A modified genetic algorithm for feature clustering

Miki AOYAGI and Kumiko TSUJI

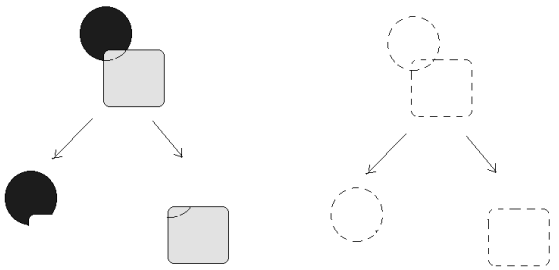
Abstract

A new genetic algorithm (GA) based on feature clustering with an energy function is proposed for obtaining optimal segmentation. In the proposed algorithm, which we call the modified genetic algorithm (MGA), the length of each genome is the number of features and each individual (genome) represents one assignment of the input-features to labels. The energy function, which is used as a fitness function for our GA, consists of local competitive and cooperative interactions among features and a global inhibition. We construct the function in order to control the local competitive and cooperative interactions, which fits to various features.

1 Introduction

In this paper, a new genetic algorithm (GA) based on feature clustering with an energy function is proposed for obtaining optimal segmentation. In general, for clustering method, each feature is labeled so that features belonging to the same cluster have the same label and features belonging to different clusters have different labels. Usually, each feature is denoted by a parameter vector. A typical vector example consists of a pixel's position and its gray (or color) level in the image. Another vector example consists of the pixel's position and its local edge orientation. The problem for the local edge orientation is called contour grouping. Figure 1(a) shows image segmentation using information of gray levels and (b) shows that using information of local edge orientations.

The GA has been applied successfully to many problems of searching for an approximate global minimum of the objective function. Therefore, it is



(a) (b)

Figure 1: (a) Image segmentation using features whose vectors consist of their positions and gray levels; regions with homogeneous gray levels have the same label and are combined. (b) Image segmentation using features whose vectors consist of their positions and local edge orientations (contour grouping problem); features that lie along a curve are combined.

natural to consider how to apply the GA to the feature clustering problem. Although there are some applications to the image segmentation, there is a few applications to the feature clustering problem[1]. The method in the paper [1] is to solve the problem of partitioning a feature group into homogeneous ones based on some measure (e.g. Figure 1(a)). However, the method is not suitable for application to contour grouping (Figure 1(b)).

2 MGA ALGORITHM

In this section, we introduce our MGA algorithm.

Energy function (Fitness function)

Consider the problem of clustering R features which are denoted by parameter vectors g_r, $1 \leq r \leq R$.

Let L be the number of labels. Consider L output layers with R (the number of features) variables in each layer. Denote the rth variable in the lth output-layer by v_{rl} which takes a value of 0 or 1. The feature g_r having the lth lavel is expressed by the activities $v_{rl} = 1$ and $v_{rl'} = 0$ for $l' \neq l$. Therefore, the feature clustering is achieved by determining v_{rl} so that each feature g_r is properly assigned to one of the output layers.

We denote a local interaction between features g_r and $g_{r'}$ by a scalar value $f_{g_r g_{r'}}$, which is a symmetric function of the feature parameters; $f_{g_r g_{r'}} = f_{g_{r'} g_r}$.

A positive value $f_{g_r g_{r'}} > 0$ means that g_r and $g_{r'}$ are very likely to belong to the same cluster, and a negative value $f_{g_r g_{r'}} < 0$ means that g_r and $g_{r'}$ are likely to belong to different clusters.

In order to measure the fitness of assignment of the input features to the output layers, let us consider the following energy function;

$$E = -\frac{1}{2}\sum_{l=1}^{L}\sum_{r=1}^{R}\sum_{r'=1,r'\neq r}^{R} f_{g_r g_{r'}} v_{rl} v_{r'l} + \frac{k}{2}\sum_{l=1}^{L}\sum_{r=1}^{R}\sum_{r'=1,r'\neq r}^{R} v_{rl} v_{r'l}, \qquad (1)$$

where $\sum_l v_{rl} = 1$, $v_{rl} \in \{0, 1\}$ and $k \geq 0$ is a control parameter that adjusts the strength of a global inhibition. This energy function is obtained by using the local interactions $f_{g_r g_{r'}}$ and global inhibition. It is similar to those used in a Potts spin model [3] and in the competitive-layer model (CLM) [4].

