PAPER
Fuzzy Entropy Based Fuzzy c-Means Clustering with Deterministic and Simulated Annealing Methods

Makoto YASUDA[†a], *Nonmember and* **Takeshi FURUHASHI**[††b], *Member*

SUMMARY This article explains how to apply the deterministic annealing (DA) and simulated annealing (SA) methods to fuzzy entropy based fuzzy c-means clustering. By regularizing the fuzzy c-means method with fuzzy entropy, a membership function similar to the Fermi-Dirac distribution function, well known in statistical mechanics, is obtained, and, while optimizing its parameters by SA, the minimum of the Helmholtz free energy for fuzzy c-means clustering is searched by DA. Numerical experiments are performed and the obtained results indicate that this combinatorial algorithm of SA and DA can represent various cluster shapes and divide data more properly and stably than the standard single DA algorithm.
key words: fuzzy c-means clustering, fuzzy entropy, Fermi-Dirac distribution, deterministic annealing, simulated annealing

1. Introduction

Statistical mechanics investigates the macroscopic properties of a physical system consisting of several elements. Recently, there has been great research interest in applying statistical mechanical models or tools to information science. Many engineering problems can be formulated as optimization problems, and the simulated annealing (SA) method [1], [2] is an efficient optimization technique for such problems. SA is a stochastic relaxation method which, by analogy with the annealing process of solids, treats a cost function as the energy of a system and, as the (pseudo) temperature is decreased, searches for a minimum randomly at a high temperature and more deterministically at a low temperature. SA is a global optimization technique for solving combinatorial optimization problems as long as a cost function is definable and the cooling is performed sufficiently slowly. SA, however, requires a very long time to find an optimal solution because it searches stochastically at each temperature and is practically an approximation algorithm.

The deterministic annealing (DA) method, on the other hand, is a deterministic variant of SA [3], [4]. DA characterizes the minimization problem of the cost function as the minimization of the Helmholtz free energy, which depends on the temperature, and tracks its minimum while decreasing the temperature and thus it can deterministically optimize the cost function at each temperature. Hence, DA is

more efficient than SA but does not guarantee the optimal solution. In addition, an effect of a cooling schedule on a quality of DA's solution is still not known. The DA method is first applied to the Shannon entropy based fuzzy clustering and proved to be effective [3].

As for the fuzzy theory, there exists a strong relationship between the membership functions of the fuzzy c-means (FCM) clustering [5] with the maximum entropy or entropy regularization methods [6], [7] and the statistical mechanical distribution functions. That is, FCM regularized with the Shannon entropy gives a membership function similar to the Maxwell-Boltzmann (or Gibbs) distribution function [3], [4], [6], and FCM regularized with the fuzzy entropy [8] gives a membership function similar to the Fermi-Dirac distribution function[*]. These membership functions are suitable for the annealing methods because they contain a parameter corresponding to the system temperature. In the present article, we focus on the Fermi-Dirac-like membership function and propose a new fuzzy c-means clustering algorithm using the DA and SA methods. As a complexity of a system increases or in the real world, it is supposed that, cluster boundaries are not so clear that fuzzy clustering is more suitable than crisp clustering.

The advantage of using the Maxwell-Boltzmann-like or Fermi-Dirac-like membership functions is that the fuzzy c-means clustering can be interpreted and analyzed from a statistical mechanical point of view [9]–[11]. Furthermore, as compared to the Maxwell-Boltzmann-like membership function, the Fermi-Dirac-like membership function has extra parameters α_ks[**], which, make it possible to represent various cluster shapes according to their unsmoothnesses, similar to former clustering methods based on the Gaussian mixture [13], the degree of fuzzy entropy [14], and the confusion degree [15], for example. α_ks strongly affect clustering results and must be optimized under a normalization constraint of FCM. On the other hand, although it is efficient, the DA method does not give appropriate values of α_ks by itself, and the DA clustering sometimes fails if α_ks are assigned improperly.

Accordingly, the objective of this article is to overcome these problems and modify the fuzzy entropy based DA method. We introduce SA into the DA clustering to optimize α_ks because, as pointed out above, both DA and SA

---

[*]These are called the Maxwell-Boltzmann-like and the Fermi-Dirac-like membership functions respectively in this article.
[**]α_k corresponds to a chemical potential in statistical mechanics [12], and k denotes a data point.

contain the parameter corresponding to the system temperature and can be naturally combined as DASA. This approach tries to combine an optimality of SA with an effectiveness of DA.

Nevertheless, this approach causes a few problems.

1. How should the initial values of α_ks be estimated under the normalization constraint?
2. How should the initial annealing temperature be estimated?
3. SA must optimize continuous α_k [16], [17].
4. SA must optimize many α_ks [18].

Linear approximations of the Fermi-Dirac-like membership function is useful in guessing the initial α_ks and the initial annealing temperature of DA.

In order to perform SA in a continuous variable domain, since an appropriate value of α_k is unknown, a normal distribution is assumed to generate a new solution in the vicinity of the current α_k.

