N \mu_{ij}^m}, \quad i \in (1, 2, \dots, C). \quad (1)$$

Update all membership values μ_{ij}^t with Eq. (2):

$$\mu_{ij} = \left[\sum_{k=1}^C \left(\frac{x_j - v_i}{x_j - v_k} \right)^{\frac{2}{m-1}} \right]^{-1}, \quad i \in (1, 2, \dots, C), j \in (1, 2, \dots, N). \quad (2)$$

Let $E^t = \max_{ij} \mu_{ij}^t - \mu_{ij}^{t-1}$, if $E^t < \epsilon$ break, else $t=t+1$;

Step 4: obtain the cluster centers.

2.2. Kernel-based fuzzy clustering algorithm(KFCM)

Define a nonlinear transformation as $\varphi: x \rightarrow \varphi(x) \in Y, x \in \mathbb{R}^d$. X is the Euclidean space and Y is the transformed space. In the new space, the purpose of KFCM is to minimize the objective function as follow:

$$J_{KFCM}(U, V) = \sum_{i=1}^C \sum_{j=1}^N \mu_{ij}^m \varphi(x_j) - \varphi(v_i)^2. \quad (3)$$

where

$$\varphi(x_j) - \varphi(v_i)^2 = K(x_j, x_j) + K(v_i, v_i) - 2K(x_j, v_i). \quad (4)$$

Actually, the transformation of $K(x,v)$ is defined as the kernel function $K(x,v) = \varphi(x)T\varphi(x)$. Moreover, we often adopt the Gaussian function as this kernel function. The Gaussian function is described as follow:

$$K(x, v) = \exp\left(\frac{-x - v^2}{\tau^2}\right). \quad (5)$$

Clearly, $K(x,x)=1$, so the Eq.(3) could be simplified as Eq.(6):

$$J_{KFCM}(U, V) = 2 \sum_{i=1}^C \sum_{j=1}^N \mu_{ij}^m (1 - K(x_j, v_i)) \quad (6)$$

Minimize Eq. (6) under the constraints on μ and v as follows:

$$\mu_{ij} = \frac{\left(\frac{1}{1 - K(x_j, v_i)}\right)^{\frac{1}{m-1}}}{\sum_{k=1}^c \left(\frac{1}{1 - K(x_j, v_k)}\right)^{\frac{1}{m-1}}}, j \in (1, 2, \dots, N). \quad (7)$$

and

$$v_i = \frac{\sum_{j=1}^N (\mu_{ij})^m K(x_j, v_i) x_j}{\sum_{j=1}^N (\mu_{ij})^m K(x_j, v_i)}, i \in (1, 2, \dots, C). \quad (8)$$

The kernel-based FCM algorithm is fully described as follow:

Step 1: set $C, t_{max}, m > 1, \varepsilon > 0$;

Step 2: initialize the membership matrix U_0 ;

Step 3: for $t=1, 2, \dots, t_{max}$ do:

Update all centers v_i^t with Eq. (8);

Update all membership values μ_{ij}^t with Eq.(7);

Let $E_t = \max_{ij} \mu_{ij}^t - \mu_{ij}^{t-1}$, if $E_t < \varepsilon$ break, else $t=t+1$;

Step 4: obtain the cluster centers.

The Gaussian function is the special case of RBF functions Eq.(9) as follow:

$$K(x, v) = \exp\left(\frac{-\sum_{i=1}^c x_i^a - v_i^{a,b}}{\tau^2}\right). \quad (9)$$

where $a=1$ and $b=2$.