The corresponding CLM energy function with the combination of the local and winner-take-all interactions and the global inhibition is

$$E_{CLM} = \frac{J}{2}\sum_{r=1}^{R}(h_r - \sum_{l=1}^{L} v_{rl})^2 + E + \frac{T}{2}\sum_{l=1}^{L}\sum_{r=1}^{R} v_{rl}^2, \qquad (2)$$

where $v_{rl} \geq 0$ is a real number. The term $J\sum_r(h_r - \sum_l v_{rl})^2$ corresponds to the winner-take-all interactions with strength J between v_{rl} and $v_{rl'}$ of the feature g_r. h_r is the significance of the detection of the feature g_r. In the MGA, this term disappears since we put $\sum_l v_{rl} = 1$. The parameter T in the last term is used to slow the local self-inhibition in order to increase grouping quality. The CLM method used gradual lowering of the inhibitory T. However, this annealing method on T does not work properly for some complex input patterns, as was shown in [2]. Since the MGA does not use the annealing method, this term $\frac{T}{2}\sum_{l,r} v_{rl}^2$ also disappears.

We explain how to initially produce each individual with a genome for our GA.

Define an iteration for a step function v_{rl} by

$$v_{rl}(\text{new}) = \begin{cases} 1 & \text{if } l = \min\{l_1 : w_{rl_1} \geq w_{rl_2} \forall l_2\} \\ 0 & \text{for the others,} \end{cases} \qquad (3)$$

where,

$$w_{rl} = \sum_{r'=1,r'\neq r}^{R} f_{g_r g_{r'}} v_{r'l}(\text{old}). \qquad (4)$$

To determine initial values $v_{rl_r}(initial)$, one of output layers l_r, $1 \leq l_r \leq L$, is chosen randomly for each feature g_r. Set $v_{rl}(initial)$ as

$$v_{rl_r}(initial) = 1, \quad v_{rl}(initial) = 0 \text{ if } l \neq l_r. \tag{5}$$

In each iteration, we modify only L variables v_{rl}, $l = 1, \cdots, L$, for the feature g_r selected randomly.

Let

$$E_{local} = -\frac{1}{2} \sum_{l=1}^{L} \sum_{r=1}^{R} \sum_{r'=1, r' \neq r}^{R} f_{g_r g_{r'}} v_{rl} v_{r'l}, \tag{6}$$

which is the first term of the energy function E (1).

Theorem *By the iteration (3) and (4), we have* E_{local}(new) $\leq$ E_{local}(old). *That is,* E_{local} *makes a decreasing sequence.*

Proof Fix r. Suppose

$$\sum_{r'=1, r' \neq r}^{R} f_{g_r g_{r'}} v_{r'm_1} = \cdots = \sum_{r'=1, r' \neq r}^{R} f_{g_r g_{r'}} v_{r'm_k}$$
$$> \sum_{r'=1, r' \neq r}^{R} f_{g_r g_{r'}} v_{r'm'}, \quad m' \neq m_1, m_2, \cdots, m_k,$$

where $1 \leq m_1 < m_2 < \cdots < m_k \leq L$, then we have v_{rm_1}(new) $= 1$, v_{rm}(new) $= 0$ where $m \neq m_1$.

Now by using

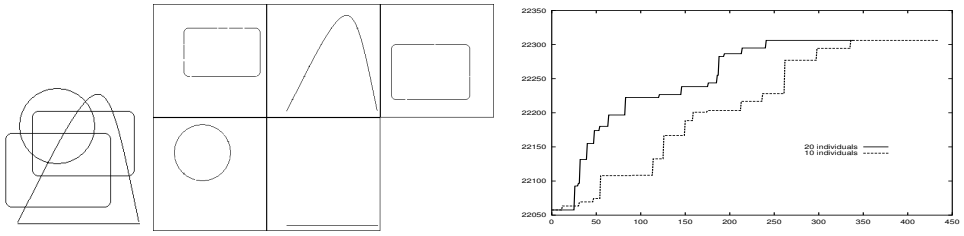
$$E_{local} = -\frac{1}{2} \sum_{l=1}^{L} \sum_{r'=1, r' \neq r}^{R} \sum_{r"=1, r" \neq r, r'}^{R} f_{g_{r'} g_{r"}} v_{r'l} v_{r"l} - \sum_{l=1}^{L} \sum_{r'=1, r' \neq r}^{R} f_{g_r g_{r'}} v_{rl} v_{r'l},$$

and

$$\sum_{l=1}^{L} \sum_{r'=1, r' \neq r}^{R} f_{g_r g_{r'}} v_{rl}(\text{new}) v_{r'l} = \sum_{r'=1, r' \neq r}^{R} f_{g_r g_{r'}} v_{r'm_1}$$
$$\geq \sum_{l=1}^{L} v_{rl}(\text{old}) \sum_{r'=1, r' \neq r}^{R} f_{g_r g_{r'}} v_{r'l},$$

we obtain E_{local}(new) $\leq$ E_{local}(old). Q.E.D

The second step : Applying GA



(a) (b) (c)

Figure 2: Grouping of image data. (a) Input image. (b) The result of the output layers at the 340th generation. 10 output layers and 20 individuals are prepared. Processing time is 4 minutes on Pentium III. (c) Graphs of the fitness value $-E$ for the elite individual in each generation with 10 and 20 individuals in a population. The convergence probability is lower for the smaller population size.