In order to perform SA in a many-variables domain, α_ks to be optimized should be selected according to a selection rule to reduce a computational time. In an early annealing stage, most α_ks are optimized. In a final annealing stage, however, only α_ks of data that are located sufficiently far from all cluster centers are optimized because their memberships might be fuzzy. Distances between the data and the cluster centers are measured by using linear approximations of the Fermi-Dirac-like membership function.

Numerical experiments show that DASA clusters data that are distributed in various shapes more properly and stably than the standard single fuzzy entropy based DA algorithm.

2. Fuzzy Entropy Regularization of FCM

First, we introduce the fuzzy entropy into the FCM clustering.

Let $X = \{x_1, \ldots, x_n\}$ ($x_k = (x_k^1, \ldots, x_k^p) \in R^p$) be a data set in the p-dimensional real space, which should be divided into c clusters. In addition, let $V = \{v_1, \ldots, v_c\}$ ($v_i = (v_i^1, \ldots, v_i^p)$) be the centers of clusters, and let $u_{ik} \in [0, 1]$ ($i = 1, \ldots, c$; $k = 1, \ldots, n$) be the membership functions. Furthermore, let

$$J = \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ik})^m d_{ik} \quad (m > 1) \tag{1}$$

be the objective function of FCM, where $d_{ik} = \|x_k - v_i\|^2$. In FCM, under the normalization constraint of

$$\sum_{i=1}^{c} u_{ik} = 1 \quad (\forall k), \tag{2}$$

the Lagrange function L_{FCM} is given by

$$L_{FCM} = J - \sum_{k=1}^{n} \eta_k \left(\sum_{i=1}^{c} u_{ik} - 1 \right), \tag{3}$$

where η_k is the Lagrange multiplier. In case of the FCM regularized by the fuzzy entropy S_{FE}, m is set to 1 and S_{FE} is introduced to L_{FCM} as

$$L = L_{FCM} - \lambda S_{FE}, \tag{4}$$

where λ can be considered to play a role of the Lagrange multiplier [19]. That is, the regularization parameter is replaced by λ [10]. This method approaches crisp or hard clustering as λ decreases to +0. The fuzzy entropy can be written as

$$S_{FE} = -\sum_{k=1}^{n} \sum_{i=1}^{c} \{u_{ik} \log u_{ik} + (1 - u_{ik}) \log(1 - u_{ik})\}. \tag{5}$$

Here, $\partial L/\partial u_{ik} = 0$ yields the following membership function

$$u_{ik} = \frac{1}{e^{\alpha_k + \beta d_{ik}} + 1}, \tag{6}$$

where $\beta = 1/\lambda$ defines the extent of the distribution [10], and $\alpha_k = -\eta_k/\lambda$. Equation (6) is formally normalized as

$$u_{ik} = \frac{1}{e^{\alpha_k + \beta d_{ik}} + 1} \bigg/ \sum_{j=1}^{c} \frac{1}{e^{\alpha_k + \beta d_{jk}} + 1}. \tag{7}$$

$\partial L/\partial v_i = 0$ gives

$$v_i = \frac{\sum_{k=1}^{n} u_{ik} x_k}{\sum_{k=1}^{n} u_{ik}}. \tag{8}$$

3. Deterministic Annealing

The membership function obtained in Eq. (6) resembles the Fermi-Dirac distribution function

$$n_l = \frac{1}{e^{\alpha + \beta \epsilon_l} + 1}, \tag{9}$$

where n_l is an average number of particles that occupy energy levels ϵ_l. This similarity makes it possible to formulate the preceding method as the DA clustering from the statistical mechanical viewpoint.

3.1 Formulation of Fuzzy Clustering with DA

The grand partition function for particles governed by the Fermi-Dirac distribution function or the Fermi-Dirac statistics takes the form

$$\Xi = \prod_l (1 + e^{-\alpha - \beta \epsilon_l}), \tag{10}$$

where $\beta = 1/(k_B T)$ (k_B is the Boltzmann constant) [12]. The relationship $F = -(1/\beta)(\log \Xi - \alpha \partial \log \Xi/\partial \alpha)$ gives the Helmholtz free energy

$$F = -\frac{1}{\beta} \left\{ \sum_l \log(1 + e^{-\alpha - \beta \epsilon_l}) + \alpha N \right\}, \tag{11}$$

where N is the total number of particles [12].

Similarly, in the FCM clustering regularized with the fuzzy entropy, in which data can belong to any cluster, the grand partition function can be written as

$$\Xi_{FE} = \prod_{k=1}^{n} \prod_{i=1}^{c} (1 + e^{-\alpha_k - \beta d_{ik}}), \qquad (12)$$

which, from the relationship $F_{FE} = -(1/\beta)(\log \Xi_{FE} - \alpha_k \partial \log \Xi_{FE}/\partial \alpha_k)$, gives the Helmholtz free energy

$$F_{FE} = -\frac{1}{\beta} \sum_{k=1}^{n} \left\{ \sum_{i=1}^{c} \log(1 + e^{-\alpha_k - \beta d_{ik}}) + \alpha_k \right\}. \qquad (13)$$

The inverse of β can be considered to represents the system or computational temperature T.