In addition to Gaussian function or RBF functions, the hyper tangent function Eq. (10) as follow:

$$K(x, v) = 1 - \tanh\left(\frac{-x - v^2}{\tau^2}\right). \quad (10)$$

3. Proposed algorithm

The idea of simulated annealing was proposed by N. Metropolis in 1953 and successfully applied to combinatorial optimization fields by S. Kirkpatrick in 1983. The algorithm of simulated annealing is based on Monte-Carlo iterative strategy for solving stochastic optimization algorithm. Its starting point is based on the physical process of annealing in solids similarity between the composition of the general problem. Simulated annealing algorithm is derived from the solid annealing principle, which heat the solid to a sufficiently high, and cool it slowly. When heated, the solid particles with the internal temperature becomes disordered state. And then internal energy increases, while cooling slowly, particles are gradually orderly. At the equilibrium of each temperature, finally at room temperature, internal energy reduce to the minimum. According to Metropolis principle, the probability of particles tending to equilibrium at the temperature of T is $\exp(-\Delta E/(kT))$, where E is the energy within the temperature of T, ΔE is the increment of E, k is the Boltzmann constant. The simulated annealing is fully described in several steps as follows:

Step1: Set the initial temperature T, initial solution state S, termination temperature T0;

Step2: If $T > T_0$, do the cycle of the implementation of Step 3 to Step 6;

Step3: Search for a new solution S' in the solution space;

Step4: Calculate the increment $\Delta E = C(S') - C(S)$, Where C(S) is the objective function value corresponding to the solution S or is called the evaluation function;

Step5: If $\Delta E < 0$ then accept S' as the new current solution, otherwise accept S' as the new current solution with the probability $\exp(-\Delta E / T)$;

Step6: If the termination condition is satisfied, the current solution is output as the optimal solution, and the program is terminated.

4. Experiments and conclusions

We take advantage of the database IRIS to test the algorithms of KFCM and SAKFCM. The database IRIS consists of 150 objectives with 4 dimensions and 3 clusters. Without priori knowledge, we will use some algorithms to cluster it. The result is showed as follows in Table 1:

Table 1 : Comparing their Cost and Time

	Final Centroids				Cost Function	Time Consume
KFCM	5.88872261	5.88872261	4.3635578	1.39708713	363.4727417	0.420999765
	1.39708713	3.05226292	5.64619541	2.05333085		
	5.00355993	3.40305848	1.48496286	0.25152364		
	6.77461004	3.0522827	5.64628004	2.05336357	363.455733	182.7610002
	5.8887788	2.76108331	4.36363987	1.39712943		
	5.0035601	3.4030558	1.48496741	0.25152569		
	5.00355971	3.40306205	1.48495682	0.25152092	363.4727417	0.420999765
	5.88864789	2.76103614	4.36344866	1.3970309		
	6.77445197	3.05223662	5.64608288	2.05328733		
SAKFCM	5.88919979	2.76123495	4.36425513	1.39744655	363.455733	182.7610002
	6.77511899	3.05243091	5.64691443	2.05360851		
	5.00356137	3.40303567	1.48500156	0.25154107		
	6.77511899	3.05243091	5.64691443	2.05360851	363.4727417	0.420999765
	5.88919979	2.76123495	4.36425513	1.39744655		
	5.00356137	3.40303567	1.48500156	0.25154107		
	5.00356137	3.40303567	1.48500156	0.25154107	363.455733	182.7610002
	5.88919979	2.76123495	4.36425513	1.39744655		
	6.77511899	3.05243091	5.64691443	2.05360851		

From Table 1, we can see that the two algorithms perform very similarly and the proposed algorithm performs a little better than the other. It should be noted that the proposed algorithm is converged to the same centroids and the other not but the similar. Unfortunately, the proposed algorithm consumes more time. In fact, we can adjust the T value to control the time. In the future, we will do more work to reduce the time.

If the data set contains a lot of duplicate data, the proposed algorithm performs well on it but the other worse. The result is as follows in Table 2.

