A Neural Interior Point Algorithm for Solving Hard and Fuzzy Clustering Problems

Hidetomo Ichihashi, Muneki Nemoto, Tomohiro Ohta, Tetsuya Miyoshi

Department of Industrial Engineering, College of Engineering,
Osaka Prefecture University
1-1 Gakuencho, Sakai, Osaka 593 Japan
ichi@ie.osakafu-u.ac.jp

Abstract This paper proposes an improved neural algorithm for solving hard clustering which is a mixed integer programming problem. The algorithm is also applicable to a fuzzy clustering with maximizing entropy by Li and Mukaidono. The interior point method using affine scaling for linear programming problems is employed. The proposed method is an improved version of the Hopfield networks, and outperforms the gradient projection method and the conventional k-means algorithm in obtaining optimum solutions.

Keywords:Interior point algorithm, Neural network, Fuzzy clustering, Affin transformation, K-means

1. Introduction

Cluster analysis is a tool for numerical classification or mathematical taxonomy. Among others K-means algorithm is a standard procedure of non-hierarchical clustering[3, 5, 6, 16] and it can be regarded as a combinatorial optimization problem which is formulated as a mixed integer programming problem. The Hopfield networks[7, 8] are familiar neural algorithms for solving combinatorial optimization problems. The deficiency of the network is their week reliability to obtain feasible and optimal solutions. At an early stage, it did not seem likely to lead to better algorithms for finding acceptably low-cost minima[10]. The conventional algorithm already in use clearly outperformed the Hopfield networks both in speed and the quality of solution found. And, careful adjustment of the value of parameters is required to raise their yield rates[2]. Therefore much effort has been made paying attention to achieving better solutions by modifying algorithms. The adaptive schemes have been proposed[20, 22, 23]. Theoretical aspects on the stability of solution in a multi-linear form of the energy function have been studied[1, 21]. It was shown that the Hopfield model is effective when it is endowed with an appropriate form and parameters of the energy function with the term restricting non-integer solution[8]. The conditions of asymptotic stability of solutions and the non-stability of non-solutions have been studied[15]. Most of these studies are based on the stability at a vertex of a hypercube of the solution space. Actually, in most cases, the local minimum of the energy function with general quadratic form exists inside the hypercube.

In this paper we propose a revised algorithm to obtain a better solution of the clustering problem by employing the interior point method using affin scaling for linear programming problems. By selecting proper values of parameters of the energy function, better solutions of the hard clustering problems are obtained with reasonable certainty. The numerical experiments reveal that the proposed method outperforms the Hopfield networks, the gradient projection method and the conventional k-means algorithm in obtaining optimum solutions. The algorithm is also effective to a fuzzy clustering problem with maximizing entropy by Li and Mukaidono[13, 14]. The problem formulation is generalized as to find out a solution taking tradeoffs between the error criterion and entropy into account.

2. Neural Clustering by Minimizing Hopfield Energy

A typical neural network mathematical model proposed by Hopfield[7] is given by a system of ordinary differential equations as:

$$\frac{du_i}{dt} = \sum_j w_{ij} v_j + h_i \tag{1}$$

with

$$v_i = s(u_i) = \frac{1}{2}(1 + \mathbf{tanh}\frac{u_i}{x_0}),$$

$$i = 1, \ldots, m \tag{2}$$

where,$\boldsymbol{u}_i$ is the intermediate state variable, $v_i \in (0,1)$ is the output of neuron i, h_i is the external input of neuron i, $w_{ij} (w_{ij} = w_{ji})$ is the synaptic connection of neuron j to neuron i, $x_0 (> 0)$ is the gain scaling parameter of the sigmoid function s. The state of a network of N neurons is represented by a vector $v = (v_1, \ldots, v_N) \in (0,1)^N$, namely by a point in a unit hypercube.

It has been shown by Hopfield that the network converges to a set of stable states and these stable states correspond to a local minimum of the following energy function as:

$$E = -\frac{1}{2}\sum_i \sum_j w_{ij} v_i v_j - \sum_i h_i v_i \tag{3}$$

E certainly decreases as the time changes, and converges to the stable equilibrium point. Strictly speaking, in case that the network operates with the initial value in a hypercube, it converges to an asymptotic stable vertex or a local minimum inside the hypercube.

