

## On Utilization of K-Means for Determination of *q*-Parameter for Tsallis-Entropy-Maximized-FCM

### Makoto Yasuda

Department of Electrical and Computer Engineering, National Institute of Technology, Gifu College, Motosu, Japan Email: yasuda@gifu-nct.ac.jp

How to cite this paper: Yasuda, M. (2017) On Utilization of K-Means for Determination of *q*-Parameter for Tsallis-Entropy-Maximized-FCM. *Journal of Software Engineering and Applications*, **10**, 605-624. https://doi.org/10.4236/jsea.2017.107033

**Received:** April 14, 2017 **Accepted:** June 20, 2017 **Published:** June 23, 2017

Copyright © 2017 by author and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). http://creativecommons.org/licenses/by/4.0/

co Open Access

## Abstract

In this paper, we consider a fuzzy c-means (FCM) clustering algorithm combined with the deterministic annealing method and the Tsallis entropy maximization. The Tsallis entropy is a q-parameter extension of the Shannon entropy. By maximizing the Tsallis entropy within the framework of FCM, membership functions similar to statistical mechanical distribution functions can be derived. One of the major considerations when using this method is how to determine appropriate q values and the highest annealing temperature,  $T_{high}$ , for a given data set. Accordingly, in this paper, a method for determining these values simultaneously without introducing any additional parameters is presented. In our approach, the membership function is approximated by a series of expansion methods and the K-means clustering algorithm is utilized as a preprocessing step to estimate a radius of each data distribution. The results of experiments indicate that the proposed method is effective and both q and  $T_{high}$  can be determined automatically and algebraically from a given data set.

## **Keywords**

Fuzzy c-Means, K-Means, Tsallis Entropy, Entropy Maximization, Entropy Regularization, Deterministic Annealing

## **1. Introduction**

Techniques from statistical mechanics can be used for the investigation of the macroscopic properties of a physical system consisting of many elements. Recently, research activities utilizing statistical mechanical models or techniques for information processing have become increasingly popular.

Rose *et al.* [1] [2] proposed deterministic annealing (DA) as a deterministic variant of simulated annealing (SA) [3]. In DA, the minimization problem for an objective function is treated as the minimization of the free energy of a system. The DA approach tracks the function's minimum with decreasing the system temperature, thus allowing the deterministic optimization of the objective function at each temperature. Hence, DA is more efficient than SA, but does not guarantee that the solution is the global optimal solution. From the viewpoint of statistical mechanics, the membership functions of the fuzzy c-means (FCM) clustering [4] with maximum entropy or entropy regularization methods [5] [6] can be seen as distribution functions from statistical mechanics. For example, FCM maximized with the Shannon entropy gives a membership function similar to the Boltzmann distribution function [1].

Tsallis [7], inspired by multi-fractal, non-extensively extended the Boltzmann-Gibbs statistics by postulating a generalized form of the entropy (the Tsallis entropy) with a generalization parameter *q*. The Tsallis entropy is proved to be applicable to the numerous systems [8] [9]. In the field of fuzzy clustering, a membership function was derived by maximizing the Tsallis entropy within the framework of FCM [10] [11] [12]. This membership function has a similar form to the statistical mechanical distribution function, and is suitable for use with annealing methods because it contains a parameter corresponding to the system temperature. Accordingly, the Tsallis entropy maximized FCM was successfully combined with the DA method as Tsallis-DAFCM in [13].

One of the major challenges with using Tsallis-DAFCM is the determination of an appropriate value for q and the highest (or initial) annealing temperature,  $T_{high}$ , for a given data set. Especially, the determination of a suitable q value is a fundamental problem for systems where the Tsallis entropy is applied. Even in physics, quite a few systems are known in which q is calculable. In the previous study [13], the values were experimentally determined, and only roughly optimized.

Accordingly, we presented a method that can determine both q and  $T_{high}$  simultaneously from a given data set without introducing additional parameters [14]. The membership function of Tsallis-DAFCM was approximated by a series expansion to simplify the function. Based on this simplified formula, both q and  $T_{high}$  could be estimated along with the membership function for a given data set. However, it was also found that the results from this method depend on the estimation of the radius of the distribution of the data or the location of clusters.

To overcome this difficulty, in this study, we propose a method that utilizes K-means [15] as a preprocessing step of the approximation method. That is, a data set is clustered by K-means roughly. We then estimate the radius of the distribution of the data set, and apply the approximation method to determine q and  $T_{high}$ .

Experiments are performed on numerical data and the Iris Data Set [16], and the results show that the proposed method can be used to determine q and

 $T_{high}$  automatically and algebraically from a data set. It is also confirmed that the data can be partitioned into clusters appropriately using these parameters.

#### 2. FCM with Tsallis Entropy Maximization

Let  $X = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} (\mathbf{x}_k = (x_k^1, \dots, x_k^p) \in \mathbb{R}^p)$  be a data set in *p*-dimensional real space, and let  $V = \{\mathbf{v}_1, \dots, \mathbf{v}_c\} (\mathbf{v}_i = (v_i^1, \dots, v_i^p))$  be the *c* distinct clusters. Let  $u_{ik} \in [0,1] (i = 1, \dots, c; k = 1, \dots, n)$  be the membership function, and let

$$I = \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ik}^{m} d_{ik} \left( 1 < m \right)$$
(1)

be the objective function of FCM, where  $d_{ik} = \| \mathbf{x}_k - \mathbf{v}_i \|$ .

On the other hand, the Tsallis entropy is defined as

$$S_q = -\frac{1}{q-1} \left( \sum_i p_i^q - 1 \right) \tag{2}$$

where  $p_i$  is the probability of the *i*th event and q is a real number [7]. The Tsallis entropy reaches the Shannon entropy as  $q \rightarrow 1$ .

