We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the
TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

BOOK CITATION INDEX
CLARIVATE ANALYTICS
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Deterministic Annealing: A Variant of Simulated Annealing and its Application to Fuzzy Clustering

Makoto Yasuda

Abstract

Deterministic annealing (DA) is a deterministic variant of simulated annealing. In this chapter, after briefly introducing DA, we explain how DA is combined with the fuzzy c-means (FCM) clustering by employing the entropy maximization method, especially the Tsallis entropy maximization. The Tsallis entropy is a q parameter extension of the Shannon entropy. Then, we focus on Tsallis-entropy-maximized FCM (Tsallis-DAFCM), and examine effects of cooling functions for DA on accuracy and convergence. A shape of a membership function of Tsallis-DAFCM depends on both a system temperature and q. Accordingly, a relationship between the temperature and q is quantitatively investigated.

Keywords: deterministic annealing, simulated annealing, free energy, fuzzy c-means clustering, entropy maximization, Shannon entropy, fuzzy entropy, Tsallis entropy

1. Introduction

Statistical mechanics investigates the macroscopic properties of a physical system consisting of many elements. Recently, great research activities of applying statistical mechanical models or tools to information engineering problems have been seen. The entropy maximization method applied to a fuzzy c-means clustering is a good example of such models.

Simulated annealing (SA) [1, 2] is one of the most commonly used optimization techniques and plays an important role in the field of engineering because many of the engineering problems can be formulated as optimization problems. SA is a stochastic relaxation method that treats an objective function as a system energy, and by analogy with the annealing process

of solids, searches for its minimum with decreasing the system temperature. SA searches randomly at a high temperature, but more deterministically at a low temperature. As long as a neighborhood can be defined and the temperature is lowered sufficiently slowly, SA is a global optimization technique for solving optimization problems. It requires a very long time to find an optimal global solution because of a stochastic search at each temperature. Thus, SA is an approximation method in practical.

Deterministic annealing (DA) is a deterministic variant of SA, which is first proposed by Rose et al. [3] for a vector quantization algorithm. DA characterizes the minimization problem of the objective function as the minimization of the free energy of the system, which depends on the temperature. DA tracks the minimum of the free energy while decreasing the temperature. Thus, it can deterministically optimize the objective function at each temperature, and is more efficient than SA but does not guarantee the optimal solution. In addition, an effect of a cooling function on the quality of a DA's solution is still unclear.

Membership functions of the fuzzy c-means (FCM) clustering [4] maximized or regularized with entropy [5, 6] have similar forms with the distribution functions that appear in statistical mechanics. For example, a membership function obtained by the Shannon entropy maximization has a similar form with the Boltzmann-Gibbs distribution function [3, 5]. Similarly, a membership function obtained by the fuzzy entropy [7] has a similar form with the Fermi-Dirac distribution function [8]. Annealing methods can be applicable to these membership functions because they contain a parameter that can be corresponded to the system temperature. The advantage of applying the entropy maximization methods to FCM is that fuzzy clustering can be analyzed from both information processing and statistical physical points of view.

Tsallis [9], by extending Boltzmann-Gibbs statistics nonextensively with a generalization parameter q, postulated a generalized formulation of entropy. The entropy is now well known as The Tsallis entropy, which, in the limit of q to 1, approaches the Shannon entropy. The Tsallis entropy is applicable to numerous fields, including physics, bioscience, chemistry, networks, computer science, and so on [10–12]. Menard et al. [13, 14] investigated fuzzy clustering in the framework of nonextensive thermostatistics, and derived the possibilistic membership function by taking the possibilistic constraint into account.

On the other hand, based on the Tsallis entropy, Yasuda [15] defined another form of entropy for FCM, and then derived the membership function by maximizing this entropy within FCM [15]. After that, by combining the membership function with DA, a new fuzzy clustering algorithm (the Tsallis-DAFCM algorithm) has been developed. Tsallis-DAFCM was proved to yield superior results in comparison with the conventional annealing methods.

Similarly to SA, a performance of Tsallis-DAFCM strongly depends on how to decrease the temperature. Among the cooling functions for SA, the very fast annealing (VFA) method decreases the temperature fastest. Thus, VFA is applied to Tsallis-DAFCM to improve its performance, and proved to be effective [16].

In spite of its performance, it remains unknown how appropriate q value and initial annealing temperature T_{high} for Tsallis-DAFCM should be determined according to the data distribution.

One of the important characteristics of the membership function of Tsallis-DAFCM is that centers of clusters are given as a weighted function of the membership function to the power of $q(u_{ik}^q)$. Furthermore, it changes its shape in a similar way by decreasing the temperature or by increasing q. In order to reveal the relationship between the temperature and q, u_{ik}^q is quantitatively analyzed using the Iris Data Set [17]. The result shows that the temperature and q affect u_{ik}^q almost inversely, suggesting that a q-incrementation algorithm is possible. This algorithm might be a solution to the initialvalue problem of Tsallis-DAFCM [18, 19].

This chapter is composed of five sections. Section 1 is this introduction. In Section 2, how DA works is explained generously, and its applications are summarized. In Section 3, each of the components of entropy-maximized FCM clustering methods is explained. In Section 4, VFA is used as the cooling function of Tsallis-DAFCM, and its effects are experimentally investigated. Effects of the temperature and q-values on the membership function of Tsallis-DAFCM are also examined. Section 5 gives a conclusion of this chapter.

2. Deterministic annealing

Vector quantization is a classification method for a big data, which is widely used in the field of image compression. Linde-Buzo-Gray [20] and self-organizing feature map [21] algorithms, for example, are well-known vector quantization algorithms. However, classification results of these algorithms depend on initial settings of the reference vector, and can easily fall into local minima.

In order to overcome the problem, by analogy with statistical mechanics, a DA method has been proposed. DA does not suffer from the initial vector problem. It usually performs better than other methods. However, it does not theoretically guarantee to find a global optimal solution. Furthermore, its performance depends on a way of decreasing the system temperature.

This section gives a brief introduction of the deterministic annealing method. In Section 2.1, a definition and major characteristics of DA are explained. In Section 2.2, DA is compared with SA. In Section 2.3, applications and modifications of DA are summarized.

2.1. Major characteristics of deterministic annealing

In DA, by analogy with statistical mechanics, the free energy F is derived from an objective function J of a problem. At a high temperature, F represents a global structure of J. As the temperature decreases, it gradually reaches J.