The energy function E for the clustering is given as:

$$\begin{aligned} E &= \frac{1}{2}\sum_{i=1}^{N}(\sum_{j=1}^{M} x_{ij} - 1)^2 + \frac{A}{2}\sum_{i=1}^{N}\sum_{j=1}^{M} x_{ij}(1 - x_{ij}) \\ &+ \frac{B}{2}\sum_{i=1}^{N}\sum_{j=1}^{M} x_{ij}||\boldsymbol{d}_i - \boldsymbol{r}_j||^2 \end{aligned} \tag{4}$$

where x_{ij} denotes membership of the pattern i to the cluster j. d_i is a numerical attribute value of the pattern j. r_j is a center of cluster j. A and B are constants. The first term represents a constrait that each pattern belongs to a cluster, and the second term restricts noninteger solutions. The third term represents the total distance from cluster centers. We solve the problem by using a nonlinear programing technique.

The parameter update rules by a simple steepest descent method is as:

$$u_{ij}^{NEW} = u_{ij}^{OLD} - \tau \frac{\partial E}{\partial u_{ij}} \tag{5}$$

$$\frac{\partial E}{\partial u_{ij}} = \frac{\partial E}{\partial x_{ij}} \cdot \frac{\partial x_{ij}}{\partial u_{ij}} \tag{6}$$

$$\begin{aligned} \frac{\partial E}{\partial x_{ij}} &= (\sum_{l=1}^{M} x_{il} - 1) + A(\frac{1}{2} - x_{ij}) + \\ &\frac{B}{2}||\boldsymbol{d}_i - \boldsymbol{r}_j||^2 \end{aligned} \tag{7}$$

$$r_{jk}^{NEW} = r_{jk}^{OLD} - \tau \frac{\partial E}{\partial r_{jk}} \tag{8}$$

$$\frac{\partial E}{\partial r_{jk}} = B \sum_{i=1}^{N} x_{ij}(r_{jk} - d_{ik}) \tag{9}$$

where τ is a small positive number. By simplifying x_{ij} as:

$$x_{ij} = \begin{cases} 1 & ; u_{ij} > 1 \\ u_{ij} & ; 0 \leq u_{ij} \leq 1 \\ 0 & ; u_{ij} < 0 \end{cases} \tag{10}$$

Eqs.(5) and (6) can be replaced by

$$x_{ij}^{NEW} = x_{ij}^{OLD} - \tau \frac{\partial E}{\partial x_{ij}} \tag{11}$$

A solution (i.e., a feasible solution which forms a partition) is obtained such that all x_{ij} equal 0 or 1, namely it is on the vertex $X = (X_{11}, \ldots, X_{NM}) \in [0,1]^{N \times M}$ of a hypercube where the first and the second terms of the energy function equal 0. Eq. (4) is a multi-linear function when A=2, and $\boldsymbol{x}$ converges to a certain vertex. But it is known that by an energy function of the multi-linear form, to obtain a better solution is relatively hard compared with the general energy function of quadratic form. Eq. (1) is not a multi-linear form when $A < 2$, and $\boldsymbol{x}$ converges to a local minimum in a hypercube of the solution space. In order to obtain a solution corresponding to a vertex of the hypercube, we introduce two threshold values θ_L and θ_U for selecting values of x_{ij} from 1 or 0. When repeating the learning rule, if $x_{ij} \geq \theta_U$ then x_{ij} is set to 1 and if $x_{ij} \leq \theta_L$ then x_{ij} is set to 0. In short, we put x_{ij} on a side face, an edge, or a vertex of the hypercube if x_{ij} is outside the interval $[\theta_L, \theta_U]$.

For choosing a proper threshold value, we focus on the energy surface between a vertex of hypercube corresponding to a solution and its adjoining vertex. Followings are mainly based on the results in [17]
.

1) If X_{ab} =1 at a vertex $\boldsymbol{X}$ corresponding to a solution.

There exist cluster b containing pattern c. For any c $(\neq b)$, $x_{ac} = 0$. Hence,

$$\frac{\partial E}{\partial x_{ab}} = [x_{ab} - 1] + A[\frac{1}{2} - x_{ab}] + \frac{B}{2}||\boldsymbol{d}_a - \boldsymbol{r}_b||^2 \tag{12}$$

and at an adjoining vertex

$$\left.\frac{\partial E}{\partial x_{ab}}\right|_{x_{ab}=0} = -1 + \frac{A}{2} + \frac{B}{2}||\boldsymbol{d}_a - \boldsymbol{r}_b||^2 \qquad (13)$$

If the K dimensional pattern d_i takes its value in a hypercube $[0,1]^K$, then

$$||\boldsymbol{d}_a - \boldsymbol{r}_b||^2 << 1 \qquad (14)$$

And, if we choose a proper negative value A, we have

$$\left.\frac{\partial E}{\partial x_{ab}}\right|_{x_{ab}=0} < 0 \qquad (15)$$

Since,

$$\left.\frac{\partial E}{\partial x_{ab}}\right|_{x_{ab}=1} = -\frac{A}{2} + \frac{B}{2}||\boldsymbol{d}_a - \boldsymbol{r}_b||^2 \qquad (16)$$