According to the principle of minimal free energy in statistical mechanics, the minimum Helmholtz free energy determines the distribution at thermal equilibrium [12]. Thus, formulating the DA clustering as a minimization of Eq. (13) leads to $\partial F_{FE}/\partial \mathbf{v}_i = 0$ at the current temperature, and again gives Eq. (8). Desirable cluster centers are obtained by calculating Eqs. (6) and (8) repeatedly.

3.2 Correspondences between Fermi-Dirac Statistics and Fuzzy Entropy Based Fuzzy Clustering

Correspondences between the FCM clustering regularized with the fuzzy entropy (FC) and the Fermi-Dirac statistics (FD) are summarized as follows:

- **Constraints:** (a) Constraint whereby the sum of all particles occupying the energy level ϵ_l is fixed, that is, $\sum_l n_l = N$ in FD corresponds with the normalization constraint in FC, and l corresponds to the cluster number i. In addition, the fact that data can belong to multiple clusters leads to the summation on k. (b) There is no constraint in FC that corresponds to $\sum_l \epsilon_l n_l = E$ in FD. Nevertheless, $\sum_{k=1}^{n} \sum_{i=1}^{c} u_{ik} d_{ik}$ must instead be minimized in FC.
- **Distribution Function:** n_l/N, the ratio of particles that occupy the energy level ϵ_l, corresponds to u_{ik}. In FD, particles cannot be distinguished from each other. In FC, however, data are distinguishable, and for that reason, u_{ik} gives the probability of data belonging to multiple clusters.
- **Entropy:** The entropy S for FD equals $-\partial F/\partial T$. Similarly, the fuzzy entropy S_{FE} equals $-\partial F_{FE}/\partial T$ in FC.
- **Energy:** The relationships $E = F + TS$ and $J_{m=1} = F_{FE} + TS_{FE}$ hold in FD and FC, respectively.

In the DA clustering, cluster distributions that minimize F_{FE} are searched at a given temperature. At a higher temperature, membership functions are widely distributed and clusters to which a datum belongs are fuzzy. By reducing the temperature according to the annealing schedule, DA achieves thermal equilibrium, which minimizes F_{FE}. At absolute zero, fuzzy clustering becomes crisp clustering. The extent of distribution, or fuzziness, can be measured by the fuzzy entropy S_{FE}.

3.3 Linear Approximation of Fermi-Dirac Distribution Function

The Fermi-Dirac distribution function can be approximated by linear functions. That is, as shown in Fig. 1, the Fermi-Dirac distribution function of the form:

$$f(x) = \frac{1}{e^{\alpha + \beta x^2} + 1} \qquad (14)$$

is approximated by the linear functions

$$g(x) = \begin{cases} 1.0 & \left(x \le \dfrac{-\alpha - 1}{\kappa} \right) \\[2mm] -\dfrac{\kappa}{2}x - \dfrac{\alpha}{2} + \dfrac{1}{2} & \left(\dfrac{-\alpha - 1}{\kappa} \le x \le \dfrac{-\alpha + 1}{\kappa} \right), \\[2mm] 0.0 & \left(\dfrac{-\alpha + 1}{\kappa} \le x \right) \end{cases} \qquad (15)$$

where $\kappa = \sqrt{-\alpha\beta}$. $g(x)$ satisfies $g(\sqrt{-\alpha/\beta}) = 0.5$ and requires that α be negative. In the same way, the Fermi-Dirac-like membership function can be approximated by linear functions.

In Fig. 2, $\Delta x = x - x_{new}$ denotes a reduction in the extent of distribution with decreasing the temperature from T to T_{new} ($T > T_{new}$). The extent of distribution also narrows with increasing α. α_{new} ($\alpha < \alpha_{new}$), which satisfies $g(0.5)_\alpha - g(0.5)_{\alpha_{new}} = \Delta x$, is obtained as

$$\alpha_{new} = -\left\{ \sqrt{-\alpha} + \sqrt{-\alpha\beta_{new}} \left(\frac{1}{\sqrt{\beta}} - \frac{1}{\sqrt{\beta_{new}}} \right) \right\}^2, \quad (16)$$

where $\beta = 1/T$ and $\beta_{new} = 1/T_{new}$. Thus, taking T to be the temperature at which the previous DA was executed and T_{new} to be the next temperature, the covariance of α_k's distribution is estimated as