Table 2 : Comparing their Cost with different initial Centroids

	Initial Centroids				Final Centroids				Cost Function
KFCM	5.4	3.4	1.7	0.2	6.386835	2.925579	5.09014	1.76924	635.2916147
	5.1	3.5	1.4	0.3	5.069074	3.312239	1.75785	0.36220	
	5.1	3.5	1.4	0.3	5.069074	3.312239	1.75785	0.36220	
	7.3	2.9	6.3	1.8	6.305033	2.892877	4.96563	1.70371	459.8143901
	7.3	2.9	6.3	1.8	6.305033	2.892877	4.96563	1.70371	
	5.5	2.4	3.8	1.1	5.006793	3.395211	1.50351	0.25915	
	6.5	3	5.2	2	5.069116	3.312164	1.75805	0.36228	635.3267639
	6.4	3.1	5.5	1.8	6.386946	2.925615	5.09030	1.76932	
	6.5	3	5.2	2	5.069116	3.312164	1.75805	0.36228	
SAKFCM	5.4	3.4	1.7	0.2	5.003561	3.403035	1.48500	0.25154	363.455733
	5.1	3.5	1.4	0.3	5.889199	2.761234	4.36425	1.39744	
	5.1	3.5	1.4	0.3	6.775118	3.052430	5.64691	2.05360	
	7.3	2.9	6.3	1.8	6.775118	3.052430	5.64691	2.05360	363.455733
	7.3	2.9	6.3	1.8	5.889199	2.761234	4.36425	1.39744	
	5.5	2.4	3.8	1.1	5.003561	3.403035	1.48500	0.25154	
	6.5	3	5.2	2	6.775118	3.052430	5.64691	2.05360	363.455733
	6.4	3.1	5.5	1.8	5.003561	3.403035	1.48500	0.25154	
	6.5	3	5.2	2	5.889199	2.761234	4.36425	1.39744	

From Table 2, we can find that the proposed algorithm is always converged to the same centroids, but the other is affected by the initial centroids. That is to say, if the dataset has too much duplicate data, the proposed algorithm can well deal with it for the reason that the proposed algorithm has some disturbance. This disturbance is also an attempt to get the global optimization. The more terrible example is that the initial centroids has too much similarity or duplicate data. The proposed algorithm can also deal with this case, but the other can't.

From the results, we can find that:

KFCM is sensitive to the initial centroids. However, the proposed algorithm is not affected by the initial centroids and almost converged to the same final centroids.

KFCM is always reach the local optimization but not the global and the proposed algorithm is just an attempt to get the global.

If not so strict, KFCM can be considered to reduce the time. Terribly, the proposed algorithm will consume more time to find the global one with more wishes.

It should be noted that the simulated annealing starts with a random solution of the problem. The temperature is represented by a variable, which is very high at the beginning and then gradually reduces. During each iteration, the algorithm will randomly select a number in the solution, and then change in a certain direction. If the cost value of the new is lower, the new will become the current solution. However, if the cost value is higher, the new solution may still be current solution, which is an attempt to avoid the local minimum. The algorithm always accepts a better solution and receives a poor solution at the beginning of the annealing process. As the annealing progresses, the algorithm can not accept the poor solution until it can only accept a better solution. In a word, the propose algorithm is an attempt to get the global optimization.

As an attempt, it is necessary to tolerate failure.