Next, we apply the Tsallis entropy maximization method to FCM [12] [13]. First, Equation (2) is rewritten as

$$S_{q} = -\frac{1}{q-1} \left( \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ik}^{q} - 1 \right)$$
(3)

Then, the objective function in Equation (1) is rewritten as

$$J_{q} = \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ik}^{q} d_{ik}$$
(4)

Under the normalization constraint of

$$\sum_{i=1}^{c} u_{ik} = 1 \; \bigl( \forall k \bigr) \tag{5}$$

the Tsallis entropy functional becomes

$$\delta S_q - \sum_{k=1}^n \alpha_k \delta \left( \sum_{i=1}^c u_{ik} - 1 \right) - \beta \sum_{k=1}^n \sum_{i=1}^c \delta \left( u_{ik}^q d_{ik} \right)$$
(6)

where  $\alpha_k$  and  $\beta$  are the Lagrange multipliers. By applying the variational method, the stationary condition for the Tsallis entropy functional yields the following membership function for Tsallis-FCM [12]:

$$u_{ik} = \frac{\left\{1 - \beta \left(1 - q\right) d_{jk}\right\}^{\frac{1}{1 - q}}}{Z}$$
(7)

where

$$Z = \sum_{j=1}^{c} \left\{ 1 - \beta \left( 1 - q \right) d_{jk} \right\}^{\frac{1}{1-q}}$$
(8)

From Equation (7), the expression for  $v_i$  becomes

$$\mathbf{v}_{i} = \frac{\sum_{k=1}^{n} u_{ik}^{a} d_{ik}}{\sum_{k=1}^{n} u_{ik}^{a}}$$
(9)

#### **3. Approximation of Membership Function**

The performance of Tsallis-DAFCM is superior to those of other entropy-based-FCM methods [12]. However, it is still unknown how to determine an appropriate q value and a highest annealing temperature  $T_{high}$  for a given data set. To tackle this problem, we first simplify the membership function using a series expansion.

### **3.1. Series Expansion of** $u_{ik}$

 $u_{ik}$  in Equation (7) can be expanded to a power of  $\beta$  as follows:

$$u_{ik} = \sum_{n=0}^{\infty} \frac{1}{n!} \frac{\partial^n u_{ik}(0)}{\partial \beta^n} \beta^n$$
(10)

When the temperature is high enough, if the series expansion up to the third order terms is used, Equation (10) becomes

$$u_{ik}' = \frac{1}{c} + \frac{-cd_{ik} + L_{1k}}{c^2}\beta + \frac{c^2qd_{ik}^2 - cqL_{2k} - 2cd_{ik}L_{1k} + 2L_{1k}^2}{2c^3}\beta^2$$
(11)

where

$$L_{1k} = \sum_{j=1}^{c} d_{jk},$$

$$L_{2k} = \sum_{j=1}^{c} d_{jk}^{2}.$$
(12)

#### 3.2. Determination of q and $T_{high}$

Based on the results in Section 3.1, we propose a method for determining both q and  $\beta$  simultaneously.

First, to ensure the convergence of Equation (10), we use the following expression for  $\beta$ :

$$\overline{L}_{1k} = \frac{1}{n} \sum_{k=1}^{n} L_{1k},$$

$$\beta = \frac{1}{N} \sum_{l=1}^{N} \frac{c}{\overline{L}_{lk}^{(l)}},$$
(13)

where N and (l) denote the maximum number of iterations, and the number of iterations to be used in the calculation of  $\overline{L}_{ik}$ , respectively.  $T_{high}$  can be calculated as  $T_{high} = 1/\beta$ .

Then, setting  $v_i = 0$  and replacing  $x_k$  with the continuous variable x, Equation (11) becomes

$$u'(\mathbf{x}) = \frac{q\beta^{2}}{2c} \|\mathbf{x}\|^{4} - \frac{(c + L_{1}(\mathbf{x})\beta)\beta}{c^{2}} \|\mathbf{x}\|^{2} + \frac{1}{c} + \frac{(2cL_{1}(\mathbf{x}) - cqL_{2}(\mathbf{x})\beta + 2L_{1}^{2}(\mathbf{x})\beta)\beta}{2c^{3}}$$
(14)

where



$$L_{1}(\mathbf{x}) = \sum_{j=1}^{c} \|\mathbf{v}_{j} - \mathbf{x}\|^{2},$$

$$L_{2}(\mathbf{x}) = \sum_{j=1}^{c} \|\mathbf{v}_{j} - \mathbf{x}\|^{4}.$$
(15)

From this equation, q can be determined as follows. By designating the range of the dataset as  $R = (R^1, \dots, R^p)$ , the maximum range of the distribution  $R_{\text{max}}$  is defined as

$$R_{\max} = R^{\Theta} \left( \Theta = \underset{1 \le \theta \le p}{\operatorname{arg\,max}} R^{\theta} \right)$$
(16)

Furthermore, by assuming that the radius r of each cluster is between  $R_{\max}/2c \le r \le R_{\max}/2$ , and  $u'(\mathbf{x})$  tends to  $u_0$  at  $\mathbf{x}' = (x^1 = 0, \dots, x^{\Theta} = r, \dots, x^p = 0)$ , Equation (14) can be solved for q. Consequently, we have the following formula for q.

$$\mathcal{G}(r) = \frac{2}{c\left(cr^{4} - L_{2}\left(\boldsymbol{x}'\right)\right)\beta} \left\{ c\left(c + L_{1}\left(\boldsymbol{x}'\right)\beta\right)r^{2} \right\} - \frac{2}{c\left(cr^{4} - L_{2}\left(\boldsymbol{x}'\right)\right)\beta} \left\{ 2L_{1}\left(\boldsymbol{x}'\right) + L_{1}^{2}\left(\boldsymbol{x}'\right)\beta \right\}, \quad (17)$$

$$q = \mathcal{G}(r).$$

It should be noted that in this equation, for simplicity,  $u_0$  is set to

$$u_0 = \frac{1}{c} \tag{18}$$

because Equation (7) tends to 1/c as  $\|\boldsymbol{x}_k\|$  goes to  $\infty$ .