Based on these characteristics, at the high temperature, DA is able to find a global minimum of F by the steepest descent method because it should have multiple local minima. When the temperature is lowered a little, F would change its shape only a little. Accordingly, by setting the previous global mimimum as an initial value of the steepest descent mehod, DA searches

the next global minimum. This procedure continues until the temperature is lowered suffi-ciently, and *F* reaches *J*. Consequently, at each temperature, DA searches the local minimum of *J* deterministically.

2.2. Comparison with simulated annealing

While decreasing the temperature, SA searches the minimum stochastically at each tempera-ture and thus requires a very long time to find an optimal solution. Hence, though theoretically, it is guaranteed to find the optimal solution, SA is practically an approximation method.

On the contrary, DA consumes less computational time because it searches the local minimum deterministically at each temperature. Furthermore, it should be noticed that, in case that multiple local minima exist at some temperature, DA might not be able to find the minimum. For this reason, even theoretically, DA is not guaranteed to find the optimal solution.

Approaches to speed up SA are mainly based on the improvement of a transition method and a cooling function including its parallelization. For example, adaptive SA (ASA) [22], which may belong to the both categories, is an implementation of very fast simulated re-annealing (VFSA) [23]. As compared with the acceleration method called fast annealing (FA) [24] in which the temperature is lowered inversely linear to a number of iterations, ASA is faster than FA. In addition, among many features included in ASA, it can get the benefit of speeding-up by simulated quenching.

In DA, it seems no comprehensive studies on this topic have been conducted.

The summary of comparison is shown in **Table 1**.

	SA	DA
Search strategy	Stochastic search based on the Metropolis algorithm.	Deterministic search based on the steepest descent algorithm.
Cooling function	Two categories of cooling functions are well-used. 1. Functions based on statistical analysis. 2. Adaptive functions depending on the problems.	Cooling functions appear in SA are used empirically.
Optimality	The global minimum can be achieved if a temperature is decreased as slow as $T \propto 1/\log(\text{iterations})$.	Not guaranteed.

Table 1. Comparison of SA and DA.

2.3. Applications and modifications of deterministic annealing

The study on DA first addressed avoidance of the poor local minima of data clustering [25]. Then it was extensively applied to various subjects such as combinational optimization problems [26], vector quantization [27, 28], maximum likelihood estimation [29], classifier design [30], and pairwise data clustering [31].

In order to cluster a large number of data, research activities that attempt to parallelize DA have become popular. Kim et al. [32] discussed the parallelization method of DA using GPU, and applied it to color image segmentation. In order to cluster a large number of bioscientific data, Fox et al. [33, 34] parallelized DA using MPI. Qiu et al. [35] compared the DA's performance using C# messaging runtime library CCR with that using MPI.

3. Application of deterministic annealing to fuzzy c-means clustering maximized with entropy

One of the important applications of DA is fuzzy clustering. In this section, we focus on fuzzy c-means (FCM) clustering. By maximizing the objective function of FCM with various entropies, membership functions similar to the statistical mechanical distribution functions are obtained. These membership functions can be easily combined with DA, because they contain a parameter corresponding to a system temperature.

In this section, first we outline the formulation of FCM. Then, we describe how to apply the entropy maximization methods to FCM. In Section 3.1, the classical FCM clustering method is introduced. In Section 3.2, various entropy maximization methods are explained. The free energies for each entropy maximization method are derived in Section 3.3. In Section 3.4, representative annealing functions are presented.

3.1. Fuzzy c-means clustering

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

$$J = \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ik}{}^m d_{ik} \quad \left(d_{ik} = \| \mathbf{x}_k - \mathbf{v}_i \|^2 ; \; m \in R; \; 1 < m \right) \tag{1}$$

be the objective function of FCM to be minimized. Under the normalization constraint of

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

the Lagrange function L is given by

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

where η_k denotes the Lagrange multiplier. $\partial L / \partial u_{ik} = 0$ gives the membership function of the form:

$$u_{ik} = \left[\sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}} \right)^{\frac{1}{m-1}} \right]^{-1}. \tag{4}$$

Similarly, v_i can be determined by $\partial L / \partial v_i = 0$ as follows:

$$\mathbf{v}_i = \frac{\sum_{k=1}^{n} u_{ik} \mathbf{x}_k}{\sum_{k=1}^{n} u_{ik}}. \tag{5}$$

Desirable cluster centers can be obtained by calculating Eq. (4) and Eq. (5) repeatedly.

3.2. Entropy maximization method for fuzzy c-means

3.2.1. Shannon entropy maximization

Shannon entropy for FCM takes the form:

$$S_{SE} = -\sum_{k=1}^{n} \sum_{i=1}^{c} u_{ik} \log u_{ik}. \tag{6}$$

By setting m to 1 in Eq. (1), under the normalization constraint of Eq. (2), the Shannon entropy functional is given by

$$\delta S_{SE} - \sum_{k=1}^{n} \alpha_k \delta \left(\sum_{i=1}^{c} u_{ik} - 1 \right) - \beta \sum_{k=1}^{n} \sum_{i-1}^{c} \delta \left(u_{ik} d_{ik} \right), \tag{7}$$

where α_k and β denote the Lagrange multipliers. The stationary condition for Eq. (7) leads to the following Gaussian membership function:

$$u_{ik} = \frac{e^{-\beta d_{ik}}}{\sum_{j=1}^{c} e^{-\beta d_{jk}}}, \tag{8}$$

and the same formula for v_i in Eq. (5).

3.2.2. Fuzzy entropy maximization

Fuzzy entropy for FCM is defined as [36]

$$S_{FE} = -\sum_{k=1}^{n}\sum_{i=1}^{c}\{u_{ik}\log u_{ik} + (1-u_{ik})\log(1-u_{ik})\}. \tag{9}$$

The fuzzy entropy functional is given by

$$\delta S_{FE} - \sum_{k=1}^{n}\alpha_k\delta\left(\sum_{i=1}^{c}u_{ik}-1\right) - \beta\sum_{k=1}^{n}\sum_{i-1}^{c}\delta\left(u_{ik}d_{ik}\right). \tag{10}$$

The stationary condition for Eq. (10) leads to the following membership function:

$$u_{ik} = \frac{1}{e^{\alpha_k + \beta d_{ik}} + 1}, \tag{11}$$

and the same formula for v_i in Eq. (5). Eq. (11) is similar to the Fermi-Dirac distribution function.