5. References

- [1] A.K.Jain, M.N.Murty, P.J. Flynn, Data clustering:are view,ACM Comput. Surv.(CSUR) 31(1999) 264–323
- [2] J. MacQueen, Some methods for classification and analysis of multivariate observations, Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 1 (14) (1967) 281–297
- [3] James C. Bezdek,William Full,FCM—the Fuzzy C-Means clustering algorithm
- [4] J.C. Dunn, A fuzzy relative of the ISODATA process and its use in detecting compact well-Separated Clusters,J. Cybernet.3(3) (1973) 32–57.
- [5] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms,Kluwer Academic Publishers,Norwell,MA, USA, 1981.
- [6] D.Gustafson, W.Kessel, Fuzzy clustering with a fuzzy covariance matrix,in:Proceedings of the IEEE Conference on Decision and Control Including the 17th Symposium on Adaptive Processes,17,1978,pp.761–766.
- [7] I. Gath, A.B.Geva, Unsupervised optimal fuzzy clustering,IEEE Trans.Pattern Anal.Mach.Intell. 11(1989) 773–780.
- [8] R.Krishnapuram, J. Kim, A note on the gustafson-kessel and adaptive fuzzy clustering algorithms,IEEE Trans.Fuzzy Syst.7(1999) 453–461.
- [9] H.-C.Liu, J.-M. Yih, S.-W.Liu, Fuzzy c-mean algorithm based on mahalanobis distances and better initial values,in: Proceedings of the 10th Joint Conference and 12th International Conference on Fuzzy Theory & Technology, 1, World Scientific, 2007, pp. 1398–1404.
- [10] A.Banerjee, S.Merugu, I.S. Dhillon, J. Ghosh, Clustering with bregman divergences,J. Mach.Learn.Res.6(2005) 1705–1749.
- [11] M.Teboulle, A unified continuous optimization framework for center-based clustering methods,J. Mach.Learn.Res.8(2007) 65–102.
- [12] M.Teboulle, P.Berkhin, I. Dhillon, Y.Guan, J. Kogan, Clustering with entropy-like k-means algorithms,in: Grouping Multidimensional Data, Springer, 2006, pp. 127–160.
- [13] S.C.D.Q. Zhang, A novel kernelized fuzzy c-means algorithm with application in medical image segmentation, Artif. Intell. Med. 32 (2004) 37–50.
- [14] D.Z.S.C. Chen, Robust image segmentation using fcm with spatial constraints based on new kernel-induced distance measure, IEEE Trans. Syst. Man Cy. B 34 (2004) 1907–1916.
- [15] Khachaturyan, A.; Semenovskaya, S.; Vainshtein, B. (1979). "Statistical-Thermodynamic Approach to Determination of Structure Amplitude Phases". *Sov.Phys. Crystallography*. 24 (5): 519–524.
- [16] Khachaturyan, A.; Semenovskaya, S.; Vainshtein, B. (1981). "The Thermodynamic Approach to the Structure Analysis of Crystals". *Acta Crystallographica (A37)*: 742–754.
- [17] Kirkpatrick, S.; Gelatt Jr, C. D.; Vecchi, M. P. (1983). "Optimization by Simulated Annealing". *Science*. 220 (4598): 671–680.
- [18] Černý, V. (1985). "Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm". *Journal of Optimization Theory and Applications*. 45: 41–51.
- [19] Semenovskaya, S.; Khachaturyan, K.; Khachaturyan, A. (1985). "Statistical Mechanics Approach to the Determination of a Crystal". *Acta Crystallographica (A41)*: 268–273.
- [20] Metropolis, Nicholas; Rosenbluth, Arianna W.; Rosenbluth, Marshall N.; Teller, Augusta H.; Teller, Edward (1953). "Equation of State Calculations by Fast Computing Machines". *The Journal of Chemical Physics*. 21 (6): 1087.
- [21] K Lenin, BR Reddy, M Suryakalavathi, Hybrid Tabu search-simulated annealing method to solve optimal reactive power problem.
- [22] M Almarashi, Robert John, A Hopgood, S Ahmadi, Learning of Interval and General Type-2 Fuzzy Logic Systems using Simulated Annealing: Theory and Practice
- [23] S. Kirkpatrick, C. D. Gelatt and M. P. Vecchi, Optimization by Simulated Annealing, Science, New Series, Vol. 220, 1983, pp. 671-680.
- [24] K Lenin, BR Reddy, M Suryakalavathi, Hybrid Tabu search-simulated annealing method to solve optimal reactive power problem
- [25] Harsha Perera, Jack Davis and Tim B. Swartz, Assessing the Impact of Fielding in Twenty20 Cricket
- [26] M. Almarashi, R. John, A. Hopgood, S. Ahmadi, Designing generalized type-2 fuzzy logic systems using interval type-2 fuzzy logic systems and simulated annealing