International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:5, No:8, 2011

Initialization method of reference vectors for improvement of recognition accuracy in LVQ

Yuji Mizuno and Hiroshi Mabuchi

Abstract—Initial values of reference vectors have significant influence on recognition accuracy in LVQ. There are several existing techniques, such as SOM and k-means, for setting initial values of reference vectors, each of which has provided some positive results. However, those results are not sufficient for the improvement of recognition accuracy. This study proposes an ACO-used method for initializing reference vectors with an aim to achieve recognition accuracy higher than those obtained through conventional methods. Moreover, we will demonstrate the effectiveness of the proposed method by applying it to the wine data and English vowel data and comparing its results with those of conventional methods.

Keywords—Clustering, LVQ, ACO, SOM, k-means.

I. Introduction

LVQ is one of learning algorithms in neural networks. Its learning algorithm is simple, yet it achieves high recognition accuracy against complicated data. The algorithm of LVQ first sets initial values of each class called a reference vector and, by updating these reference vectors, forms an ideal cluster. Thus, recognition accuracy in LVQ will greatly be affected by initial values of reference vectors. Existing studies used to set initial values of reference vectors include: a method by which randomly selected learning data is appointed as initial values of reference vectors [1][2]; an SOM-used method for initializing reference vectors [3]; and a k-means-used method for initializing reference vectors [4]. Each of these methods, then, has achieved some positive results. However, those results are not sufficient for the improvement of recognition accuracy.

It is known that reference vectors that greatly affect recognition accuracy are located near the decision boundary. To improve recognition accuracy, it can be assumed, therefore, that setting initial values of reference vectors near the decision boundary and learning them is essential [3].

Conventional methods, such as SOM and k-means, however, can not classify data derived from near the decision boundary.

To initialize reference vectors, the Ant Colony Optimization algorithm (ACO), a technique to solve combinatorial optimization problems, is used in this study. Generally in the case of the Quadratic Assignment Problem (QAP), the evaluation function of ACO uses the distance between two arbitrary locations and the sum of the cost of facilities. Evaluation values become larger as these values in distance and cost become smaller. In this study, learning data is made to correspond to locations and reference vectors to facilities, and the partial evaluation function is a product of the sums of the Euclidean distances between an arbitrary location and another arbitrary location in a different class and between an arbitrary facility and another arbitrary facility in a different class. Furthermore, we assume that if the Euclidean distance between one arbitrary set of data and another arbitrary set of data in a different class is minimal, these two sets of data are derived from near the decision boundary. If this is the case, the inverse of an exponential function with the Euclidean distance as the base is added to the partial evaluation function. Defining the function obtained as a result of the aforementioned procedure as an evaluation function and applying it, this paper proposes a method for initializing reference vectors, which are located near the decision boundary and thus have significant influence on recognition accuracy.

This paper then demonstrates the effectiveness of the proposed method by applying it to the wine data and English vowel data obtained from the UCI Machine Learning Repository [18] and comparing its results with those of conventional methods.

Yuji Mizuno is Rhythm-based Brain Computation Unit RIKEN BSI-TOYOTA Collaboration Center RIKEN Brain Science Institute, 2-1, Hirosawa, Wako, Saitama 351-0198 Japan (e-mail: ymizuno@brain.riken.jp).

Hiroshi Mabuchi is with the Iwate Prefectural University Faculty of Software and Information Science, 152-52, Sugo, Takizawa, Iwate, 020-0193, Japan (e-mail: mabu@iwate-pu.ac.jp).

II. LVQ

Currently defined LVQ models are LVQ1, LVQ2.1, LVQ3 and so on [3][5]. The following sections provide explanations of each model.

A. LVQ1

Let the reference vector nearest to the input vector x be w_c. When w_c and x belong to the same class,

$$w_c(t+1) = w_c(t) + \alpha(x(t) - w_c(t)) \tag{1}$$

where w_c is brought closer to x. When w_c and x belong to different classes,

$$w_c(t+1) = w_c(t) - \alpha(x(t) - w_c(t)) \tag{2}$$

where w_c is further removed from x. Here, t is time and α is the learning coefficient $0 < \alpha < 1$.

B. LVQ2.1

Let the two reference vectors nearest to the input vector x be w_i and w_j, and let both w_i and w_j be updated simultaneously. Here, one of the two reference vectors must belong to the same class as x and the other to a different class.