$$\Delta\alpha = \alpha_{new} - \alpha, \qquad (17)$$
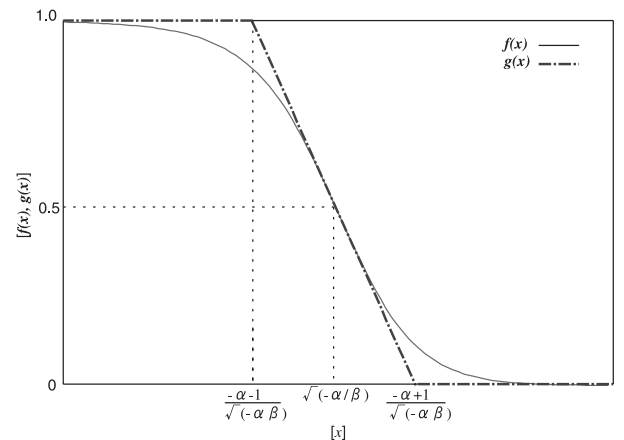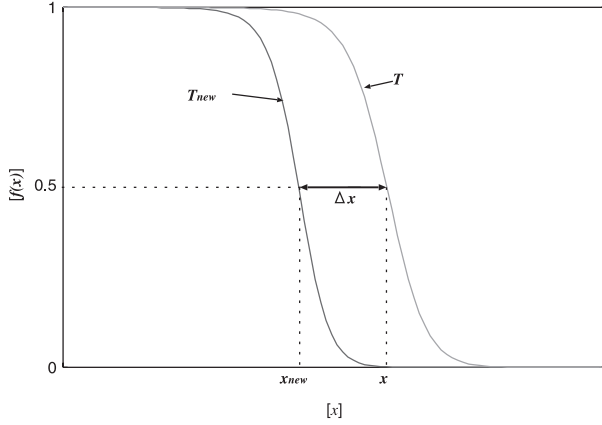
by assuming its normal distribution.

Fig. 1 The Fermi-Dirac distribution function $f(x)$ and its linear approximation functions $g(x)$.

Fig. 2 Decreasing of extent of the Fermi-Dirac distribution function from x to x_{new} with decreasing the temperature from T to T_{new}.

3.4 Initial Estimation of α_k and Annealing Temperature

Before executing DA, it is very important to estimate the initial values of α_ks and the initial annealing temperature in advance.

From Fig. 1, the distances between a data point and cluster centers are averaged as

$$\mathcal{L}_k = \frac{1}{c} \sum_{i=1}^{c} \|\boldsymbol{x}_k - \boldsymbol{v}_i\|, \tag{18}$$

and this gives

$$\alpha_k = -\beta(\mathcal{L}_k)^2. \tag{19}$$

With given initial clusters distributed widely enough, Eq. (19) overestimates α_k, so that α_k must be adjusted by decreasing its value gradually.

Furthermore, Fig. 1 gives the width of the Fermi-Dirac distribution function as wide as $2(-\alpha + 1)/(\sqrt{-\alpha\beta})$, which must be equal to or smaller than that of the data distribution width ($= 2R$). This condition leads to

$$2\frac{-\alpha + 1}{\sqrt{-\alpha\beta}} = 2R. \tag{20}$$

As a result, the initial value of β_{low} or the initial annealing temperature T_{high} is roughly determined as

$$\beta_{low} \simeq \frac{4}{R^2} \quad \left(T_{high} \simeq \frac{R^2}{4}\right). \tag{21}$$

4. Simulated Annealing

The cost or objective function for SA is

$$E(\alpha_k) = J_{m=1} + K \sum_{k=1}^{n} \left(\sum_{i=1}^{c} u_{ik} - 1\right)^2, \tag{22}$$

where K is a constant large enough to satisfy the normalization constraint.

In order to optimize each α_k by SA, its neighbor α_k^{new} (a displacement from the current α_k) is generated as a normal distribution with a mean of 0 and a covariance of $\Delta\alpha_k$, as defined in Eq. (17).

The SA's initial temperature $T_0(= 1/\beta_0)$ is determined such that the acceptance probability becomes

$$\mathrm{e}^{-\beta_0\{E(\alpha_k)-E(\alpha_k^{new})\}} = 0.5 \quad (E(\alpha_k) - E(\alpha_k^{new}) \geq 0). \tag{23}$$

By selecting α_ks to be optimized from data that locate far from all cluster centers, the computational time of SA can be shortened. After annealing has done to some extent, the boundary of the transition region can be easily obtained with the linear approximations of the membership function. Thus, from Fig. 1, data that have distances greater than $\sqrt{-\alpha_k/\beta}$ from each cluster center are selected at the final SA stage.

4.1 Simulated Annealing Algorithm

The SA algorithm is stated as follows:
1 **Initialization:** Calculate an initial temperature T_0 from Eq. (23). Set a current temperature T to T_0. Set an iteration count t to 1. Calculate a covariance $\Delta\alpha_k$ for each α_k by Eq. (17).
2 Select α_ks to be optimized, if necessary.
3 Calculate neighbors of current α_ks.
4 Apply the Metropolis algorithm [20] to the selected α_ks using Eq. (22) as the cost function and calculating u_{ik} without the normalization constraint by Eq. (6).
5 If $t_{SAmax} < t$ is satisfied, then return. Otherwise, decrease the temperature as $T = T_0/\log(t+1)$, increment t, and go back to **2**.