#### 4. Proposed Algorithm

By combining the method presented in the previous section with Tsallis-DAFCM, we proposed the following fuzzy c-means clustering algorithm [14]. In this algorithm, the number of clusters in the data is assumed to be known in advance.

In the first algorithm shown in Figure 1, the parameters q and  $\beta = 1/T_{high}$  for a given data set are determined (N is the maximum number of iteration. In Equation (17),  $L_1(\mathbf{x}')$  and  $L_2(\mathbf{x}')$  are approximated by  $\overline{L}_{1k}$  and  $\overline{L}_{2k}$ , respectively.).

The second algorithm is the conventional Tsallis-DAFCM algorithm [12].

1) Set the temperature reduction rate  $T_r$ , and the thresholds for convergence  $\delta_1$  and  $\delta_2$ .

2) Generate c initial clusters at random locations. Set the current temperature T to  $1/\beta$ .

3) Calculate  $u_{ik}$  using Equation (7).

4) Calculate the cluster centers using Equation (9).

5) Compare the difference between the current centers and the centers of the previous iteration obtained using the same temperature  $v'_i$ . If the convergence condition  $\max_{1 \le i \le c} ||v_i - v'_i|| < \delta_1$  is satisfied, then go to Step 2.6. Otherwise re-



Figure 1. Processing flow of the conventional method.

turn to Step 2.3.

6) Compare the difference between the current centers and the centers of the previous iteration obtained using a lower temperature v''. If the convergence condition  $\max_{1 \le i \le c} \|v_i - v''_i\| < \delta_2$  is satisfied, then stop. Otherwise decrease the temperature;  $T = T \cdot T_r$ , and return to Step 2.3.

The experimental results in [14] confirmed that the first algorithm can determine  $\beta$  desirably. However, they also revealed that q from this algorithm strongly depends on the estimation of the radius r in Equation (17). Accordingly, as shown in **Figure 2**, the first algorithm is divided in two parts. The first one determines  $\beta$ . In the second part, the K-means algorithm is utilized to calculate r by assuming that each data point belongs to its nearest cluster.

## **5. Experiments**

To examine the effectiveness of the proposed algorithm, we conducted two experiments.



Figure 2. Processing flow of the proposed method.

#### 5.1. Experiment 1

The first experiment examined whether appropriate q and  $\beta = 1/T_{high}$  values can be determined for a given data set, and the relation between the number of iterations N and the parameters q and  $\beta$ .

In this experiment, data sets containing (a) three clusters and (b) five clusters were used, as shown in **Figure 3**. Each cluster follows a normal distribution, and contains 2, 250 data points.

Dependencies of the maximum, minimum, mean and standard deviation of  $\beta$ , a mean radius of the data distribution and q for Figure 3(a) on the number of iterations N are summarized in Table 1, Table 2 and Table 7. Figure 4 shows the plots of the maximum, minimum, and mean of q. In these tables,  $r_{\text{max}}$  and  $r_{\text{mean}}$  denote  $R_{\text{max}}/2$  and the mean of  $R_{\text{max}}/2c$  and  $R_{\text{max}}/2$ , respectively.  $r_{\text{max}}^{k}$  and  $r_{\text{mean}}^{k}$ , on the other hand, denote the maximum and mean radius of the distribution obtained by K-means, respectively.

In **Table 7**, the value of q for  $r_{\text{max}}$  for example is calculated using Equation (17) as  $q = \vartheta(r_{\text{max}})$ . Based on the results in **Table 1**, the value of q was calculated by



**Figure 3.** Numerical data ( $C_x$  denotes the cluster number). (a) c = 3; (b) c = 5.



**Figure 4.** Maximum, minimum, and mean of q (c = 3,  $\beta = 5.35$  le -06).

| N      | Maximum   | Minimum   | Mean      | Std. deviation |
|--------|-----------|-----------|-----------|----------------|
| 10     | 5.624e-06 | 3.877e-06 | 5.147e-06 | 6.568e-07      |
| 100    | 6.098e-06 | 3.446e-06 | 5.350e-06 | 6.024e-07      |
| 1000   | 6.121e-06 | 2.793e-06 | 5.353e-06 | 5.775e-07      |
| 10,000 | 6.207e-06 | 2.497e-06 | 5.351e-06 | 5.914e-07      |
|        |           |           |           |                |

**Table 1.** Maximum, minimum, mean, and standard deviation of  $\beta$  (c = 3).

**Table 2.** Maximum, minimum, mean, and standard deviation of  $r_{\text{max}}^k$  and  $r_{\text{mean}}^k$  (c = 3).

|                           | Ν      | Maximum | Minimum | Mean  | Std. deviation |
|---------------------------|--------|---------|---------|-------|----------------|
|                           | 5      | 200.4   | 196.8   | 198.8 | 1.7            |
|                           | 10     | 204.6   | 196.8   | 200.1 | 2.2            |
| $r_{_{ m max}}^{^{ m k}}$ | 100    | 203.8   | 195.7   | 198.8 | 1.7            |
|                           | 1000   | 305.9   | 195.0   | 199.3 | 6.2            |
|                           | 10,000 | 313.4   | 194.8   | 199.6 | 8.1            |
|                           | 5      | 197.5   | 190.3   | 192.0 | 2.8            |
| $r_{\rm mean}^{ m k}$     | 10     | 193.3   | 190.7   | 193.8 | 2.7            |
|                           | 100    | 198.3   | 189.5   | 193.1 | 3.0            |
|                           | 1000   | 292.8   | 188.1   | 194.1 | 8.8            |
|                           | 10,000 | 298.0   | 188.2   | 194.0 | 8.6            |

fixing  $\beta$  to its mean value 5.351e–06.  $R_{\text{max}}$ ,  $r_{\text{max}}$ , and  $r_{\text{mean}}$  for Figure 3(a) are 860.0, 430.0 and 286.7, respectively.