The purpose of the second step is to reach the globally minimal value of the energy function E (1) by using GA. In other words, our aim is to obtain the state of image processing, where split contours generated by the first step are gathered and combined into proper groups. In this step, the global inhibition is taken into account.

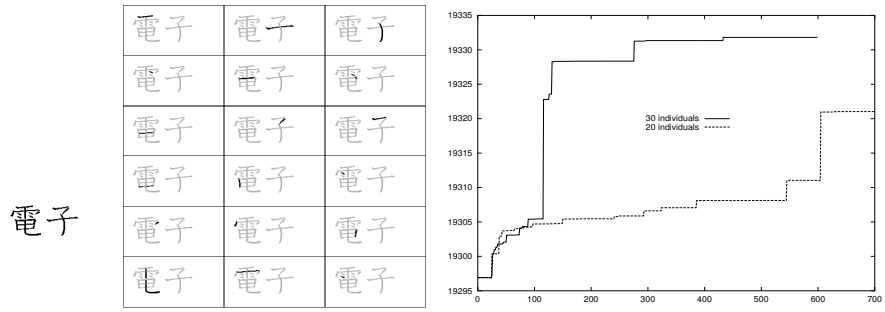
A genome of each individual contains a set of R parameters $(l_1, l_2, \cdots, l_R)$. Each parameter l_r implies the output layer's number of the feature g_r. An individual represents one assignment of features to the output layers. The fitness function of GA is the energy function defined by (1), which is rewritten as follows.

$$E = -\frac{1}{2}\sum_{r=1}^{R}\sum_{r'=1, r'\neq r, l_{r'}=l_r}^{R} f_{g_r g_{r'}} + k\sum_{l=1}^{L}\sum_{r=1, l_r=l}^{R} v_{rl_r}.$$

We propose four kinds of special mutations. The mutations and the algorithm were listed in [2]. Moreover, comprisons of MGA, the standard GA and CLM were shown in [2].

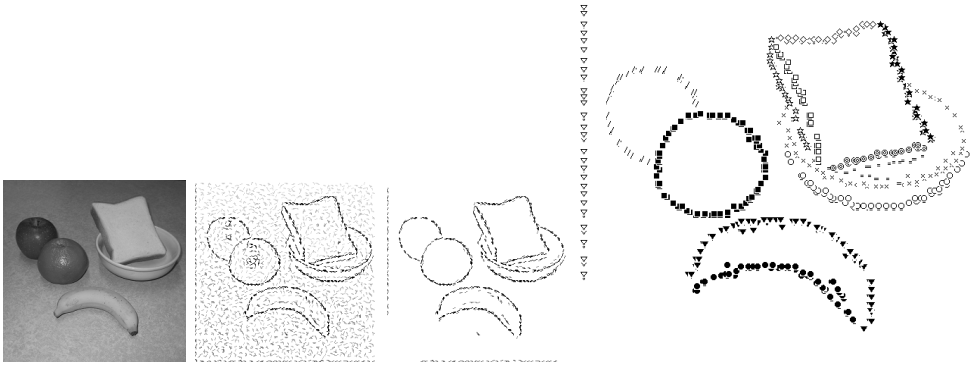
Numerical method to obtain a(r)**, a local interaction** $f_{g_r g_{r'}}$ **and a coefficient of global inhibition** k

The pairwise local interaction $f_{g_r g_{r'}}$ between edge features is depicted in Figure 5. This interaction field is similar to the cocircular interaction employed in [4]. In [4], Wersing, Steil, & Ritter used only coarse images, i.e., they assumed that only coarsely sampled features are active in numerical experiments. Therefore, their interaction field is relatively wide and is not



(a) (b) (c)

Figure 3: Grouping of image data. (a) Input image. (b) The result of the output layers at the 577th generation. 30 output layers and 30 individuals are prepared. Processing time is 9 minutes. The gray curves show the original image. (c) Graphs of the fitness value $-E$ for the elite individual in each generation with 20 and 30 individuals in a population.



(a) (b) (c) (d)

Figure 4: Grouping of image data. (a) Input image. (b) Edge features. (c) Image after noise removal with $EPS = 3$. (d) The result of the output layers at the 601st generation. 30 output layers and 40 individuals are prepared. Different symbols denote activity in different output layers. Processing time is 10 minutes.