3.2.3. Tsallis entropy maximization

The Tsallis entropy for FCM is defined as [15]

$$S_{Tsallis} = -\frac{1}{q-1}\left(\sum_{k=1}^{n}\sum_{i-1}^{c}u_{ik}{}^{q} - 1\right), \tag{12}$$

where $q \in R$ is a real number. In case of the Tsallis entropy maximization, the objective function should be rewritten as

$$J_{Tsallis} = \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ik}{}^q d_{ik}. \tag{13}$$

Accordingly, the Tsallis entropy functional is given by

$$\delta S_{Tsallis} - \sum_{k=1}^{n} \alpha_k \delta \left(\sum_{i=1}^{c} u_{ik} - 1 \right) - \beta \sum_{k=1}^{n} \sum_{i-1}^{c} \delta \left(u_{ik}{}^q d_{ik} \right). \tag{14}$$

The stationary condition for Eq. (14) leads to the membership function of the form:

$$u_{ik} = \frac{\{1 - \beta(1-q)d_{ik}\}^{\frac{1}{1-q}}}{Z}, \tag{15}$$

where

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

v_i is defined as

$$\mathbf{v}_i = \frac{\sum_{k=1}^{n} u_{ik}{}^q \mathbf{x}_k}{\sum_{k=1}^{n} u_{ik}}. \tag{16}$$

3.3. Free energy for entropy maximized fuzzy c-means

In each entropy, maximization methods introduced in Section 3.2, β can be regarded as the inverse of the system temperature T^{-1}. This feature makes it possible to apply DA, and Shannon-entropy-maximized FCM with DA (Shannon-DAFCM, hereafter), fuzzy-entropy-maximized FCM with DA (fuzzy-DAFCM, hereafter), and Tsallis-entropy-maximized FCM with DA (Tsallis-DAFCM, hereafter) have been developed.

3.3.1. Free energy for Shannon-DAFCM

In Shannon-DAFCM, by analogy with statistical mechanics [37], the sum of the states (the partition function) for the grand canonical ensemble of FCM can be expressed as

$$\xi = \prod_{k=1}^{n} \sum_{i=1}^{c} e^{-\beta d_{ik}} \tag{17}$$

The free energy is derived as

$$F = -\frac{1}{\beta} \log \xi = -\frac{1}{\beta} \sum_{k=1}^{n} \log \sum_{i=1}^{c} e^{-\beta d_{ik}}. \tag{18}$$

Stable thermal equilibrium requires a minimization of the free energy. By formulating deterministic annealing as a minimization of the free energy, $\partial F / \partial v_i = 0$ yields the same expression for v_i as that in Eq. (5).

3.3.2. Free energy for fuzzy-DAFCM

Similarly, in fuzzy-DAFCM, the grand partition function for the grand canonical ensemble for FCM can be written as

$$\Xi = \prod_{k-1}^{n} \prod_{i=1}^{c} \left(1 + e^{-\alpha_k - \beta d_{ik}}\right). \tag{19}$$

The free energy is calculated as

$$F = -\frac{1}{\beta} \left(\log \Xi - \alpha_k \frac{\partial \log \Xi}{\partial \alpha_k} \right) = -\frac{1}{\beta} \sum_{k=1}^{n} \left\{ \sum_{i=1}^{c} \log\left(1 + e^{-\alpha_k - \beta d_{ik}}\right) + \alpha_k \right\}. \tag{20}$$

$\partial F / \partial v_i = 0$ yields the same expression for v_i as that in Eq. (5).

3.3.3. Free energy for Tsallis-DAFCM

In Tsallis-DAFCM, the free energy can be derived as

$$F = J_{Tsallis} - TS_{Tsallis} = -\frac{1}{\beta} \sum_{k=1}^{n} \frac{Z^{1-q} - 1}{1-q}, \tag{21}$$

where T is a system temperature. $\partial F / \partial v_i = 0$ yields the same expression for v_i as that in Eq. (16).

3.4. Effect of annealing temperature on clustering

3.4.1. Dependency of shape of membership function on temperature

While reducing the system temperature T, DA achieves thermal equilibrium at each temperature by minimizing the free energy. Thus, DA searches a cluster distribution that minimizes the free energy at each temperature. When the temperature is high, the membership functions distribute widely. This makes clusters to which a data belong fuzzy. In case of Tsallis-DAFCM, when q is nearly equal to 2, the width of the membership function is almost proportional to $\sqrt{T}$. On the contrary, at the low temperature, fuzzy clustering approaches hard clustering. The relationship $F = J_{\text{Tsallis}} - TS_{\text{Tsallis}}$ in Eq. (21) suggests that, at the higher temperature, the larger entropy state or chaotic state is caused by a widening of the extent of the membership function.

3.4.2. Cooling function

In SA, the temperature decreases according to a cooling function or an annealing schedule. The representative cooling functions for SA [38] are:

(I) Exponential function

$$T = T_{high} r^t, \tag{22}$$

where T_{high} is the highest initial temperature, r is a parameter which defines a temperature reduction rate, and t is a number of iterations of temperature reduction.