If A is negative, then

$$\left.\frac{\partial E}{\partial x_{ab}}\right|_{x_{ab}=1} \geq 0 \qquad (17)$$

Consequently, x_{ij} at a local minimum of E exists inside the unit interval $[0,1]$. Since $X_{ab} = 1$, if A decreases, the solution x_{ij} moves left. If we set A such that the local minimum point is close to 0.5, x_{ij} is hard to converge to 0 or 1. Hence, it is expected to find a better solution of $\boldsymbol{r}_j$. The local minimum solution is

$$x_{ab} = \frac{1 - \frac{A}{2} - \frac{B}{2}||\boldsymbol{d}_a - \boldsymbol{r}_b||^2}{1-A} \qquad (18)$$

The objective of clustering is to minimize the distance from the cluster centers. Therefore B should be positive large. Then the x_{ij} which minimizes E comes closer to X_{ab}=0 which is far from the solution X_{ab}=1.

2) If x_{ab}=1 at a vertex X corresponding to a solution.

At a feasible solution, if pattern $\boldsymbol{a}$ does not belong to cluster $\boldsymbol{b}$, then $\boldsymbol{x}_{ab}$=0 and we have

$$\frac{\partial E}{\partial x_{ab}} = x_{ab} + A[\frac{1}{2} - x_{ab}] + \frac{B}{2}||\boldsymbol{d}_a - \boldsymbol{r}_b||^2 \qquad (19)$$

$$\left.\frac{\partial E}{\partial x_{ab}}\right|_{x_{ab}=0} = \frac{A}{2} + \frac{B}{2}||\boldsymbol{d}_a - \boldsymbol{r}_b||^2 \approx 0 \qquad (20)$$

$$\left.\frac{\partial E}{\partial x_{ab}}\right|_{x_{ab}=1} = 1 - \frac{A}{2} + \frac{B}{2}||\boldsymbol{d}_a - \boldsymbol{r}_b||^2 > 0 \qquad (21)$$

where A is negative and $|A| << 1$. From Eqs. (20) and (21) , when A=0, The local minimum of E exist at the left side of $\boldsymbol{x}_{ab}$=0 and

$$x_{ab} = \frac{-\frac{A}{2} - \frac{B}{2}||\boldsymbol{d}_a - \boldsymbol{r}_b||^2}{1-A} \qquad (22)$$

Since, $\partial x_{ab}/\partial A < 0$, when A decreases from 0, the local minimum point moves right. The above consideration aims to operate the positions of local minimum by the value of A. By setting A to a little smaller value than zero, x_{ij} does not converge to 0 or 1 quickly and hence better solutions are expected to obtain.

3. Neural Clustering by Affin Scaling Interior Point Method

A modified version of Hopfield network by using the gradient projection method of Rosen[19] is developed to clarify the effectiveness of the affin scaling interior point method. The gradient projection method and the affin scaling interior point method search for a solution being confined in a feasible region. The difference among these two approaches is whether it uses the affin scaling or not. We formulate the clustering problem as:

$$min \quad E(\boldsymbol{x},\boldsymbol{r}) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{M} x_{ij}||\boldsymbol{d}_i - \boldsymbol{r}_j||^2$$

$$+\frac{A}{2}\sum_{i}^{N}\sum_{j}^{M} x_{ij}(1-x_{ij}) \qquad (23)$$

$$s.t. \quad \sum_{j=1}^{M} x_{ij} = 1, \quad i = 1,\cdots,N \qquad (24)$$

$$x_{ij} \geq 0, \quad i = 1,\cdots,N, j = 1,\cdots,M \qquad (25)$$

We have applied the gradient projection method of Rosen in order to search for solutions in the simplexes defined by the linear constraints Eq. (24). As we see in the simulation study in section 3, most of the elements of vector $\boldsymbol{x}$ quickly tend to zero resulting in a local minimum solution. To improve this deficiency we apply the interior point method of linear programming problem to the clustering problem of Eqs.(23) $\sim$ (25). Though Eqs.(23) $\sim$ (25) are a nonlinear programming problem, the constraints are linear and the feasible region is the N simplexes of $M-1$ dimensions. The affin scaling interior point method of I.I.Dikin[4] is known as a simple and effective algorithm of linear programming.