5. Combinatorial Algorithm of Deterministic and Simulated Annealing

The SA algorithms is combined with the DA algorithm as follows:
1 **Initialization for DASA:** Set a threshold of the convergence test δ_1. Set maximum iteration counts max_1 and max_2, and an iteration count $t_{SA} = 1$ for the SA algorithm. In addition, set a maximum iteration count max_3 for DASA.
2 **Initialization for DA:** Set a rate at which the temperature is lowered T_{rate}, and set a threshold for the convergence test δ_2. Calculate the initial temperature T_{high} by Eq. (21) and set the current temperature $T = T_{high}$. Place c clusters randomly, and estimate initial α_ks by Eq. (19). After that, adjust the estimated initial α_ks to satisfy the normalization constraint Eq. (2). Also set an iteration count of the DA algorithm t_{DA} to 1.
3 Calculate u_{ik} by Eq. (7) and $\boldsymbol{v}_i$ by Eq. (8).
4 **Convergence test for DA:** Compare the difference between the current cost (objective) value $J_{m=1}$ and that obtained at the previous iteration $\hat{J}_{m=1}$. If $\|J_{m=1} - \hat{J}_{m=1}\|/J_{m=1} < \delta_2 \cdot T/T_{high}$ is satisfied, then goto **5**. Otherwise, increment t_{DA}, and go back to **3**.

5 Execution of SA: Set $t_{SAmax} = max_1$, and execute the SA algorithm, and add a number of times of iterations to t_{SA}.

6 Convergence test for DASA: Compare the difference between the current cost (objective) value $\bar{J}_{m=1}$ and that obtained at the previous iteration $\hat{\bar{J}}_{m=1}$. If $\|\bar{J}_{m=1} - \hat{\bar{J}}_{m=1}\|/\bar{J}_{m=1} < \delta_1$ or $max_3 < t_{DA}$ is satisfied, then goto **7**. Otherwise, decrease the temperature as $T = T * T_{rate}$, and go back to **3**.

7 Execution of final SA: Set $t_{SAmax} = max_2$, execute the SA algorithm, add a number of times of iterations to t_{SA}, and then stop.
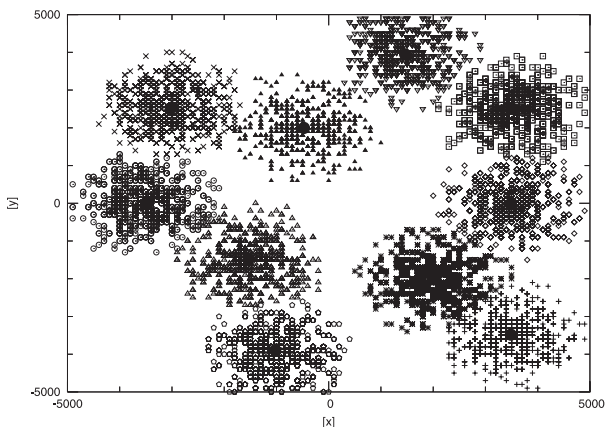
6. Experiments

To demonstrate the effectiveness of the proposed algorithm, numerical experiments were carried out. DASA's results were compared with those of single DA.

We set $\delta_1 = 0.01$, $\delta_2 = 0.5$, $T_{rate} = 0.8$, $max_1 = 500$, $max_2 = 20000$, and $max_3 = 10^\dagger$. We also set R in Eq. (20) to 350.0 for experimental data 1~3, and 250.0 for experimental data 4.

In Experiment 1, 11,479 data points were generated as ten equally sized normal distributions. Figure 3 shows a fuzzy clustering result by DASA. Single DA similarly clusters these data.

In Experiment 2-1, three differently sized normal distributions consisting of 2,249 data points shown in Fig. 4 were used. Figure 4 (0) shows the initial clusters obtained by the initial estimation of α_ks and the annealing temperature. Figures 4 (1)~(6a) show a fuzzy clustering process of DASA. At the high temperature in Fig. 4 (1), as described in Sect. 3.2, the Fermi-Dirac-like membership functions were widely distributed and clusters to which data belong were fuzzy. However, with decreasing temperature (from Fig. 4 (2) to Fig. 4 (5)), the distribution became less and less fuzzy. After executing DA and SA alternately, the clusters in Fig. 4 (6a) were obtained. Then, data to be optimized by SA were selected by the criterion stated in

Fig. 3 Experimental result 1. (Fuzzy clustering result obtained by DASA. Black circles indicate centers of clusters.)

Sect. 4, and SA was executed. The final result of DASA in Fig. 4 (6b) shows that data were desirably clustered (The DASA algorithm without the data selection rule gives a similar result). In contrast, because of the randomness of the initial cluster positions and the difficulty of good estimation of the initial α_ks, single DA becomes unstable, and sometimes gives satisfactory results, as shown in Fig. 4 (6c), and sometimes not, as shown in Fig. 4 (6d). By comparing Figs. 4 (6b) to 4 (6c), it is found that, due to the optimization of α_ks by SA, the resultant cluster shapes of DASA are far less smooth than those of single DA. These rough membership functions are suitable to represent various fuzzy sets [15].