(II) Inversely linear function

$$T = \frac{T_{high}}{t}. \tag{23}$$

(III) Inversely logarithmic function

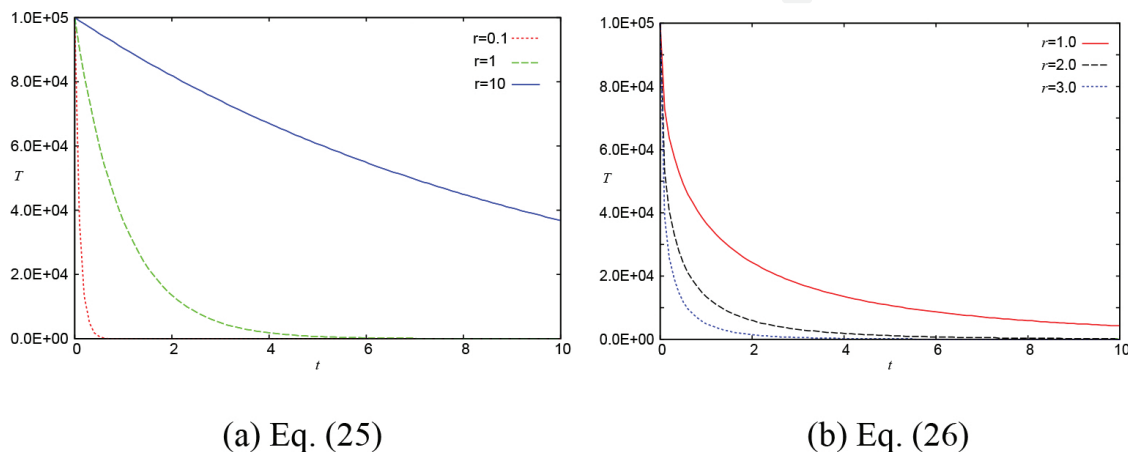$$T = \frac{T_{high}}{\ln t}. \tag{24}$$

(IV) Inversely exponential function

$$T = \frac{T_{high}}{e^{rt}}. \tag{25}$$

(V) Very fast annealing

Rosen [39] proposed another inversely exponential function known as VFA. VFA decreases the temperature exponentially in a similar way to ASA:

$$T = \frac{T_{high}}{e^{rt^{(1/D)}}},\tag{26}$$

where D is a dimension of a state space.[1] **Figure 1** compares plots of Eq. (25) and Eq. (26).

(a) Eq. (25)   (b) Eq. (26)

Figure 1. Plots of (a) Eq. (25) and (b) Eq. (26) ($T_{high} = 1.0 \times 10^5$, $D = 2$).

4. Tsallis-entropy-maximized fuzzy c-means clustering with deterministic annealing

In this section, we focus on Tsallis-DAFCM, and its important experimental results are explained. In Section 4.1, we present the Tsallis-DAFCM clustering algorithm. In Section 4.2, how VFA affects Tsallis-DAFCM is experimentally investigated. In Section 4.3, effects of the temperature and q-values on the membership function are examined.

4.1. Tsallis-DAFCM clustering algorithm

The Tsallis-entropy-maximization method, fuzzy c-means clustering, and the deterministic annealing method can be combined as the following Tsallis-DAFCM clustering algorithm [15]:

1. Set the number of clusters c, the highest temperature T_{high}, the temperature reduction rate m, and the threshold of convergence test δ_1 and δ_2;

---

[1] In clustering, a state space refers to an input space of a data set. For example, if a data set consists of 3 attributes, D is set to 3.

2. Generate c clusters at random positions. Set current temperature T to T_{high};

3. Calculate the membership function u_{ik} by Eq. (15);

4. Calculate the centers of clusters v_i by Eq. (17);

5. Compare the difference between the current cluster centers v_i and the cluster centers obtained in the previous iteration $\bar{v}_i$. If the convergence condition $max_{1 \leq i \leq c}\left|\left|\mathbf{v}_i - \bar{\mathbf{v}}_i\right|\right| < \delta_1$ is satisfied, then go to **6**, otherwise go back to **3**;

6. Compare the difference between the current cluster centers and the cluster centers obtained at the previous temperature $\bar{\bar{v}}_i$. If the convergence condition $max_{1 \leq i \leq c}\left|\left|\mathbf{v}_i - \bar{\bar{\mathbf{v}}}_i\right|\right| < \delta_2$ is satisfied, then stop, otherwise decrease the temperature using the cooling function, and go back to **3**.

4.2. Effect of cooling functions

In general, the temperature should be reduced gradually in DA. However, this takes a long time to converge. If VFA is applicable to Tsallis-DAFCM, it is of great advantage to this method. Accordingly, VFA is tested as a cooling function of Tsallis-DAFCM.

In this subsection, both Shannon- and Tsallis-DAFCM are examined.

4.2.1. Experiment 1

In experiment 1, the numerical data composed of five clusters and 2000 data points are used, as shown in **Figure 2**. The parameters are set as follows: c = 10, δ_1 = 50, δ_2 = 2, q = 1.5, T_{high} = 1.0×10⁶ or 1.0×10⁵.
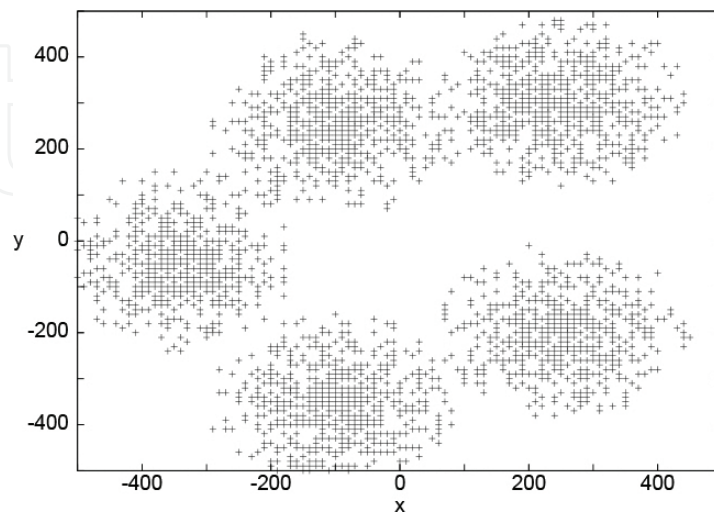

Figure 2. Numerical data.

First, the inversely exponential cooling function is applied to Tsallis-DAFCM. The changes of β parameterized by r are plotted in **Figure 3**. In case of $r = 1000.0$, when $T_{high} = 1.0\times10^5$, as T is lowered from **Figures 3 (A)** to **(D)**, data are clustered gradually and desirably. In case of $r = 10.0$ and $r = 1.0$ (**Figures 3 (E)** and **(F)**, respectively), it is observed that u_{ik} and v_i converge more rapidly.
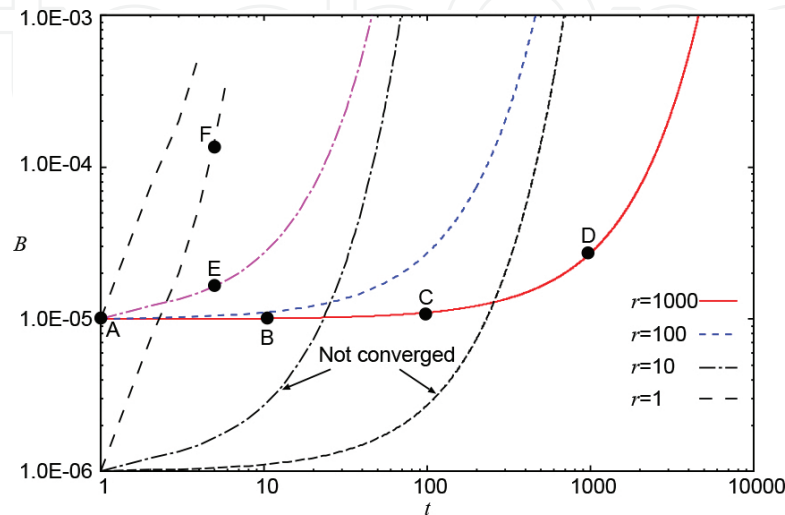
Figure 3. Increasing of β by inversely exponential cooling function.