An initial feasible solution is chosen as

$$x_{ij} = \frac{1}{M}, \qquad i = 1,\cdots,N, \quad j = 1,\cdots,M \qquad (26)$$

By the affin scaling method, the search for solution is carried out keeping $x_{ij} > 0$ with relatively large step size. Let $\boldsymbol{x}^* > \mathbf{o}$ be a feasible interior point, and a transformation which maps any $\boldsymbol{x}$ to $\boldsymbol{y} > \mathbf{o}$ be

$$y_{ij} = \frac{x_{ij}}{x^*_{ij}}, \qquad i = 1, \cdots, N, \quad j = 1, \cdots, M \quad (27)$$

Let D^*_i be a $M \times M$ diagonal matrix whose diagonal elements are $x^*_{ij}, \quad j = 1, \cdots, M$,then we have

$$\boldsymbol{y}_i = (D^*_i)^{-1}\boldsymbol{x}_i \quad (28)$$

Since

$$\boldsymbol{x}_i = D^*_i \boldsymbol{y}_i \quad (29)$$

the objective function (23) can be written as:

$$E(\boldsymbol{y}, \boldsymbol{r}) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{M} x^*_{ij} y_{ij} ||\boldsymbol{d}_i - \boldsymbol{r}_j||^2$$

$$+\frac{A}{2}\sum_{i=1}^{N}\sum_{j=1}^{M} x^*_{ij} y_{ij}(1 - x^*_{ij} y_{ij}) \quad (30)$$

The constraints (24) and (25) are written as:

$$\sum_{j=1}^{M} x^*_{ij} y_{ij} \; = \; 1, \qquad i = 1, \cdots, N \quad (31)$$

$$y_{ij} \; \geq \; 0 \quad (32)$$

Since Eq.(27) maps the feasible interior point $\boldsymbol{x}^*$ to $(1, 1, \cdots, 1)^T$, if $\boldsymbol{x} = \boldsymbol{x}^*$, $\boldsymbol{y} = (1, 1, \cdots, 1)^T$ is a feasible solution of Eqs.(30) $\sim$ (32), which satisfies Eq.(31). We apply the gradient projection method to Eqs.(30) $\sim$ (32). By projecting steepest descent direction $-\nabla_y E$ to the subspace defined by Eq.(31), we have

$$\Delta\boldsymbol{y}_i = -\nabla_{yi}E + (\frac{\nabla_{yi}E^T\boldsymbol{x}^*_i}{\boldsymbol{x}^{*T}_i\boldsymbol{x}^*_i})\boldsymbol{x}^*_i, \boldsymbol{i} = 1, \cdots, N \quad (33)$$

Eq.(33) is obtained by the Gram-Schmidt orthogonalization process. When $\boldsymbol{y}_i = \boldsymbol{I}$, the element $\partial E/\partial y_{ij}$ of $\nabla_{yi}E$ is

$$\frac{\partial E}{\partial y_{ij}} = \frac{1}{2}x^*_{ij}||\boldsymbol{d}_i - \boldsymbol{r}_j||^2 + \frac{A}{2}x^*_{ij}(1 - 2x^*_{ij}) \quad (34)$$

$\boldsymbol{y}_i$ is updated as:

$$\boldsymbol{y}_i = \boldsymbol{I} + \alpha\frac{\Delta\boldsymbol{y}_i}{||\Delta\boldsymbol{y}_i||} \quad (35)$$

When $0 < \alpha < 1$, $\boldsymbol{y} > 0$ and $\boldsymbol{y}$ is a feasible interior point. By applying the inverse transformation of Eq.(27) to $\boldsymbol{y}_i$ of Eq.(35) we have a new feasible solution $\boldsymbol{x}_i, i = 1, \cdots, N$ of Eqs.(23)$\sim$(25) as:

$$\boldsymbol{x}_i = \boldsymbol{x}^*_i + \alpha D^*_i\frac{\Delta\boldsymbol{y}_i}{||\Delta\boldsymbol{y}_i||} \quad (36)$$

The cluster center $\boldsymbol{r}$ is updated by the steepest descent method. The algorithm of affin scaling clustering is as follows:

Step 0 Initialize, set an initial feasible interior point.
$x_{ij} := \frac{1}{M}, \quad i = 1, \cdots, N, \quad j = 1, \cdots, M$

Step 1 Update $\boldsymbol{x}$ by Eqs. (33) and (36) .
Compute Δr and update cluster center $\boldsymbol{r}$ as:
$\boldsymbol{r} := \boldsymbol{r} + \beta\Delta r$

Step 2 If $x_{ij} > \theta_U$ then set $x_{ij} := 1$, and $x_{ik} := 0$ for all x_{ik} (k $\neq$ j) assosiated with the i-th simplex.

Step 3 If the iteration number equals a certain predetermined number, set the value A close to 0.

Step 4 If all x_{ij} are 0 or 1 then proceed to Step 5. Otherwise proceed from Step 1.

Step 5 Compute M cluster centers and terminate the iteration.