Then, let the Euclidean distance between x and w_i or w_j be d_i and d_j respectively, and update the reference vectors by equations (4) and (5) only if equation (3) holds.

$$min(\frac{d_i}{d_j}, \frac{d_j}{d_i}) > s, \quad s = \frac{1-\omega}{1+\omega} \tag{3}$$

$$w_i(t+1) = w_i(t) + \alpha(x(t) - w_i(t)) \tag{4}$$

$$w_j(t+1) = w_j(t) - \alpha(x(t) - w_j(t)) \tag{5}$$

C. LVQ3

LVQ2.1 does not take post-learning values of reference vectors into account and therefore will suffer a pattern misclassification problem as learning continues. LVQ3 was developed to solve this problem. When w_i and w_j, two reference vectors nearest to x, belong to the same class as x, LVQ3 updates these vectors by equations (6) and (7).

$$w_i(t+1) = w_i(t) + \epsilon\alpha(x(t) - w_i(t)) \tag{6}$$

$$w_j(t+1) = wj(t) + \epsilon\alpha(x(t) - w_j(t)) \tag{7}$$

Here, ϵ is the smallest positive real number.

Of the aforementioned models, LVQ2.1 is used in this study.

D. Conventional methods for initializing reference vectors

Initial values of reference vectors greatly affect recognition accuracy. For this reason, there have been numerous studies with regard to how to initialize reference vectors. Prevailing studies include a technique which randomly sets initial values of reference vectors [1][2], an SOM-used method for initializing reference vectors [3] and a k-means-used method for initializing reference vectors [4]. With an SOM-used method for setting reference vectors, if learning is conducted without consideration to data classes, there is a possibility that reference vectors may belong to a different class as neighborhood learning data [6]. With the SOM learning algorithm for each class, it is also possible that the distribution of initial values of reference vectors will not fit into the entire learning data [7]. A k-means-used method is also a learning algorithm without a teacher and thus possesses similar problems as SOM. To obtain higher recognition accuracy, classification of data from near the decision boundary is important as well. However, because the three methods described above do not correspond to such data, the data can not be classified correctly. To solve these problems, we propose a method for initializing reference vectors using ACO.

III. ACO

ACO is an optimization technique employing foraging behavior patterns of ants and is a metaheuristic for optimization problems. The ACO algorithms include Ant System (AS) [8][9], Rank-based Ant System [10], MAX-MIN Ant System (MMAS)[11] and Ant Colony System (ACS) [12]. In this chapter, the Ant System, the original algorithm for all ACO algorithms, and AS-QAP [13][14], an application of AS to the Quadratic Assignment Problem (QAP), used in this study are discussed.

A. Ant System

AS is an algorithm to solve the Traveling Salesman Problem (TSP). Given several cities and their pairwise distances, TSP is a combinatorial optimization problem in which the shortest possible route from one city to all the others, before returning to the departure city, is obtained.

In AS, an artificial ant (an agent) randomly chooses the next destination among unvisited cities based on pairwise distances and pheromone information and creates a tour route by visiting each city exactly once. The probability $p_{i.j}^k$ for the agent k traveling from the cities i to j is expressed by equation (8).

$$p_{i,j}^k = \frac{[\tau_{i,j}]^\alpha [\eta_{i,j}]^\beta}{\sum_{l \in N^k} [\tau_{i,l}]^\alpha [\eta_{i,l}]^\beta} \tag{8}$$

Here, $\tau_{i,j}$ is the amount of pheromones (pheromone value) laid between the cities i and j, N^k a set of unvisited nodes of the agent k and $\eta_{i,j}$ the heuristic information. The inverse distance $d_{i,j}$ between the cities i and j as shown below is commonly used in TSP.

$$\eta_{i,j} = \frac{1}{d_{i,j}}$$

α and β are positive real numbers and serve as important parameters for pheromone information and heuristic information respectively.

Once each agent successfully created a tour route, the pheromone value between cities is updated by equations (9) and (10).

$$\tau_{i,j} \leftarrow (1-\rho)\tau_{i,j} + \sum_{k=1}^{m} \Delta\tau_{i,j}^k \tag{9}$$

$$\Delta\tau_{i,j}^k = \begin{cases} \frac{1}{f^k} & \text{if } {}_{i,j} \in \ T^k \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

T^k denotes a set of paths along the tour routes which were created by the agents. Also, ρ denotes the pheromone evaporation rate. This indicates that the pheromone value is higher when the total distance of the tour route is lower.

By repeating computations described above until the end condition is satisfied, a solution is then obtained.

B. AS-QAP

In this study, we use the QAP-AS algorithm, an application of AS to solve the QAP, for setting reference vectors. This section, then, provides explanations for ACO for solving the QAP [13]-[16].