From **Table 1**, it can be seen that the maximum of  $\beta$  tends to increase and the minimum of  $\beta$  tends to decrease with increasing *N*. However, when *N* become 100 or more, the mean of  $\beta$  does not depend on *N*.

From **Table 2**, it can be seen that the mean of  $r_{max}^{k}$  and  $r_{mean}^{k}$  hardly depends on *N*, though the standard deviation becomes larger when *N* become 1, 000 or more. This is caused by a very seldom misclassification of K-means.

Comparing the results in **Table 7**, it can be found that, when r is set to  $r_{\text{max}}^k$  or  $r_{\text{mean}}^k$ , *q* has smaller standard deviations, and the magnitude of the change in the mean values of *q* is comparatively small. This shows that *q* can be calculated stably by performing K-means first. It is also can be found that the maximum of *q* increases with increasing *N*, because of the random locations of clusters. Even though  $r_{\text{max}}$  overestimates the mean radius of the clusters, clustering can be performed properly in this case.

Accordingly, q has little impact on clustering in this experiment.

Dependencies of the maximum, minimum, mean and standard deviation of  $\beta$ , a mean radius of the data distribution and q for Figure 3(b) on the number of iterations N are summarized in Table 3, Table 4 and Table 8. Figure 5 shows the plots of the maximum, minimum, and mean of q.



**Figure 5.** Maximum, minimum, and mean of q (c = 5,  $\beta = 3.608e - 06$ ).

**Table 3.** Maximum, minimum, mean, and standard deviation of  $\beta$  (*c* = 5).

| Ν      | Maximum   | Minimum   | Mean      | Std. deviation |
|--------|-----------|-----------|-----------|----------------|
| 10     | 5.878e-06 | 2.686e-06 | 4.030e-06 | 9.264e-07      |
| 100    | 5.088e-06 | 2.466e-06 | 3.618e-06 | 5.801e-07      |
| 1000   | 6.738e-06 | 2.316e-06 | 3.608e-06 | 6.117e-07      |
| 10,000 | 7.060e-06 | 2.118e-06 | 3.608e-06 | 6.320e-07      |
|        |           |           |           |                |

**Table 4.** Maximum, minimum, mean, and standard deviation of  $r_{max}^{k}$  and  $r_{mean}^{k}$  (c = 5).

|                               | Ν      | Maximum | Minimum | Mean  | Std. deviation |
|-------------------------------|--------|---------|---------|-------|----------------|
|                               | 5      | 343.2   | 116.1   | 236.0 | 99.0           |
| $r_{\max}^k$                  | 10     | 353.4   | 116.1   | 229.4 | 93.7           |
|                               | 100    | 356.4   | 116.0   | 193.8 | 91.0           |
|                               | 1000   | 394.2   | 115.7   | 203.8 | 92.3           |
|                               | 10,000 | 395.0   | 115.7   | 198.4 | 91.6           |
|                               | 5      | 189.6   | 115.3   | 150.4 | 33.2           |
|                               | 10     | 151.7   | 115.1   | 121.9 | 13.1           |
| <b>r</b> <sup>k</sup><br>mean | 100    | 190.2   | 115.2   | 134.7 | 24.5           |
|                               | 1000   | 249.5   | 115.1   | 138.3 | 30.0           |
|                               | 10,000 | 279.5   | 115.0   | 134.4 | 29.5           |

 $R_{\text{max}}$ ,  $r_{\text{max}}$ , and  $r_{\text{mean}}$  for Figure 3(b) are 860.0, 430.0 and 258.0, respectively. Based on the results in Table 3, the value of q was calculated by fixing  $\beta$ to 3.608e-06.

Comparing these results with those in Table 1, Table 2 and Table 7, it can be found that q for c=5 has larger standard deviations than those for c=3. This is caused by an increase in the number of combinations of data points and clusters.



In **Table 8**, it can be seen that q for  $r_{\text{max}}^k$  has the largest standard deviations. This is considered to be caused by the significant standard deviations of  $r_{\text{max}}^k$  shown in **Table 2**, suggesting a variation of the estimation of the radius of the distribution. On the other hand, q for  $r_{\text{mean}}^k$  has the smallest standard deviations.

Substituting the values of  $\beta$  and q in **Table 3** and **Table 8** directly, **Figure 6** and **Figure 7** compare the membership function for the cluster  $C_2$ ,

$$u_{2}(\mathbf{x}) = \frac{\left\{1 - \beta(1 - q) \|\mathbf{x} - \mathbf{v}_{2}\|\right\}^{\frac{1}{1 - q}}}{\sum_{j=1}^{5} \left\{1 - \beta(1 - q) \|\mathbf{x} - \mathbf{v}_{j}\|^{2}\right\}^{\frac{1}{1 - q}}}$$
(19)

with

$$u_{2}'(\mathbf{x}) = \frac{q\beta^{2}}{2c} \|\mathbf{x} - \mathbf{v}_{2}\|^{4} - \frac{\left(c + L_{1}(\mathbf{x} - \mathbf{v}_{2})\beta\right)\beta}{c} \|\mathbf{x} - \mathbf{v}_{2}\|^{2} + \frac{1}{c} + \frac{\left(2cL_{1}(\mathbf{x} - \mathbf{v}_{2}) - cqL_{2}(\mathbf{x} - \mathbf{v}_{2}) + 2L_{1}^{2}(\mathbf{x} - \mathbf{v}_{2})\beta\right)\beta}{2c^{3}}.$$
(20)

for c = 5,  $r = r_{\text{max}}^{k}$  and  $r_{\text{mean}}^{k}$ , and N = 10 and 10,000. In the equations,  $v_{j}$  is set to each of the cluster coordinates in **Figure 3(b)**. The data projections on the *xz* and *yz* planes are also plotted.

The figures show no significant difference between  $u_2(\mathbf{x})$  and  $u'_2(\mathbf{x})$  and between  $r^k_{\max}$  and  $r^k_{\max}$ .



**Figure 6.** Comparisons of the membership functions calculated by Equations (19) and (20) (c = 5,  $r = r_{max}^{k}$ , N = 1,010,000).