However, when $T_{high} = 1.0\times10^6$, the algorithm fails to converge with $r = 100.0$ and 10.0 (expressed by "Not converged" in **Figure 3**) because the initial distribution of u_{ik} is too wide. This result indicates that it is important to set both T_{high} and r properly.

(a) Shannon-DAFCM
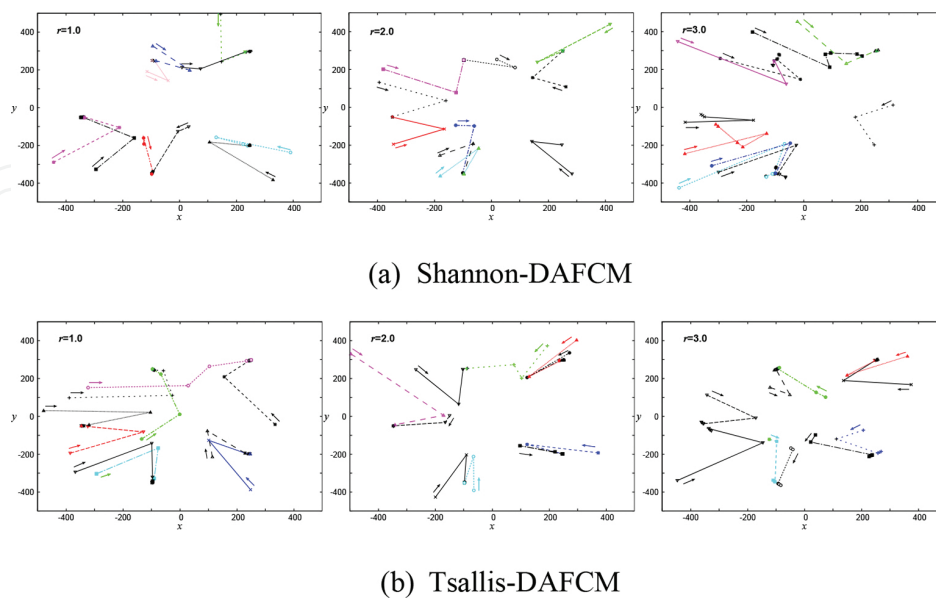
(b) Tsallis-DAFCM

Figure 4. Shifts of cluster centers during clustering obtained by (a) Shannon-DAFCM and (b) Tsallis-DAFCM.

In order to clarify the adaptability of VFA as a cooling function of DA, numerical experiments on Shannon- and Tsallis-DAFCM are performed. The shifts of cluster centers with decreasing temperature are illustrated in **Figures 4 (a)** and **(b)**. Initially, clusters are located randomly. Then, at the higher temperature, β is comparatively small and clusters move to near the center of gravity of data because the membership functions are extend over the data area and become extremely uniform. As T is lowered, contrarily, the membership functions become narrower and the associations of data to the clusters become less fuzzy. In this process, in Shannon-DAFCM, the clusters move to their nearest local data distribution centers. However, in Tsallis-DAFCM, the clusters can move a long distance to optimal positions because the membership functions have gentle base slopes.

Figures 5 (a) and **(b)** illustrates the three-dimensional plots of u_{ik} in the progress of Shannon- and Tsallis-DAFCM clustering combined with VFA. When the temperature is as high as 3.7×10^4, roughness of u_{ik} of Tsallis-DAFCM is smaller than that of Shannon-DAFCM. After that, the shapes of both membership functions do not change greatly, because VFA reduces the temperature extremely only at the early annealing stage. When the temperature is lowered to 1.3×10^4, both methods cluster data desirably.