4.Numerical Experiments

Six sets of patterns are considered. Let the dimension of $\boldsymbol{d}$ equal two, the number of patterns N=10 and the number of clusters M=4. All patterns $\boldsymbol{d}$ are chosen from unit square $[0, 1] \times [0, 1]$. Initial cluster centers are chosen from random numbers with the same seed for every 100 trials. In the Hopfield net clustering, we set $B = 3.0$, $\theta = (0.01, 0.70)$. Initial values of x_{ij} are all 0.25. In the gradient projection clustering we set θ_U=0.6 by the preparatory experiments. Table1 shows the number of trials the best and the second best solutions are obtained for the first set of patterns, corresponding to the values of A. Fig.1 compares the results by the Hopfield net and the gradient projection method where the parameters A and B are chosen properly. "best" means the number of successes to obtain optimal solutions out of 100 trials. "second" means the number of obtaining the second best solutions. In the affin scaling clustering we set θ_U=0.6 by the preparatory experiments. The value of A is increased by multiplying 0.2 at the 100-th iteration. Fig.2 shows the results of the affin scaling clustering with best choice of A and the results of the K_means method [10]. If the coefficient A is chosen properly (about -0.1 in the experiment) the optimal solution is obtained with reasonable certainty. The mean iteration number and mean computational time are 2000 times and 1.5 second respectively by the Hopfield net clustering, 600 times and 0.5 second respectively by the gradient projection clustering, 100 times and 0.15 second respectively by the affin scaling clustering, and 2 times and 0.05 second respectively by the K_means clustering. The computer used is NEC9821XA.

5. Interior Point Algorithm for Fuzzy Clustering with Maximizing Entropy

Fuzzy c-means(FCM) is a generalized version of K-means hard clustering. The problem is that N patterns should be divided into M fuzzy clusters. For each pattern $\boldsymbol{d} \in D$, the grade of membership should satisfies the condition of a fuzzy partition.

$$\mu_1(\boldsymbol{d}) + \cdots + \mu_M(\boldsymbol{d}) = 1, \quad \forall \boldsymbol{d} \in D \tag{37}$$

where $\mu_1, \cdots, \mu_M$ are the non-negative membership of M non-empty fuzzy sets (clusters) $A_1, \cdots, A_M$ respectively.

Let $x_{ij} = \mu_j(\boldsymbol{d}_i)$ be the degree of membership to A_j. The formulation by Bezdek is by minimization of the objective function

$$E = \sum_i \sum_j (x_{ij})^s \parallel \boldsymbol{d}_i - \boldsymbol{r}_j \parallel^2 \tag{38}$$

E is the within-group sum-of-square-error. When $s = 1$ the problem is reduced to the traditional K-means clustering. The weighting exponent $s > 1$ controls the extent of membership sharing between fuzzy clusters. Although the FCM is a very popular data analysis technique in recent years, it was pointed out by Li and Mukaidono [13, 14, 17] that a certain lack of clarity remains in the "misterious number" s. They introduced the maximum-entropy inference method in fuzzy clustering. The problem is to find the membership x_{ij} and cluster center $\boldsymbol{r}_j$ which maximize the entropy while satisfying the condition of fuzzy partition, i.e., the memberships sum to one. They formulated an optimization problem as:

$$max \quad -\sum_{i=1}^{N}\sum_{j=1}^{M} x_{ij}\log(x_{ij}) \tag{39}$$

$$s.t. \quad \sum_{j=1}^{M} x_{ij} = 1, \tag{40}$$

$$\sum_{i=1}^{N}\sum_{j=1}^{M} x_{ij}||d_i - \boldsymbol{r}_j||^2 = L \tag{41}$$

By setting L to a certain predetermined value, to maximize the entropy subject to Eqs.(40) and (41), the Lagrangian multipliers method is skillfully used[14, 17].

We reformulate the clustering problem by Li and Mukaidono as follows:

$$min \quad E = \sum_{i=1}^{N}\sum_{j=1}^{M} x_{ij}||d_i - r_j||^2$$

$$+C\sum_{i=1}^{N}\sum_{j=1}^{M} x_{ij}\log(x_{ij}) \tag{42}$$

$$s.t. \quad \sum_{j=1}^{M} x_{ij} = 1, \quad i = 1, \cdots, N \tag{43}$$

$$x_{ij} \geq 0 \tag{44}$$

The term to have non-integer solution is ommitted here. C is a newly introduced positive coefficient to specify the tradeoff between the within-group sum-of-square-error and the entropy. C is more general and easy to specify the value than s in Eq.(38) and L in Eq.(41). The second term of Eq.(42) has the similar meaning as that of Eq.(23), since their minimization gives noninteger solutions.