**Figure 7.** Comparisons of the membership functions calculated by Equations (19) and (20) (c = 5,  $r = r_{mean}^{k}$ , N = 1,010,000).

Compared with the clusters in Figure 3(a), those in Figure 3(b) are not aligned in a straight line. However, the results for c = 5 are as accurate as those for c = 3. As a result, the maximum error factor is considered to be  $\beta$ . Since the clusters in Figure 3(a) are aligned in a straight line,  $\beta$  cannot be determined optimally by locating clusters randomly as does in the algorithm in Figure 1.

From these results, it can be confirmed that N = 100 is sufficient to determine both  $\beta$  and q for the data sets in Figure 1.

#### 5.2. Experiment 2

In this experiment, the Iris Data Set [16], which comprises data from 150 iris flowers with four-dimensional vectors, is used. The three clusters to be detected are Versicolor, Virginia and Setosa, and the parameters in the algorithm in **Figure 2** are set as follows:  $\delta_1 = 0.02$ , and  $\delta_2 = 0.01$ , and  $T_r = 0.8$ .

 $R_{\rm max}$ ,  $r_{\rm max}$ , and  $r_{\rm mean}$  are 5.90, 2.95 and 1.97, respectively.

#### 5.2.1. Determination of Parameters

The maximum, minimum, mean, and standard deviation of  $\beta$ ,  $r_{\text{max}}^k$ ,  $r_{\text{mean}}^k$  and q are summarized in **Table 5**, **Table 6** and **Table 9**. Figure 8 shows the plots of the maximum, minimum, and mean of q. Based on the results in **Table 5**, the value of q was calculated by fixing  $\beta$  to 1.076e-01.

From Table 5, it can be seen that a dependency of  $\beta$  on N is same as those in Table 1 and Table 3. Table 6 shows that the mean of  $r_{\text{max}}^k$  and  $r_{\text{mean}}^k$ 



**Figure 8.** Maximum, minimum, and mean of q for the Iris data set ( $\beta = 1.076e - 01$ ).

**Table 5.** Maximum, minimum, mean, and standard deviation of  $\beta$  for the Iris data set.

| N      | Maximum   | Minimum   | Mean      | Std. deviation |
|--------|-----------|-----------|-----------|----------------|
| 10     | 1.455e-01 | 8.040e-02 | 1.097e-01 | 1.554e-02      |
| 100    | 1.455e-01 | 7.409e-02 | 1.081e-01 | 1.765e-02      |
| 1000   | 1.810e-01 | 5.978e-02 | 1.075e-01 | 1.872e-02      |
| 10,000 | 1.857e-01 | 5.893e-02 | 1.076e-01 | 1.949e-02      |

**Table 6.** Maximum, minimum, mean, and standard deviation of  $r_{max}^{k}$  and  $r_{mean}^{k}$  for the Iris data set.

|                       | N      | Maximum | Minimum | Mean  | Std. deviation |
|-----------------------|--------|---------|---------|-------|----------------|
|                       | 5      | 3.855   | 3.855   | 3.855 | 0.000          |
|                       | 10     | 3.855   | 3.855   | 3.855 | 0.000          |
| $r_{\max}^k$          | 100    | 4.935   | 3.855   | 3.866 | 0.107          |
|                       | 1000   | 4.935   | 3.855   | 3.861 | 0.083          |
|                       | 10,000 | 4.935   | 3.855   | 3.862 | 0.085          |
|                       | 5      | 2.066   | 2.066   | 2.066 | 0.000          |
|                       | 10     | 2.066   | 2.066   | 2.066 | 0.000          |
| $r_{\rm mean}^{ m k}$ | 100    | 2.066   | 1.849   | 2.064 | 0.022          |
| inai                  | 1000   | 2.066   | 1.849   | 2.063 | 0.026          |
|                       | 10,000 | 2.066   | 1.849   | 2.063 | 0.027          |
|                       |        |         |         |       |                |

can be calculated regardless of the value of N.

**Table 9** shows that the standard deviations of q for  $r_{\text{max}}^k$  and  $r_{\text{mean}}^k$  are smaller than those of  $r_{\text{max}}$  and  $r_{\text{mean}}$  showing the effectiveness of the proposed method.

It can be found that these tables show that the proposed method gives similar results to those in the Section 5.1, and N = 5 to 10 is sufficient to determine  $\beta$ ,  $r_{\text{max}}^{\text{k}}$ ,  $r_{\text{mean}}^{\text{k}}$ , and q. In the algorithm shown in **Figure 1**, it is unnecessary to

repeat the calculations of the means of  $\beta$  and q the same number of times N .

It is also found that not only the estimations of the radius are important to improve the accuracy because  $r_{mean}^{k}$  gives superior result compared with those of  $r_{max}^{k}$ . For this reason, a preprocessing method that can estimate the location of clusters quickly, such as the Canopy method [17] might be suitable for the proposed method to be more effective.

#### 5.2.2. Clustering Accuracy

The maximum and mean number of data points misclassified by the previous method [14], the proposed method, and Tsallis-DAFCM in 1, 000 trials are summarized in Table 10 and Figure 9.  $T_{high} = 1/\beta$  is fixed to 1/1.076e-01 =9.294. In Tsallis-DAFCM, as a typical value, q is changed from 1.2 to 2.8.

Even though the experiment was repeated 1000 times, the results obtained with the proposed method were almost identical.

By comparing the mean number of misclassified data points of the proposed method with those of the previous method, it can be confirmed the results from both methods are not significantly different when  $r = r_{max}$  and  $r = r_{max}^{k}$  or when  $r = r_{\text{mean}}$  and  $r = r_{\text{mean}}^{k}$ .

By comparing the mean number of misclassified data points of the proposed method with those of Tsallis-DAFCM, it can be confirmed the proposed method can get slightly better results. By examining the maximum number of misclassified, we see that Tsallis-DAFCM misclassifies data more often than does the proposed method.