(a) Shannon-DAFCM
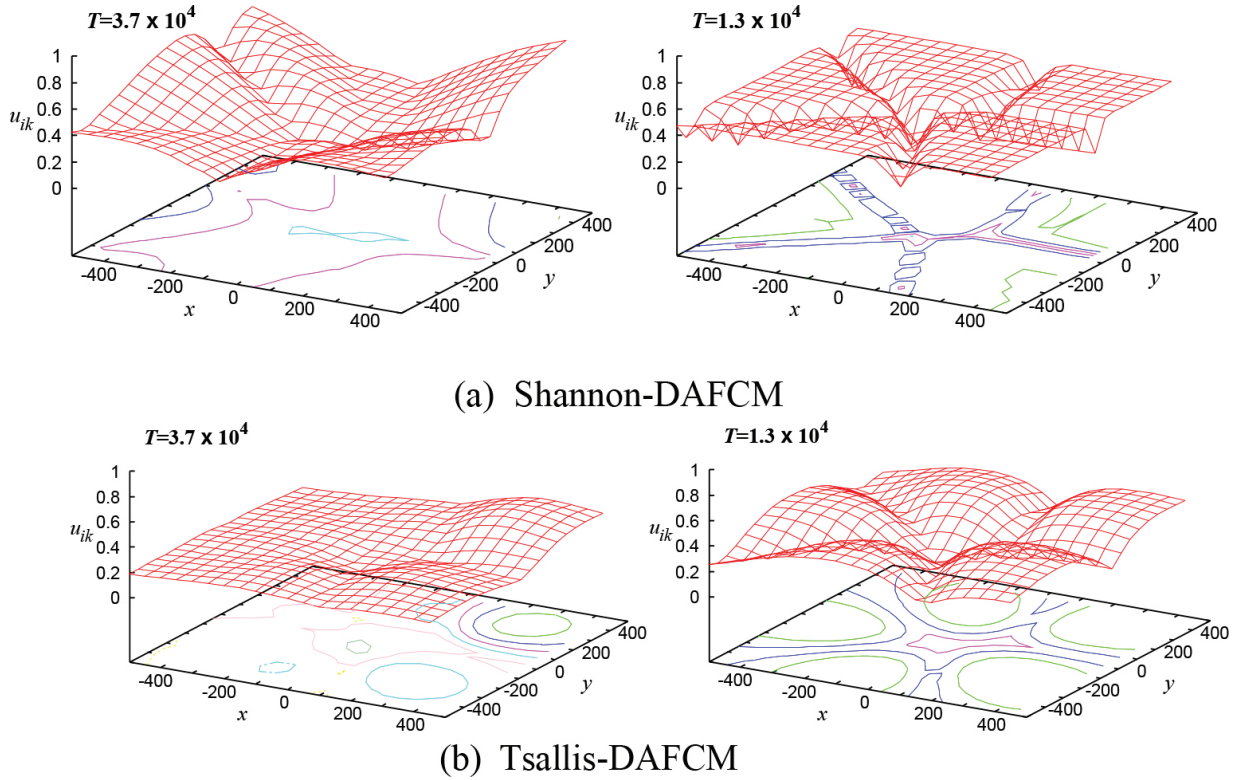
(b) Tsallis-DAFCM

Figure 5. Initial and final landscapes of u_{ik} of (a) Shannon-DAFCM and (b) Tsallis-DAFCM.

Consequently, because Tsallis-DAFCM has gentle slope in the region far from the origin, clusters can move long distance to optimal positions stably. This feature makes it possible to reduce the temperature rapidly. Thus, VFA is suitable as a cooling function of Tsallis-DAFCM.

On the other hand, final cluster positions obtained by Shannon-DAFCM tend to depend on their initial positions.

4.2.2. Experiment 2

In experiment 2, the Iris Data Set [17] consisting of 150 four-dimensional vectors of iris flowers is used. Three clusters of flowers detected are Versicolor, Virginia, and Setosa. Each cluster consists of 50 vectors. VFA is used as a cooling function of DA. The parameters are set as follows: $c = 3$, $\delta_1 = 0.1$, $\delta_2 = 0.01$, $q = 1.5$, $T_{high} = 2.0$.

The minimum, maximum, and average values of misclassified data points of 100 trials are summarized in **Table 2**. It can be seen that Shannon-DAFCM gives slightly better results than Tsallis-DAFCM. However, it is also confirmed that Tsallis-DAFCM gives its best results when the temperature reduction rate r is set to 1 or 2, though the best result for Shannon-DAFCM is obtained only when $r = 2$. Furthermore, variances of Tsallis-DAFCM are smaller than those of Shannon-DAFCM. These features indicate that a wide range of r values are applicable to Tsallis-DAFCM. On the other hand, with larger r values, Shannon-DAFCM becomes unstable.

Shannon-DAFCM				Tsallis-DAFCM			
r	Min.	Max.	Ave.	r	Min.	Max.	Ave.
1	15	16	15.01	1	14	14	14.00
2	11	13	12.97	2	14	15	14.01
3	11	14	13.59	3	14	15	14.68

Table 2. Misclassified data points of Iris Data Set (100 trials).
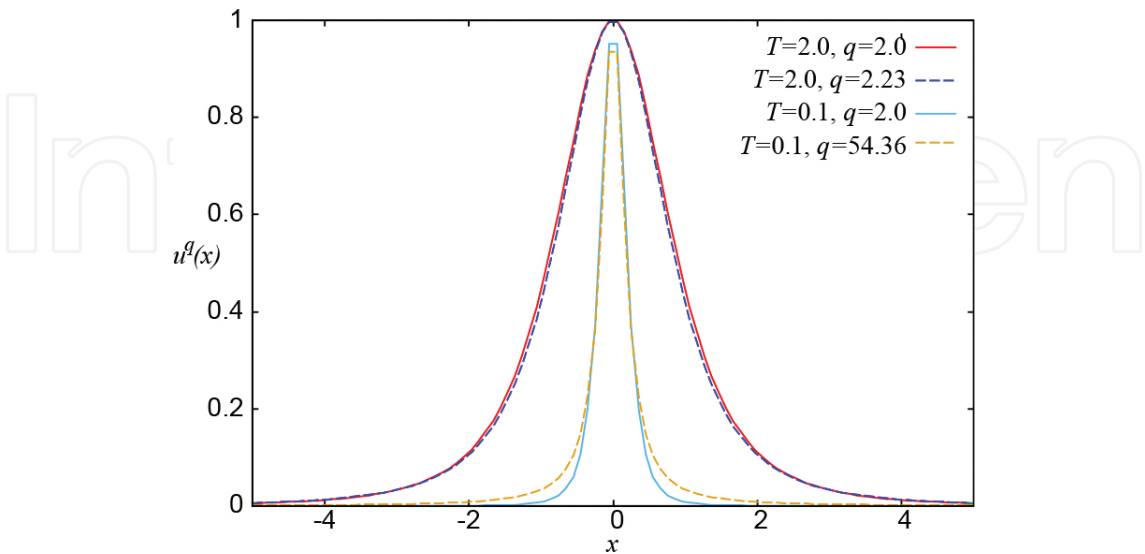

Figure 6. Plots of $u^q(x)$.

4.3. Dependencies of the membership function on temperature and q

In Eq. (16), it can be seen that u_{ik}^q plays an important role as a weight value to each x_k, and it determines v_i. For this reason, dependencies of u_{ik}^q on T and q are to be investigated.

In this subsection, for simplicity, v_i is set to be 0, because this makes the denominator of Eq. (15) become the sum of the same formulas of its numerator. **Figure 6** illustrates the numerator of u_{ik}^q (expressed by $u^q(x)$) as a function of x_k, parameterized by T and q, respectively. In order to plot the shape of u_{ik}^q as a function of the distance between the cluster center and various data points, in this figure, x_k is replaced to a continuous variable x. **Figure 6** confirms that the extent of u_{ik}^q becomes narrower with increasing q. On the contrary, as the temperature de-creases, the distribution becomes narrower.