Changes of the costs of DASA ($J_{m=1}$ for the DA stage and Eq. (22) for the SA stage[††], respectively) are plotted as a function of iteration in Fig. 5, and the both costs decrease with increasing iteration. In this experiment, the total iteration of the SA stage t_{SA} was approximately 3,615, while that of the DA stage (t_{DA} in the combinatorial DASA algorithm) was only 8. Accordingly, the amount of simulation time of DASA was mostly consumed in the SA stage.

In Table 1, computational times of final SA with or without the data selection rule is compared[†††]. It is confirmed that the time reduction effect is achieved by this method.

In Experiment 2-2, in order to examine the effectiveness of SA introduced in DASA, Experiment 2 was conducted ten times, as shown in Table 2, where *ratio* listed in the second row is the ratio of data optimized at the SA stage. That is,

$$ratio = \frac{\text{number of } \alpha_k s \text{ selected randomly from } S_{\alpha_k}}{\text{number of } S_{\alpha_k}},$$
(24)

where S_{α_k} is the set of α_ks selected in the SA algorithm. "UP" means to increase *ratio* as $1.0 - 1.0/n$, where n is the number of executions of the SA stage, and "DOWN" means to decrease *ratio* as $1.0/n$. The results are judged as "Success" or "Failure" from a human viewpoint[††††]. From Table 2, it is concluded that DASA always clusters the data properly if *ratio* is large enough ($0.6 < ratio$), whereas, as listed in the last column, single DA succeeds by 50%.

In Experiments 3 and 4, two elliptic distributions consisting of 2,024 data points, and two horseshoe-shaped distributions consisting of 1,380 data points were used, respectively. Figures 6 and 7 show DASA's clustering results. It is found that DASA can cluster these data properly. In Experiment 3, the percentage of success of DASA is 90%, whereas that of single DA is 50%. In Experiment 4, the percentage of success of DASA is 80%, whereas that of single DA is 40%.

[†]These parameters have not been optimized for experimental data.

[††]K was set to 1×10^{15} in Eq. (22).

[†††]Executions were conducted on an AMD Athlon(tm) 64×2 Dual Core Processor.

[††††]No close case was observed in this experiment.
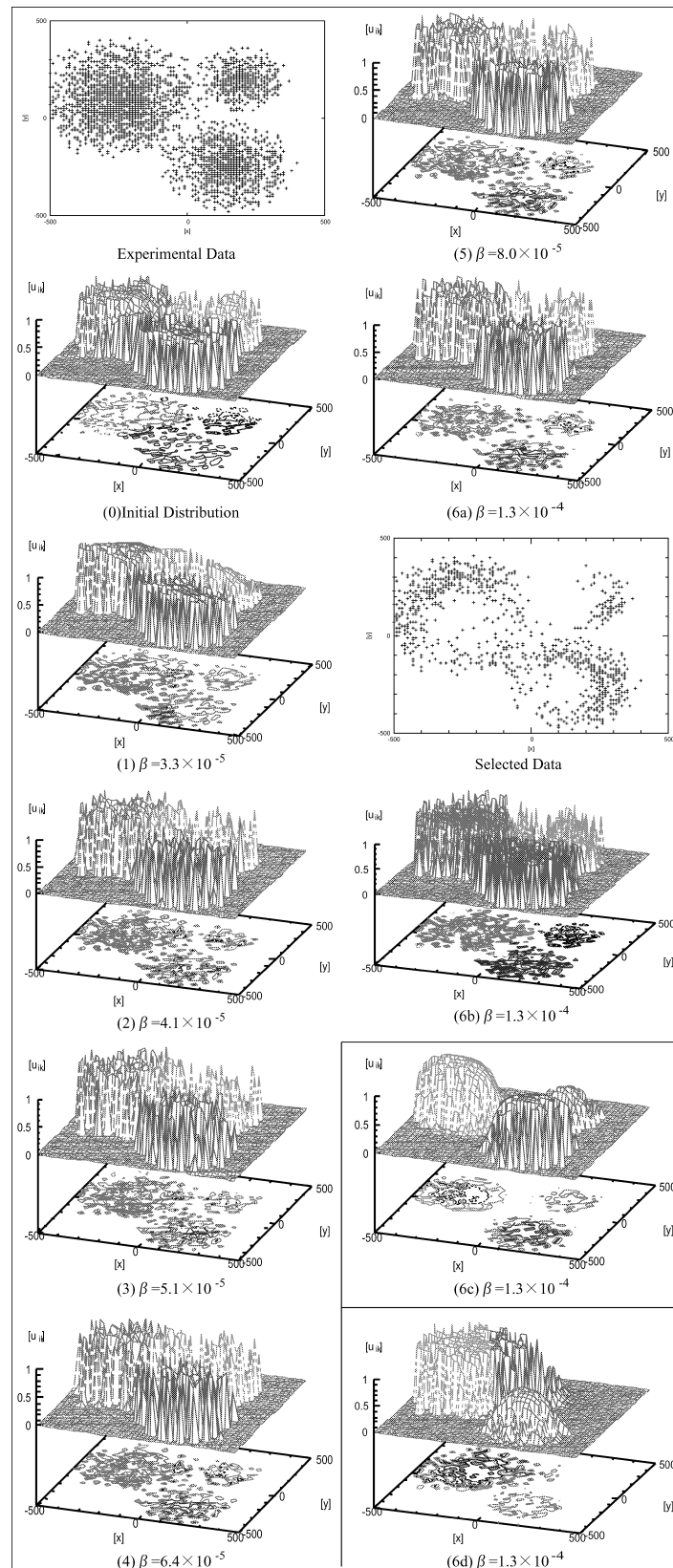
Fig. 4 Experimental result 2-1. (Fuzzy clustering result obtained by DASA and single DA. "Experimental Data" are given data distributions. "Selected Data" are data selected for final SA by the selection rule. (1)~(6a) and (6b) are results using DASA. (6c) and (6d) are results using single DA (success and failure, respectively). Data plotted on the xy plane show the cross sections of u_{ik} at 0.2 and 0.8.)
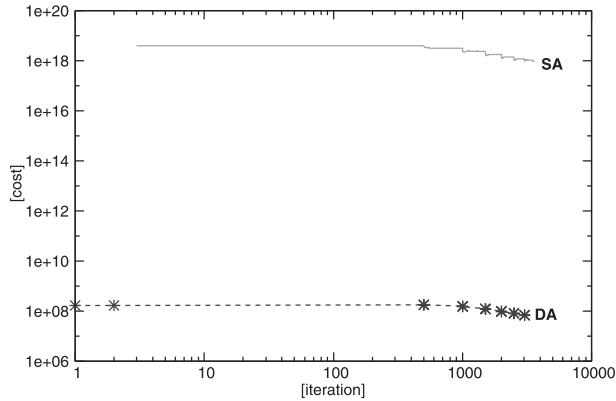
Fig. 5 Experimental result 2-1. (Change in the cost of DASA as a function of iteration counts of DA and SA. $J_{m=1}$ for the DA stage and $J_{m=1} + K \sum_{k=1}^{n} \left(\sum_{i=1}^{c} u_{ik} - 1 \right)^2$ for the SA stage against $t_{DA} + t_{SA}$.)