By applying affin transformation $y_{ij} = x_{ij}/x^*_{ij}$ we have

$$min \quad E = \sum_{i=1}^{N}\sum_{j=1}^{M} x^*_{ij}y_{ij}||d_i - r_j||^2$$

$$+C\sum_{i=1}^{N}\sum_{j=1}^{M} x^*_{ij}y_{ij}\log(x^*_{ij}y_{ij}) \tag{45}$$

$$s.t. \quad \sum_{j=1}^{M} x^*_{ij}y_{ij} = 1, i = 1, \cdots, N \tag{46}$$

$$y_{ij} \geq 0 \tag{47}$$

The orthogonal projection of the steepest descent direction $-\nabla_y E$ to the subspace parallel with the subspace defined by Eq.(46) is obtained as:

$$\Delta\boldsymbol{y}_i = -\nabla_{yi}E + (\frac{\nabla_{yi}E^T\boldsymbol{x}^*_i}{\boldsymbol{x}^{*T}_i\boldsymbol{x}^*_i})\boldsymbol{x}^*_i, i = 1, \cdots, N \tag{48}$$

When $\boldsymbol{y}_i = \boldsymbol{I}$ the element $\partial E/\partial y_{ij}$ of $\nabla_{yi}E$ is:

$$\frac{\partial E}{\partial y_{ij}} = x^*_{ij}||\boldsymbol{d}_i - \boldsymbol{r}_j||^2 + Ax^*_{ij}log(x^*_{ij}y_{ij}) + Ax^*_{ij} \tag{49}$$

where $\boldsymbol{I}$ is a M-dimensional vector$(1, 1, \cdots, 1)^T$. Then the update rule is as:

$$\boldsymbol{y}_i = \boldsymbol{I} + \alpha\frac{\Delta\boldsymbol{y}_i}{||\Delta\boldsymbol{y}_i||} \tag{50}$$

If $0 < \alpha < 1$, $\boldsymbol{y} > 0$ and $\boldsymbol{y}$ is a feasible interior point. By applying the inverse transformation $y_{ij} = x_{ij}/x^*_{ij}$, we have a new feasible interior point $\boldsymbol{x}_i, i = 1, \cdots, N$ as:

$$\boldsymbol{x}_i = \boldsymbol{x}^*_i + \alpha D^*_i \frac{\Delta \boldsymbol{y}_i}{||\Delta \boldsymbol{y}_i||} \tag{51}$$

Cluster center $\boldsymbol{r}$ is updated by the element of $\Delta \boldsymbol{r}$, i.e.,

$$\frac{\partial E}{\partial r_{jk}} = 2 \sum_{i=1}^{N} x_{ij}(d_{ik} - r_{jk}) \tag{52}$$

$$\boldsymbol{r} := \boldsymbol{r} + \beta \Delta \boldsymbol{r} \tag{53}$$

6. Application to a data set of psychological feeling for T–shirt

Image Technology or Kansei Engineering is defined as “a translation system of a consumer’s image or feeling into real design components” [18]. If the customer has images or feeling towards a product and can represent them in an adjective-type word, the product designer is able to identify the detail design items through this procedure of Image Technology. In this section we analyse which components of T–shirt, such as colors, width of lines and intervals between lines, give ”loudly” or ”gaily” feeling. We used several kinds of T–shirt which are drawn on a computer display (shown in Fig.3). We chose three attributes, width of lines(x_1), intervals between lines(x_2) and color of lines(x_3) as characteristic components of T–shirt and set the attribute values as follows:

$\boldsymbol{x}_{1,j}$:	Width	$1.25, 2.50, 3.75, 5.00(cm)$
$\boldsymbol{x}_{2,j}$:	Interval	$1.25, 2.50, 3.75, 5.00(cm)$
$\boldsymbol{x}_{3,j}$:	Color	$Blue, Cyan, Lightblue$

where the values of width and intervals were set assuming that the length of T–shirt was 50 cm. These considerations define the 48 types of T–shirt. The quantified numerical values of feeling of loudness were obtained through the pairwise comparisons procedure. Numerical values of comspicuous feeling were obtained from 40 subjects who were all 19~20 years old female students. We have applied proposed fuzzy clustering algorithm to the data set. By some preparatory simulation study, we set the learning rate of cluster center and membership to 0.002 and 0.01, respectively. The number of clusters is eight. The clustering results about typical 4 clusters are shown in Table 2. By increasing the value of C to some proper value(C=0.2), fuzzy clusters which represents its characteristics clearly were obtained as:

Loudly, Gaily:
(Cluster 1)
  Line Width: wide
  Interval: narrow
  Color: deep
(Cluster 2)
  Line Width: wide
  Interval: wide
  Color: deep