4.3.1. Quantitative relationship between temperature and q

As stated in the previous subsection, T and q inversely affect the extent of u_{ik}^q, which changes in a similar way with decreasing T or increasing q. In order to examine the quantitative relationship between T and q in more detail, they are changed independently as follows:

First, we define

$$u^q(x,T,q) = \left\{ 1 - \frac{1-q}{T} x^2 \right\}^{\frac{q}{1-q}} . \tag{27}$$

Then, Eq. (27) is calculated by fixing T and q to some constants T_0 and q_0. Next, by decreasing T, we search the q values that minimize the sum of squares of the residuals of the following two functions:

$$\sum_{S=0}^{Smax} \left| u^q(\Delta x S, T_0, q_0) - u^q(\Delta x S, T, q) \right|^2 . \tag{28}$$

In the following calculations, the parameters are set as follows: $T_{high}(= T_0) = 2.0$; the domain of x is $0 \leq x \leq 100$; the number of sampling points of the sum of residuals $S_{max} = 10{,}000$; $\Delta x = 0.01$.

For $q (= q_0)$ values of 1.01, 2.0, 6.0, 10.0, and for T decreasing from T_{high}, the q values that minimize Eq. (28) (expressed by q_{min}) are shown in **Figure 7 (a)**. **Figure 7 (b)**, on the other hand, shows the results of cases in which q is set to 2.0 and T is lowered from $T_{high} = 2.0, 20.0, 100.0, 200.0$. Approximate curves plotted in **Figure 7** are obtained by fitting the data to the following exponential function:
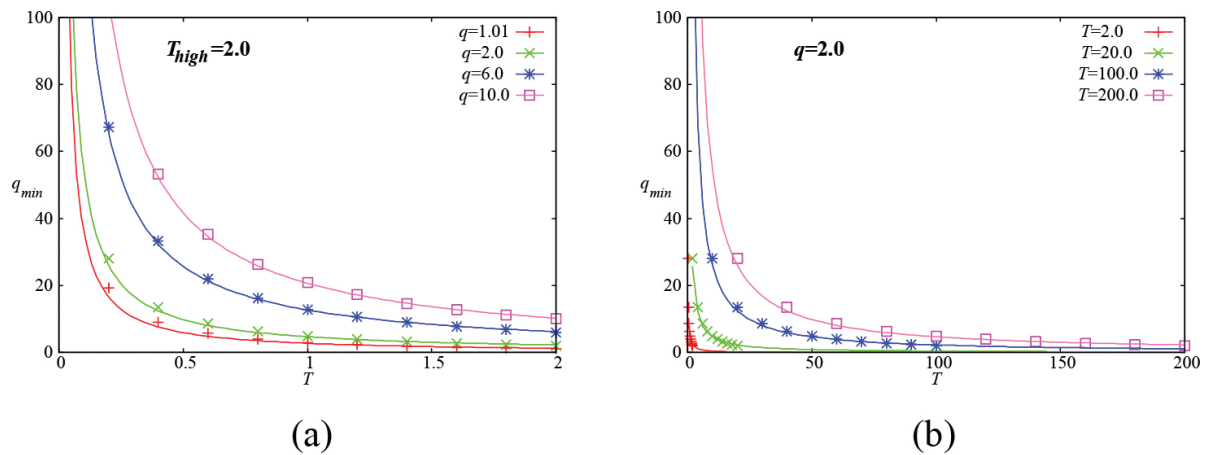
Figure 7. Plots of q_{min} as a function of T parameterized by (a) q and (b) T_{high}.

$$q_{\min} = aT^{-b}, \tag{29}$$

where a and b denote the fitting parameters. Optimal values for these parameters obtained by the least squares method are summarized in **Table 3** and **Table 4**. From these tables, it is concluded that b is nearly equal to 1.0 indicating that q is inversely proportional to T. In addition, it can be seen that, though b does not change its value much, a increases with increasing T.

q	a	b
1.01	2.71	1.126
2	4.67	1.066
6	12.65	1.023
10	20.64	1.014

Table 3. Parameters of approximate curves (T_{high} = 2.0).

T_{high}	a	b
2	4.67	1.066
20	54.33	1.066
100	302.06	1.066
200	632.35	1.066

Table 4. Parameters of approximate curves (q = 2.0).

As a result, by using the approximate relationship of T and $q_{\min}$, instead of annealing or T-reduction, q-incrementation clustering might be possible [18].

5. Conclusion

In this chapter, we first explained the major characteristics of DA and compared it with SA. DA is a variant of SA and searches a minimum deterministically. Thus, generally it is more efficient than SA. We then explained how DA could be applied to the fuzzy c-means clustering by employing the entropy maximization method.

After that, by focusing on Tsallis-entropy-maximized FCM combined with DA (Tsallis-DAFCM), an effect of VFA on DA was examined. VFA reduces the temperature extremely only at the early annealing stage, and the experimental result showed that this feature improved the performance of Tsallis-DAFCM because it has gentle slope in the region far from the origin and clusters can move long distance to optimal positions from the beginning.

The Tsallis entropy is an extension of the Shannon entropy with a generalization parameter q. A shape of a membership function of Tsallis-DAFCM strongly depends on both the temperature and q. Accordingly, a relationship between the temperature and q was quantitatively investigated, and it was experimentally confirmed that they affected the area covered by the membership function almost inversely. Based on the result, a development of a q-incrementation algorithm is our future subject.

Author details

Makoto Yasuda

Address all correspondence to: yasuda@gifu-nct.ac.jp

National Institute of Technology, Gifu College, Gifu, Japan

References

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

[2] S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi. Optimization by simulated annealing. Science. 1983; 220: 671–680.

[3] K. Rose, E. Gurewitz, B. C. Fox. A deterministic annealing approach to clustering. Pattern Recognition Letters. 1990; 11(9): 589–594.

[4] J.C. Bezdek. Pattern Recognition with Fuzzy Objective Function Algorithms. New York: Prenum Press; 1981.

[5] R.-P. Li, M. Mukaidono. A maximum entropy approach to fuzzy clustering. In: Proceedings of the 4th IEEE International Conference on Fuzzy Systems (FUZZ-IEEE/IFES '95); 1995 March 20–24; Yokohama, Japan; 1995. p. 2227–2232.

[6] S. Miyamoto, M. Mukaidono. Fuzzy c-means as a regularization and maximum entropy approach. In: Proceedings of the 7th International Fuzzy Systems Association World Congress; 1997 June 25–29; Prague, Czech Republic; 1997. p. 86–92.

[7] A. DeLuca, S. Termini. A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory. Information and Control. 1972; 20: 301–312.

[8] L. D. Landau, E. M. Lifshitz. Statistical Physics (Part 1). Oxford: Butterworth Heinemann; 1980.