Table 1 Experimental result 2-2. (Comparison of computational times [sec] of final SA with/without data selection) (average of 10 trials).

|	
without data selection	0.856
with data selection	0.781

Table 2 Experimental result 2-2. (Comparison of numbers of successes and failures of fuzzy clustering using DASA for *ratio* = 0.3, 0.6, 1.0, 1.0 − 1.0/n (UP), 1.0/n (DOWN) and single DA. (n is the number of execution times of the SA stage).)

	DASA					DA
ratio	0.3	0.6	1.0	UP	DOWN	
Success	6	9	10	6	7	5
Failure	4	1	0	4	3	5
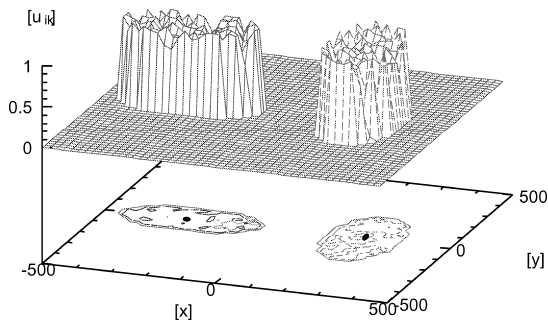
Fig. 6 Experimental result 3. (Fuzzy clustering result of elliptic distributions obtained using DASA. Data plotted on the xy plane show the cross sections of u_{ik} at 0.2 and 0.8. Black circles indicate centers of clusters.)

These experimental results demonstrate the advantage of DASA over single DA. Nevertheless, DASA suffers from two disadvantages. First, it takes so long to execute SA repeatedly that, instead of Eq. (6) or (7), it might be better to use its linear approximation functions as the membership function. Second, since α_ks differ from each other, it is difficult to interpolate them.
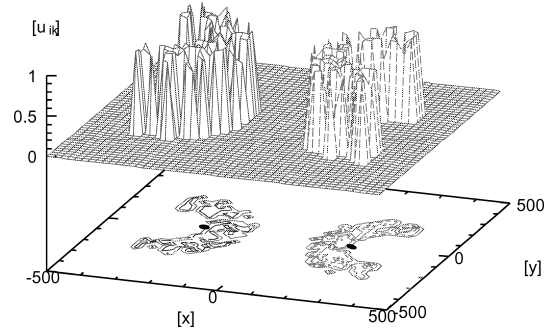
Fig. 7 Experimental result 4. (Fuzzy clustering result of horseshoe-shaped distributions obtained using DASA. Data plotted on the xy plane show the cross sections of u_{ik} at 0.2 and 0.8. Black circles indicate centers of clusters.)

7. Conclusions

In the present article, by combining the deterministic and simulated annealing methods, we proposed a new statistical mechanical fuzzy c-means clustering algorithm (DASA). Numerical experiments showed the effectiveness and the stability of DASA.