Quiet:
(Cluster 3)
  Line Width: thin
  Interval: more or less wide
  Color: deep
(Cluster 4)
  Line Width: thin
  Interval: wide
  Color: light

7.Concluding remarks

In this paper we have shown that the better solutions of hard clustering problem can be obtained with reasonable certainty by the proposed solution algorithm for its continuous relaxation problem. Experiments reveal that the affin scaling interior point method outperforms the revised Hopfield net method and the gradient projection method. Although the additional term for controlling non-integer solution is effective, since the Hopfiel net algorithm is an exterior point method, its computation time is sufficiently large and the clustering results are not satisfactory. The revised algorithm of gradient projection method search for the solution in the feasible region. Since the elements of solution vector tend to quickly converge to the boundary of feasible region ($x_{ij} = 0$) , the clustering results are unsatisfactory. Therefore we have applied affin scaling interior point method to confine the solution in a feasible region and obtained better cluster centers. By this novel algorithm better solutions can be obtained with relatively short computation time.

The proposed algorithm can be easily applied to a generalized version of the fuzzy clustering formulation with maximizing entropy. Caracteristic features of T-shirt design were clarified by the fuzzy clustering.

References

[1] Abe, S., Determination of Weights of the Hopfield Neural Networks and Its Evaluation, Transactions of Information Processing Society of Japan, Vol. 34, No. 1, 1993, pp. 21-28 (in Japanese).

[2] Iyer, S.V.B., Niranjan, M. and Fallside, F., A Theoretical Investigation into the Performance

of the Hopfield model, IEEE Trans. Neural Networks, 1, 1990, pp. 204-215.

[3] Bezdek, J.C., Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum, NewYork, 1981.

[4] Dikin, I.I., Iterative Solution of Problems of Linear and Quadratic Programming, Soviet Mathematics Doklady, Vol.8, 1967, pp. 674-675 .

[5] Duda, R.O. and Hart, P.E., Pattern Classification and Scene Analysis, Wiley, New York, 1973.

[6] Forgy, E.W., Cluster Analysis of multivariate data, efficiency vs. interpretability of classifications, Biometrics, Vol.21, 1965, pp. 768-769 (Abstract).

[7] Hopfield, J.J., Neurons with Graded Response Have Collective Computational Properties Like Those of Two-state Neurons, Proceedings of National Academy of Science of the United States of America, Vol. 81, 1984, pp. 3088-3092.

[8] Hopfield, J.J. and Tank, D.W., Neural Computation of Decisions in Optimization Problems, Biol. Cybern., Vol. 52, 1987, pp. 141-152 .

[9] Jacoby, S.L.S., Kowalik, J.S., Pizzo, J.T., Iterative Methods for Nonlinear Optimization Problems, Prentice-Hall, New Jersey, 1972.

[10] Johnson, D., More Approaches to the Traveling Salesman Guide, Nature, Vol. 330, No. 10, 1987, pp. 525-525.

[11] Kita, H., The Hopfield Neural Network and Simulated Annealing, Journal of Japan Society for Artificial Intelligence, Vol.7, No.6, 1992, pp. 970-979 (in Japanese).

[12] Kita, H., Odani, H., Nishikawa, Y., Optimization by Means of a Neural Network Model - An Application to the Placement Problem, Transactions of the Institute of Systems, Control and Information Engineers, Vol. 4, No. 1, 1991, pp. 48-56 (in Japanese).

[13] Li, R.P. and Mukaidono, M., Fuzzy Modelling and Clustering Neural Network, Journal of Control and Cybernetics, Vol.25, No.2, 1996, pp.225-242.

[14] Li, R.P. and Mukaidono, M., A Maximum Entropy Approach to Fuzzy Clustering, Proc. of the 4th IEEE Intern. Conf. on Fuzzy Systems (FUZZ-IEEE/IFFES'95), Yokohama, Japan, March 20-24, 1995, pp. 2227-2232.

[15] Matsuda, S., The stability of the Solution in Hopfield Neural Network, Transactions of the Institute of Electronics, Information and Communication Engineers, Vol.J77-D-II, No. 7, 1994, pp. 1366-1374 (in Japanese).

[16] McQueen, J.B., Some Methods of Classification and Analysis of Multivariate Observations, Proc. of 5th Berkeley Symposium on Math. Stat. and Prob., 1967, pp. 281-297 .

[17] Mukaidono, M. and Li, R.P., Gaussian Clustering Method Based on Maximum-fuzzy-entropy interpretation, Fuzzy Sets and Systems(to appear).

[18] Nagamacni, M., An Image Technology Expert System and Its Application to Design Consultation, International Journal of Human–Computer Interaction, Vol.3, No.3, 1991, pp.267-279.