These results confirm that appropriate values of q and  $T_{high} = 1/\beta$  for the Iris Data Set can be estimated by the proposed method. Setting  $r = r_{mean}^{k}$  is most suitable for this data set.

#### 5.2.3. Computational Time

**Figure 10** compares the mean of computational times of  $\beta$  and q, and clus-



Figure 9. Maximum, minimum, and mean numbers of misclassified data points for the Iris Data Set of the previous method, the proposed method and Tsallis-DAFCM  $(\beta = 1.076e - 01, T_{high} = 9.294).$ 





**Figure 10.** Mean of computational times of  $\beta$ , q, and clustering for the Iris Data Set. (a)  $\beta$ ; (b) q; (c) Clustering.

tering in 1000 trials (Executions were conducted on an Intel(R) Core(TM)2 Duo CPU E6550 @ 2.33 GHz).

**Figure 10(a)** shows that the computational time of  $\beta$  does not depend on r and increases proportionally to N because, as can be seen from Equations

(12) and (13), the value of  $\beta$  is determined independently of r and  $\overline{L}_{\mu}$  is calculated N times.

**Figure 8(b)**, on the other hand, shows that the calculation of q for  $r_{max}^{k}$ sometimes takes time suggesting that, in this case,  $r_{max}^{k}$  becomes too large to give an appropriate q value.

**Figure 10(c)** shows that that when r is set to  $r_{mean}$  or  $r_{mean}^{k}$ , clustering can be conducted quickly and stably.

#### 5.3. Evaluation of the Proposed Algorithm

From the experimental results in 5.1 and 5.2, the effectiveness of the proposed algorithm using K-means can be evaluated as follows:

- 1)  $r_{max}^{k}$  and  $r_{mean}^{k}$  can be obtained with very small variances without consuming much computational time;
- 2) q can be determined with a very small variance using  $r_{mean}^{k}$  without consuming;
- 3) Much computational time;
- 4) The numerical data sets and the Iris Data Set can be clustered desirably using  $r_{\text{mean}}^{k}$ .

#### 6. Conclusions

The Tsallis entropy is a q-parameter extension of the Shannon entropy. FCM with the Tsallis entropy maximization has a proper characteristic for clustering, especially when it is combined with DA as Tsallis-DAFCM. The extent of its membership function strongly depends on the parameter q and the initial annealing temperature  $T_{high}$ .

In this study, we proposed a method for approximating the membership function of Tsallis-DAFCM which, by using the K-means method as a preprocessing step, determines q and  $\beta = 1/T_{high}$  automatically and algebraically from a given data set.

Experiments were performed on the numerical data sets and the Iris Data Set, and showed that the proposed method can more accurately and stably determine q and  $\beta$  algebraically than the previous method without consuming much computational time. It was also confirmed that the data can be partitioned into clusters appropriately using these parameters.

In the future, as described in 5.1, we first intend to explore ways to improve the accuracy of the estimates for  $\beta$  and q by using other rough clustering methods. We then intend to examine the effectiveness of the method using very complicated real world data set [18].

#### References

- Rose, K. (1998) Deterministic Annealing for Clustering, Compression, Classifica-[1] tion, Regression, and Related Optimization Problems. Proceedings of the IEEE, 86, 2210-2239. https://doi.org/10.1109/5.726788
- Rose, K., Gurewitz, E. and Fox, B.C. (1990) A Deterministic Annealing Approach to [2]



Clustering. Pattern Recognition Letters, 11, 589-594.

- [3] Kirkpatrick, S., Gelatt, C.D. and Vecchi, M.P. (1983) Optimization by Simulated Annealing. *Science*, 220, 671-680. <u>https://doi.org/10.1126/science.220.4598.671</u>
- Bezdek, J.C. (1981) Pattern Recognition with Fuzzy Objective Function Algorithms. Prenum Press, New York. <u>https://doi.org/10.1007/978-1-4757-0450-1</u>
- [5] Honda, K. and Ichihashi, H. (2007) A Regularization Approach to Fuzzy Clustering with Nonlinear Membership Weights. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 11, 28-34. https://doi.org/10.20965/jaciii.2007.p0028
- [6] Kanzawa, Y. (2012) Entropy-Based Fuzzy Clustering for Non-Euclidean Relational Data and Indefinite Kernel Data. *Journal of Advanced Computational Intelligence* and Intelligent Informatics, 16, 784-792. https://doi.org/10.20965/jaciii.2012.p0784
- Tsallis, C. (1988) Possible Generalization of Boltzmann-Gibbs Statistics. Journal of Statistical Physics, 52, 479-487. <u>https://doi.org/10.1007/BF01016429</u>
- [8] Abe, S. and Okamoto, Y. (2001) Nonextensive Statistical Mechanics and Its Applications. Springer, Berlin.
- [9] Gell-Mann, M. and Tsallis, C. (2004) Nonextensive Entropy—Interdisciplinary Applications. Oxford University Press, New York.
- [10] Menard, M., Courboulay, V. and Dardignac, P. (2003) Possibilistic and Probabilistic Fuzzy Clustering: Unification within the Framework of the Non-Extensive Thermostatistics. *Pattern Recognition*, **36**, 1325-1342.
- [11] Menard, M., Dardignac, P. and Chibelushi, C.C. (2004) Non-Extensive Thermostatistics and Extreme Physical Information for Fuzzy Clustering. *International Journal* of Computational Cognition, 2, 1-63.
- [12] Yasuda, M. (2010) Deterministic Annealing Approach to Fuzzy C-Means Clustering Based on Entropy Maximization. *Advances in Fuzzy Systems*, 2011, Article ID: 960635.
- [13] Yasuda, M. and Orito, Y. (2014) Multi-Q Extension of Tsallis Entropy Based Fuzzy C-Means Clustering. *Journal of Advanced Computational Intelligence and Intelli*gent Informatics, 18, 289-296. https://doi.org/10.20965/jaciii.2014.p0289
- [14] Yasuda, M. (2016) Approximate Determination of Q-Parameter for FCM with Tsallis Entropy Maximization. Proceedings of the Joint 8th International Conference on Soft Computing and 17th International Symposium on Advanced Intelligent Systems, 700-705. https://doi.org/10.1109/scis-isis.2016.0151
- [15] MacQueen, J. (1967) Some Methods for Classification and Analysis of Multivariate Observations. *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, 281-297.
- [16] UCI Machine Learning Repository (1998) Iris Data Set. <u>http://archive.ics.uci.edu/ml/datasets/Iris/</u>
- [17] McCallum, A., Nigam, K. and Ungar, L.H. (2000) Efficient Clustering of High Dimensional Data Sets with Application to Reference Matching. *Proceedings of the* 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 169-178.
- [18] De, T., Burnet, D.F. and Chattopadhyay, A.K. (2016) Clustering Large Number of Extragalactic Spectra of Galaxies and Quasars through Canopies. *Communication* in Statistics—Theory and Methods, 45, 2638-2653. https://doi.org/10.1080/03610926.2013.848286