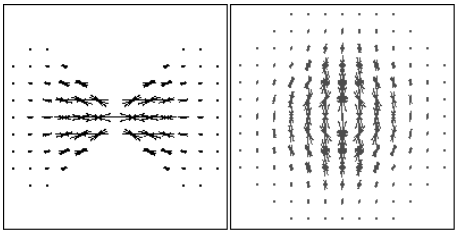


Figure 5: Local interaction for contour grouping; edges $g_{r'}$ in the left picture have $f_{g_r g_{r'}} > 0$ and ones in the right picture have $f_{g_r g_{r'}} < 0$, where g_r is the horizontal edge at the center. Length of each edge codes interaction strength.

suitable for our cases.

In all our examples, each edge orientation of a feature is generated by applying the least square method to 5×5 pixels surrounding the position of the feature. Local interactions $f_{g_r g_{r'}}$ are defined as follows.

Denote the position parameter of g_r by $n(r)$.

Let $d = n(r) - n(r')$ be the difference vector for two positions of $g_r, g_{r'}$, and a_d the angle of d.

Let $0 \leq \mathrm{da}_1 < \pi$ be

$$\mathrm{da}_1 = \mathrm{a}(r') - \mathrm{a}(r) \bmod \pi.$$

For $c = \mathrm{a}_d - \mathrm{a}(r) \bmod \pi$ $(0 \leq c < \pi)$, put

$$\mathrm{da}_2 = \begin{cases} \mathrm{a}(r') - \mathrm{a}(r) - c \bmod \pi & \text{if } c < \pi/4 \text{ or } c > 3\pi/4, \\ \mathrm{a}(r') - \mathrm{a}(r) - (\pi/2 - c) \bmod \pi & \text{if } \pi/4 \leq c \leq \pi/2, \\ \mathrm{a}(r') - \mathrm{a}(r) - (3\pi/2 - c) \bmod \pi & \text{if } \pi/2 \leq c \leq 3\pi/4, \end{cases}$$

where $0 \leq \mathrm{da}_2 < \pi$.

Let P be a constant which controls the spatial range. Define their local interaction by

$$\begin{aligned} & f_{g_r g_{r'}} \\ = & \begin{cases} \{b(\mathrm{da}_1) + b(\mathrm{da}_2)\} \exp(-|d|^2/P) & \text{if } \mathrm{da}_1 < \pi/4 \text{ or } \mathrm{da}_1 > 3\pi/4, \\ -0.5 & \text{if } \pi/4 \leq \mathrm{da}_1 \leq 3\pi/4, \end{cases} \end{aligned} \tag{7}$$

where

$$b(x) = \begin{cases} -\frac{4}{\pi}x + 1 & 0 \leq x < \pi/2, \\ \frac{4}{\pi}x - 3 & \pi/2 \leq x \leq \pi. \end{cases}$$

$b(x)$ is a piecewise linear function of x with a corner at $\pi/2$, which attains its minimum -1 at $\pi/2$ and its maximum 1 at 0 and π.

If two features g_r and $g_{r'}$ have the same orientation, then $b(\mathrm{da}_1)$ is equal to 1 and contributes to increasing its local interaction. If g_r lies at right angles to $g_{r'}$, then $b(\mathrm{da}_1)$ is equal to -1 and contributes to decreasing its local interaction.

da_2 is adopted to admit excitatory interaction if two features can be connected by a curve with a small curvature. The parameters da_1 and c contribute to increasing or decreasing its local interaction.

$P = 25$ in (7) is used for Figures 2 and 3. $P = 100$ is used for Figure 4. The coefficient of the global inhibition is computed by letting $k = 0.3/R$ in (1), which controlls the spatial range.

References

[1] P. Andrey & P. Tarroux, "Unsupervised segmentation of Markov random field modeled textured images using selectionist relaxation," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, no. 3, pp. 252–262, 1998.

[2] M. Aoyagi and K. Tsuji, "A modified genetic algorithm for image segmentation based on feature clustering," to appear in International Symposium on Information Theory and its Applications, (Parma, Italy), 2004.

[3] C. Peterson and B. Soderberg, "A new method for mapping optimization problems onto neural networks," *Int. J. Neural Systems,* vol. 1, no. 1, pp. 3-22, 1989.

[4] H. Wersing, J. J. Steil and H. Ritter, "A competitive-layer model for feature binding and sensory segmentation," *Neural Computation,* vol. 13, pp. 357-387, 2001.

Miki AOYAGI
Department of Mathematics
Sophia University
7-1 Kioi-cho, Chiyoda-ku,
Tokyo, 102-8554 JAPAN

Kumiko TSUJI
Fukuoka Junior College
Teikyo University
4-3-124 Shin-Katsutachimachi
Ohmuta 836 JAPAN