[19] Rosen, J.B., The Gradient Projection Methed for Non-linear Programming, Part I, Linear Constraints, SIAM Journal., Vol. 8, 1960, pp. 181-217 .

[20] Storolz, P., Merging Constrained Optimization with Deterministic Annealing to Solve Combinatorically Hard Problem Adv. in Neural Information Proc. Systems, Vol.4, 1992, pp. 1025-1032.

[21] Uesaka, Y., Mathematical Aspects of Neurodynamics for Combinatorial Optimization, Transactions of the Institute of Electronics, Information and Communication Engineers, Vol. E74, No. 6, 1991, pp. 1368-1372.

[22] Urahama, K., Nishiyuki, H., Neural Algorithms for Placement Problems, Proceedings of 1993 International Joint Conference on Neural Networks, 1993, pp. 2421-2424.

[23] Urahama, K., Mathematical Programming Formulation for Neural Combinatorial Optimization Algorithms, Transactions of the Institute of Electronics, Information and Communication Engineers, Vol.J78-A, No. 1, 1995, pp. 47-53 (in Japanese).

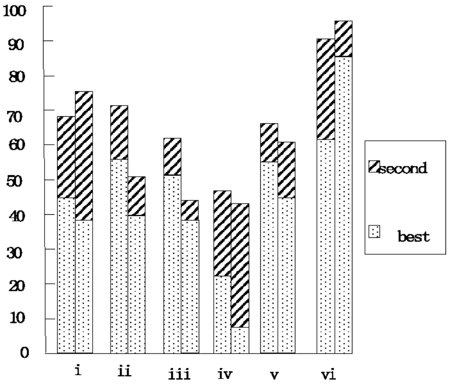

Figure 1: Comparisons of clustering result (Left:Hopfield Neural Net, Right : Gradient Projection Method)

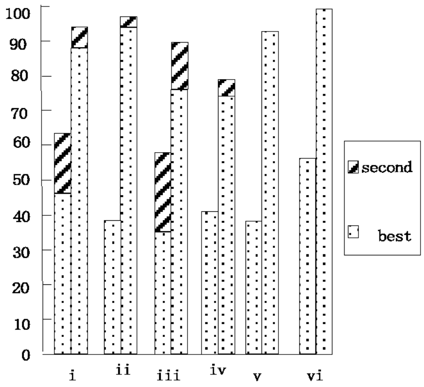

Figure 2: Comparisons of clustering result (Left : K_means, Right : Affin Scaling Interior Point Method)

Table 1 Number of trials optimal solutions are obtained

A	best	second
0.10	1	9
0.05	1	16
0.00	2	25
-0.05	27	41
-0.10	77	15
−0.11	**88**	**7**
-0.15	87	7
-0.20	23	74

Table 2 Results of fuzzy clustering for T-shirt data

(a) Cluster number (b) Data number
(c) Membership
(d) The largest membership cluster
(e) Width (f) Interval (g) Color (h) Loudness

(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
1	27	0.458	1	0.666	0.333	1	0.739
1	25	0.340	1	0.666	0	0.5	0.784
1	42	0.316	2	1	0.666	1	0.773
1	12	0.311	7	0.333	0	1	0.773
1	37	0.309	6	1	0	0.5	0.500
1	40	0.307	1	1	0.333	0.5	0.659
1	28	0.249	1	0.666	0.333	0.5	0.648
2	45	0.858	2	1	1	1	0.557
2	33	0.803	2	0.666	1	1	0.659
2	42	0.560	2	1	0.666	1	0.773
2	21	0.507	2	0.333	1	1	0.545
2	30	0.490	2	0.666	0.666	1	0.773
2	46	0.427	8	1	1	0.5	0.568
2	34	0.308	8	0.666	1	0.5	0.330
2	18	0.306	2	0.333	0.666	1	0.545
3	9	0.924	3	0	1	1	0.023
3	6	0.918	3	0	0.666	1	0.000
3	3	0.710	3	0	0.333	1	0.216
3	8	0.480	3	0	0.666	0.5	0.080
3	11	0.476	3	0	1	0.5	0.091
3	21	0.343	2	0.333	1	1	0.545
3	5	0.343	5	0	0.333	0.5	0.102
4	10	0.860	4	0	1	0	0.159
4	7	0.694	4	0	0.666	0	0.080
4	23	0.680	4	0.333	1	0	0.261
4	20	0.498	4	0.333	0.666	0	0.284
4	11	0.461	3	0	1	0.5	0.091
4	22	0.369	4	0.333	1	0.5	0.364
4	8	0.358	3	0	0.666	0.5	0.080
4	4	0.310	5	0	0.333	0	0.170
4	19	0.250	4	0.333	0.666	0.5	0.409