## Appendix

**Table 7.** Maximum, minimum, mean, and standard deviation of q (c = 3,  $\beta = 5.351e - 06$ ).

|                        | Ν      | Maximum | Minimum | Mean  | Std. deviation |
|------------------------|--------|---------|---------|-------|----------------|
|                        | 5      | 1.825   | 1.165   | 1.437 | 0.225          |
| r <sub>max</sub>       | 10     | 2.089   | 1.290   | 1.688 | 0.249          |
|                        | 100    | 2.823   | 1.029   | 1.667 | 0.320          |
|                        | 1000   | 5.699   | 1.006   | 1.631 | 0.374          |
|                        | 10,000 | 9.884   | 1.001   | 1.632 | 0.412          |
|                        | 5      | 2.256   | 1.820   | 2.043 | 0.160          |
|                        | 10     | 2.279   | 1.959   | 2.038 | 0.106          |
| ₿<br>mean              | 100    | 2.612   | 1.501   | 2.072 | 0.235          |
|                        | 1000   | 2.749   | 1.255   | 2.050 | 0.247          |
|                        | 10,000 | 2.889   | 1.198   | 2.060 | 0.244          |
|                        | 5      | 2.415   | 2.394   | 2.405 | 0.006          |
|                        | 10     | 2.415   | 2.372   | 2.397 | 0.014          |
| $r_{\max}^k$           | 100    | 2.421   | 2.374   | 2.404 | 0.010          |
|                        | 1000   | 2.427   | 1.888   | 2.401 | 0.031          |
|                        | 10,000 | 2.430   | 1.827   | 2.400 | 0.040          |
|                        | 5      | 2.461   | 2.410   | 2.449 | 0.019          |
|                        | 10     | 2.459   | 2.404   | 2.438 | 0.020          |
| $r_{\rm mean}^{\rm k}$ | 100    | 2.461   | 2.404   | 2.440 | 0.021          |
|                        | 1000   | 2.476   | 1.909   | 2.435 | 0.046          |
|                        | 10,000 | 2.474   | 1.913   | 2.436 | 0.046          |

Table 8. Maximum, minimum, mean, and standard deviation of q ( c=5 ,  $\beta = 3.608e - 06$  ).

|                  | Ν      | Maximum | Minimum | Mean  | Std. deviation |
|------------------|--------|---------|---------|-------|----------------|
|                  | 5      | 2.266   | 1.042   | 1.616 | 0.498          |
| r <sub>max</sub> | 10     | 2.106   | 1.122   | 1.650 | 0.330          |
|                  | 100    | 2.596   | 1.018   | 1.784 | 0.281          |
|                  | 1000   | 2.400   | 1.020   | 1.781 | 0.238          |
|                  | 10,000 | 8.514   | 1.002   | 1.789 | 0.259          |
|                  | 5      | 2.884   | 2.308   | 2.562 | 0.184          |
|                  | 10     | 2.816   | 1.218   | 2.514 | 0.181          |
| $r_{\rm mean}$   | 100    | 3.125   | 1.890   | 2.489 | 0.227          |
|                  | 1000   | 3.381   | 1.855   | 2.490 | 0.213          |
|                  | 10,000 | 4.355   | 1.730   | 2.482 | 0.220          |
|                  | 5      | 3.452   | 2.245   | 2.790 | 0.543          |
|                  | 10     | 3.452   | 1.860   | 2.786 | 0.567          |
| $r_{\max}^k$     | 100    | 3.453   | 1.841   | 3.005 | 0.525          |
|                  | 1000   | 3.454   | 1.765   | 2.946 | 0.535          |
|                  | 10,000 | 3.454   | 1.760   | 2.979 | 0.529          |
|                  | 5      | 3.456   | 2.966   | 3.240 | 0.223          |
|                  | 10     | 3.457   | 3.205   | 3.421 | 0.077          |
| $r_{mean}^{k}$   | 100    | 3.488   | 2.962   | 3.361 | 0.151          |
|                  | 1000   | 3.492   | 2.630   | 3.341 | 0.183          |
|                  | 10,000 | 3.496   | 2.455   | 3.344 | 0.181          |