[9] C. Tsallis. Possible generalization of Boltzmann-Gibbs statistics. Journal of Statistical Physics. 1988; 52(1–2): 479–487.

[10] S. Abe, Y. Okamoto, editors. Nonextensive Statistical Mechanics and Its Applications. New York: Springer; 2001.

[11] M. Gell-Mann, C. Tsallis, editors. Nonextensive Entropy—Interdisciplinary Applications. New York: Oxford University Press; 2004.

[12] C. Tsallis, editor. Introduction to Nonextensive Statistical Mechanics. New York: Springer; 2009.

[13] M. Menard, V. Courboulay, P.-A. Dardignac. Possibilistic and probabilistic fuzzy clustering: unification within the framework of the non-extensive thermostatistics. Pattern Recognition. 2003; 36(6): 1325–1342.

[14] M. Menard, P. Dardignac, C. C. Chibelushi. Non-extensive thermostatistics and extreme physical information for fuzzy clustering. International Journal of Computational Cognition. 2004; 2(4): 1–63.

[15] M. Yasuda. Entropy maximization and very fast deterministic annealing approach to fuzzy c-means clustering. In: Proceedings of the 5th Joint International Conference on So Computing and 11th International Symposium on Intelligent Systems; 8–12 December 2010; Okayama, Japan; 2010. p. 1515–1520.

[16] M. Yasuda. Deterministic annealing approach to fuzzy c-means clustering based on entropy maximization. Advances in Fuzzy Systems. 2011; 960635: 9.

[17] UCI Machine Learning Repository: Iris Data Set [Internet]. 1988. Available from: http://archive.ics.uci.edu/ml/datasets/Iris

[18] M. Yasuda. Q-increment deterministic annealing fuzzy c-means clustering using Tsallis entropy. In: Proceedings of the 11th International Conference on Fuzzy Systems and Knowledge Discovery (FSKS); 19–21 August 2014; Xiamen, China; 2014. p. 31–35.

[19] M. Yasuda. Quantitative analyses and development of a q-incrementation algorithm for FCM with Tsallis entropy maximization. Advances in Fuzzy Systems. 2015; 404510: 7.

[20] Y. Linde, A. Buzo, R. M. Gray. An algorithm for vector quantizer design. IEEE Transaction on Communication. 1980; Com-28(1): 84–95.

[21] T. Kohonen. Self-Organizing Maps. 3rd ed. New York: Springer; 2000.

[22] L. Ingber. Adaptive Simulated Annealing. In: H. A. Oliveira, A. Petraglia Jr., L. Ingber, M. A. S. Machado, M. R. Petraglia (Eds.), Stochastic Global Optimization and Its Applications with Fuzzy Adaptive Simulated Annealing. New York: Springer; 2012. p. 33–61.

[23] L. Ingber. Very fast simulated re-annealing. Mathematical Computer Modelling. 1989; 12(8): 967–973.

[24] H. Szu, R. Hartley. Fast simulated annealing. Physics Letters A. 1987; 122(3–4): 157–162.

[25] K. Rose. Deterministic annealing for clustering, compression, classification regression, and related optimization problems. Proceedings of the IEEE. 1998; 86(11): 2210–2239.

[26] K. Rose, E. Gurewitz, G. C. Fox. Constrained clustering as an optimization method. IEEE Transaction on Pattern Analysis and Machine Intelligence. 1993; 15(8): 785–794.

[27] J. Buhmann, H. Kuhnel. Vector quantization with complexity costs. IEEE Transaction on Information Theory. 1993; 39(4): 1133–1143.

[28] K. Rose, E. Gurewitz, G. C. Fox. Vector quantization by deterministic annealing. IEEE Transaction on Information Theory. 2002; 38(4): 1249–1257.

[29] N. Ueda, R. Nakano. Mixture density estimation via EM algorithm with deterministic annealing. In: Proceedings of 1994 IEEE Neural Networks for Signal Processing; 6–8 September 1994; Ermioni, Greek. IEEE; 1994. p. 69–77.

[30] D. Miller, A. V. Rao, K. Rose, A. Gersho. A global optimization technique for statistical classifier design. IEEE Transaction on Signal Processing. 1996; 44: 3108–3122.

[31] T. Hofmann, J. Buhmann. Pairwise data clustering by deterministic annealing. IEEE Transaction on Pattern Analysis and Machine Intelligence. 1997; 19: 1–14.

[32] E. Kim, W. Wang, H. Li, X. Huang. A parallel annealing method for automatic color cervigram image segmentation. In: Medical Image Computing and Computer Assisted Intervention, MICCAI-GRID 2009 HPC Workshop; 2009 September 20–24; London, UK; 2009.

[33] G. C. Fox, D. R. Mani, S. Pyne. Detailed results of evaluation of DAVS(c) deterministic annealing clustering and its application to LC-MS data analysis. 2013. Available from: http://grids.ucs.indiana.edu/ptliupages/publications/DAVS2.pdf

[34] G. C. Fox, D. R. Mani. Parallel deterministic annealing clustering and its application to LC-MS data analysis. In: 2013 IEEE International Conference on Big Data; 6–9 October 2013; Silicon Valley, USA. IEEE; 2013. p. 665–673.

[35] X. Qiu, G. C. Fox, H. Yuan, S. Bae, G. Chrysanthakopoulos, F. Nielsen. Performance of multicore systems on parallel data clustering with deterministic annealing. In: International Conference on Computer Science 2008; 23–25 June 2008; Krakow, Poland. Springer-Verlag; 2008. p. 407–416.

[36] M. Yasuda, T. Furuhashi, S. Okuma. Statistical mechanical analysis of fuzzy clustering based on fuzzy entropy. IEICE Transaction on Information and Systems. 2007; E90-D(6): 883–888.

[37] L. E. Reichl. A Modern Course in Statistical Physics. New York: John Wiley & Sons; 1998.

[38] S. M. Sait, H. Youssef. Iterative Computer Algorithms with Applications in Engineering: Solving Combinatorial Optimization Problems. U.S.A.: Wiley-IEEE Computer Society Press; 2000.

[39] B. E. Rosen. Function optimization based on advanced simulated annealing. In: Proceedings of the IEEE Workshop on Physics and Computation (PhysComp '92); 2–4 October 1992; Dallas, U.S.A. IEEE; 1992. p. 289–293.