However, as stated at the end of Sect. 6, DASA has problems to be considered. In addition, a major problem of the fuzzy c-means methodologies is that they do not give a number of clusters by themselves. Thus, a method such as [11], which can determine the number of clusters automatically, should be combined with DASA.

Future research will include experiments and examinations of the properties of DASA, especially on the adjustment of its parameters, its annealing scheduling problem, and its applications.

References

[1] E. Aarts and J. Korst, Simulated Annealing and Boltzmann Machines, John Wiley & Sons, Chichester, 1989.

[2] S. Kirkpatrick, C.D. Gelatt, and M.P. Vecchi, "Optimization by simulated annealing," Science, vol.220, pp.671–680, 1983.

[3] K. Rose, E. Gurewitz, and B.C. Fox, "A deterministic annealing approach to clustering," Pattern Recognit. Lett., vol.11, no.9, pp.589–594, 1990.

[4] K. Rose, "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems," Proc. IEEE, vol.86, no.11, pp.2210–2239, 1998.

[5] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Prenum Press, New York, 1981.

[6] R.-P. Li and M. Mukaidono, "A maximum entropy approach to fuzzy clustering," Proc. 4th IEEE Int. Conf. Fuzzy Systems (FUZZ-IEEE/IFES '95), pp.2227–2232, 1995.

[7] S. Miyamoto and M. Mukaidono, "Fuzzy c-means as a regularization and maximum entropy approach," Proc. 7th Int. Fuzzy Systems Association World Congress, vol.II, pp.86–92, 1997.

[8] A. DeLuca and S. Termini, "A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory," Information and Control, vol.20, pp.301–312, 1972.

[9] K. Rose, E. Gurewitz, and B.C. Fox, "Statistical mechanics and phase transition in clustering," Phys. Rev. Lett., vol.65, no.8, pp.945–948, 1990.

[10] M. Yasuda, T. Furuhashi, M. Matsuzaki, and S. Okuma, "Fuzzy clustering using deterministic annealing method and its statistical mechanical characteristics," Proc. 10th IEEE Int. Conf. Fuzzy Systems, P294, 2001.

[11] M. Yasuda, T. Furuhashi, and S. Okuma, "Phase transitions in fuzzy clustering based on fuzzy entropy," Journal of Advanced Computational Intelligence and Intelligent Informatics, vol.7, no.3, pp.370–376, 2003.

[12] L.E. Reichl, A Modern Course in Statistical Physics, John Wiley & Sons, New York, 1998.

[13] A.P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithms," Journal of Royal Stat. Soc., Series B, vol.39, pp.1–38, 1977.

[14] D. Tran and M. Wagner, "Fuzzy entropy clustering," Proc. 9th IEEE Int. Conf. Fuzzy Systems (FUZZ-IEEE2000), vol.1, pp.152–157, 2000.

[15] T. Esaki, T. Hashiyama, and Y. Tsukamoto, "Fuzzy c-means clustering with regularization by confusion degree," Journal of Japan Society for Fuzzy Theory and Intelligent Informatics, vol.18, no.4, pp.609–618, 2006.

[16] A. Corana, M. Marchesi, C. Martini, and S. Ridella, "Minimizing multimodal functions of continuous variables with the simulated annealing algorithm," ACM Trans. Math. Softw., vol.13, no.3, pp.262–280, 1987.

[17] P.R. Wang, "Continuous optimization by a variant of simulated annealing," Computational Optimization and Applications, vol.6, pp.59–71, 1996.

[18] P. Siarry, "Enhanced simulated annealing for globally minimizing functions of many-continuous variables," ACM Trans. Math. Softw., vol.23, no.2, pp.209–228, 1997.

[19] M. Yasuda, T. Furuhashi, and S. Okuma, "Entropy based fuzzy c-means clustering: Analogy with statistical mechanics," Journal of Japan Society for Fuzzy Theory and Intelligent Informatics, vol.17, no.4, pp.468–476, 2005.

[20] N. Metropolis, "Equation of state calculation by fast computing machines," Journal of Chem. Physics, vol.21, pp.1087–1092, 1953.

Takeshi Furuhashi received the Ph.D. degree in engineering from Nagoya University, Japan, in 1985. From 1987 to 1990, he was a research associate of Nagoya University. From 1990 to 2000, he was an Associate Professor of Nagoya University. From 2001 to 2003, he was a Professor of Mie University, Japan. In 2004, he joined Nagoya University, and is now a Professor of the Department of Computational Science and Engineering. His research interests include soft computing and emergent soft computer. He is a member of the IEEE, Japan Society for Fuzzy Theory and Systems, Society of Instrument and Control Engineers.

Makoto Yasuda received the B.S. degree in physics from Kwansei Gakuin University, Japan, in 1980 and his Ph.D. degree in engineering from Nagoya University, Japan, in 2002. In 1980, he joined Oki Electric Industry, Japan. From 1997 to 2001, he was with Nagoya Industrial Research Institute, Japan. Since 2002, he has been with Gifu National College of Technology, Japan, and is now an Associate Professor of the Department of Electrical and Computer Engineering. His current research interests include multi-agent technology, soft computing and statistical mechanics of information.