|                        | Ν      | Maximum | Minimum | Mean  | Std. deviation |
|------------------------|--------|---------|---------|-------|----------------|
|                        | 5      | 1.606   | 1.314   | 1.515 | 0.110          |
|                        | 10     | 1.829   | 1.057   | 1.412 | 0.268          |
| r <sub>max</sub>       | 100    | 1.952   | 1.011   | 1.382 | 0.218          |
|                        | 1000   | 1.993   | 1.011   | 1.408 | 0.209          |
|                        | 10,000 | 2.000   | 1.010   | 1.410 | 0.205          |
|                        | 5      | 1.920   | 1.649   | 1.821 | 0.093          |
|                        | 10     | 1.953   | 1.628   | 1.767 | 0.099          |
| $r_{\rm mean}$         | 100    | 1.984   | 1.270   | 1.730 | 0.160          |
|                        | 1000   | 1.998   | 1.201   | 1.744 | 0.164          |
|                        | 10,000 | 2.000   | 1.170   | 1.746 | 0.163          |
|                        | 5      | 1.069   | 1.013   | 1.044 | 0.022          |
|                        | 10     | 1.105   | 1.013   | 1.050 | 0.030          |
| $r_{ m max}^{ m k}$    | 100    | 1.962   | 1.013   | 1.062 | 0.095          |
|                        | 1000   | 1.962   | 1.013   | 1.057 | 0.056          |
|                        | 10,000 | 1.962   | 1.013   | 1.056 | 0.055          |
|                        | 5      | 1.876   | 1.876   | 1.876 | 0.000          |
|                        | 10     | 1.876   | 1.876   | 1.876 | 0.000          |
| $r_{\rm mean}^{\rm k}$ | 100    | 1.993   | 1.876   | 1.877 | 0.012          |
|                        | 1000   | 1.993   | 1.876   | 1.878 | 0.012          |
|                        | 10,000 | 1.993   | 1.876   | 1.877 | 0.012          |

**Table 9.** Maximum, minimum, mean, and standard deviation of q for the Iris Data Set ( $\beta = 1.076e - 01$ ).

**Table 10.** Maximum, minimum, and mean numbers of misclassified data points for the Iris Data Set of the previous method, the proposed method and Tsallis-DAFCM ( $\beta = 1.076e - 01$ ,  $T_{high} = 9.294$ ).

| Method                      | Ν      | q     | Maximum | Minimum | Mean  |
|-----------------------------|--------|-------|---------|---------|-------|
|                             | 5      | 1.515 | 14      | 14      | 14.00 |
|                             | 10     | 1.412 | 16      | 15      | 16.00 |
| Previous method $(r_{max})$ | 100    | 1.382 | 16      | 16      | 16.00 |
|                             | 1000   | 1.408 | 16      | 14      | 16.00 |
|                             | 10,000 | 1.410 | 17      | 16      | 16.00 |
|                             | 5      | 1.821 | 13      | 13      | 13.00 |
|                             | 10     | 1.767 | 13      | 13      | 13.00 |
| Previous method $(r_{men})$ | 100    | 1.730 | 14      | 13      | 13.05 |
| i iraii -                   | 1000   | 1.744 | 13      | 13      | 13.00 |
|                             | 10,000 | 1.746 | 13      | 13      | 13.00 |

| Continued                     |        |       |    |    |       |
|-------------------------------|--------|-------|----|----|-------|
|                               | 5      | 1.044 | 17 | 17 | 17.00 |
|                               | 10     | 1.050 | 17 | 17 | 17.00 |
| Proposed method $(r^k)$       | 100    | 1.062 | 17 | 17 | 17.00 |
| x max /                       | 1000   | 1.057 | 17 | 17 | 17.00 |
|                               | 10,000 | 1.056 | 17 | 17 | 17.00 |
|                               | 5      | 1.876 | 16 | 13 | 13.00 |
| <b>D</b>                      | 10     | 1.876 | 13 | 13 | 13.00 |
| Proposed method $(r_{max}^k)$ | 100    | 1.877 | 13 | 13 | 13.00 |
| - mean -                      | 1000   | 1.877 | 13 | 13 | 13.00 |
|                               | 10,000 | 1.877 | 13 | 13 | 13.00 |
|                               |        | 1.2   | 16 | 16 | 16.00 |
|                               |        | 1.6   | 16 | 14 | 15.91 |
| Tsallis-DAFC                  | М      | 2.0   | 14 | 14 | 14.00 |
|                               |        | 2.4   | 27 | 13 | 13.01 |
|                               |        | 2.8   | 26 | 13 | 13.01 |

# Table 11. Computational times of $\beta$ , q, and clustering for the Iris data set.

|                            | Ν      | β     | q      | Clustering |
|----------------------------|--------|-------|--------|------------|
|                            | 5      | 0.000 | 0.000  | 0.049      |
|                            | 10     | 0.000 | 0.000  | 0.059      |
| r <sub>max</sub>           | 100    | 0.000 | 0.000  | 0.060      |
|                            | 1000   | 0.016 | 0.023  | 0.059      |
|                            | 10,000 | 0.156 | 0.242  | 0.059      |
|                            | 5      | 0.000 | 0.000  | 0.047      |
|                            | 10     | 0.000 | 0.000  | 0.047      |
| $r_{ m mean}$              | 100    | 0.000 | 0.000  | 0.047      |
|                            | 1000   | 0.016 | 0.023  | 0.047      |
|                            | 10,000 | 0.156 | 0.219  | 0.047      |
|                            | 5      | 0.000 | 0.000  | 0.068      |
|                            | 10     | 0.000 | 0.023  | 0.067      |
| $r_{ m max}^{ m k}$        | 100    | 0.000 | 0.188  | 0.068      |
|                            | 1000   | 0.016 | 1.695  | 0.068      |
|                            | 10,000 | 0.156 | 17.586 | 0.068      |
|                            | 5      | 0.000 | 0.000  | 0.047      |
|                            | 10     | 0.000 | 0.000  | 0.047      |
| $r_{_{ m mean}}^{^{ m k}}$ | 100    | 0.000 | 0.000  | 0.047      |
|                            | 1000   | 0.016 | 0.039  | 0.047      |
|                            | 10,000 | 0.156 | 0.430  | 0.047      |

💸 Scientific Research Publishing 🕂

# Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc. A wide selection of journals (inclusive of 9 subjects, more than 200 journals) Providing 24-hour high-quality service User-friendly online submission system Fair and swift peer-review system Efficient typesetting and proofreading procedure Display of the result of downloads and visits, as well as the number of cited articles Maximum dissemination of your research work

Submit your manuscript at: <u>http://papersubmission.scirp.org/</u> Or contact <u>jsea@scirp.org</u>