The QAP is a problem to assign n facilities to n locations and find a permutation of the minimum value computed by equation (11).

$$f(\phi) = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{D}_{ij}\mathrm{F}_{\phi(i)\phi(j)} \tag{11}$$

Here, D is the pairwise distances between locations and F the costs between facilities in an $n \times n$ matrix. ϕ is a permutation of $1, 2, \ldots, n$. The QAP is a generalized optimal location

problem which assigns all facilities to different locations with the goal of minimizing both the distances between locations and the sum of the costs of facilities.

In AS-QAP, suppose that agents (artificial ants) deposit pheromones between locations and facilities. Tracing pheromones, an agent chooses and assigns a facility among unassigned facilities until all facilities are assigned to each location and creates permutations. The probability $p_{i,j}^k$ for the agent k choosing an unassigned facility j from the location i is expressed by equation (12).

$$p_{i,j}^k \frac{[\tau_{i,j}]}{\sum_{l \in N^k} [\tau_{i,l}]} \tag{12}$$

Because equation (12) does not employ the heuristic value, it is equivalent to equation (8) with $\eta(i,j) = 1$.

Once each agent successfully created a permutation, the pheromone value between locations and facilities is updated by equations (13) and (14).

$$\tau_{i,j} \leftarrow (1-\rho)\tau_{i,j} + \sum_{k=1}^{m} \Delta\tau_{i,j}^k \tag{13}$$

$$\Delta\tau^k(i,j) = \begin{cases} \frac{1}{f(\phi^k)} & \text{if } \phi_i^k = j \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

By repeating computations described above until the end condition is satisfied, a solution is then obtained.

Using this method, this study attempts to obtain higher evaluation values by assigning a reference vector (facility) to certain learning data (location).

IV. ACO-used method for initializing reference vectors

In the context of recognition problems, it is important that misclassification of data from near the decision boundary is kept minimal for the improvement of recognition accuracy. By setting initial values of reference vectors near the decision boundary and learning them, it can be expected that misclassification of data from near the boundary is reduced in LVQ. In this study, we will use the ACO algorithm to set initial values of reference vectors near the decision boundary. In the proposed method, it is assumed that if the Euclidean distance between one arbitrary set of data and another arbitrary set of data in a different class is minimal, these two sets of data are derived from near the decision boundary and are elements of the partial evaluation function.

First, we have allocated locations and facilities in the QAP to learning data and reference vectors respectively and written an equation (15) for the QAP.

$$f(\phi) = \mathrm{D}_i\mathrm{F}_{\phi(j)} \tag{15}$$

Here, D is the sum of the Euclidean distance between the location i and an arbitrary location in a different class while F is the sum of the Euclidean distance between the facility j and an arbitrary facility in a different class. Using the method discussed in (III-B), this study assigns a reference vector (facility) to learning data (location) and sets an initial value of a reference vector that represents the facility with the highest pheromone value. The number of reference vectors is fixed when determining reference vector initial values.

In this study, pheromones are updated by equations (16) and (17). For updating pheromones, the inverse $\frac{1}{d^n}$ of an exponential function with the sum of the minimum Euclidean distance between a set of data from the conventional update function (partial evaluation function) (13) and another set of data in a different class as the base is added to the pheromone update function. d is expressed in equation (18).

$$\tau_{i,j} \leftarrow (1-\rho)\tau_{i,j} + \sum_{k=1}^{\mathrm{m}} \Delta\tau_{i,j}^k + \frac{1}{d^n} \tag{16}$$

$$\Delta\tau_{i,j}^k = \begin{cases} \frac{1}{f(\phi^k)} & \text{if } \phi_i^k = j \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

$$d = \sum_{c1=1}^{\mathrm{n}} \sum_{c2=1}^{\mathrm{n}} min \left(\sqrt{\sum_{i=1}^{dim} (x_{c1i} - x_{c2i})^2} \right) \quad c1 \neq c2 \tag{18}$$

In equation (18), x_{c1} denotes the class $c1$ data, x_{c2} the class $c2$ data, n the number of classes and dim the number of dimensions of data. In this study, $\frac{1}{d^n}$ is added to the evaluation function of ACO; as the n value becomes larger, the distance decreases if its value is smaller than 1 but increases if its value is 1 or larger. It can be said, therefore, that the distance d has significant influence. To determine the optimal n value, we used 1 to 10 as values for n and examined the change in recognition accuracy. Figures 1 and 2 show recognition accuracy for wine data and vowel data respectively when the range of numbers 1 to 10 is used in the place of n.

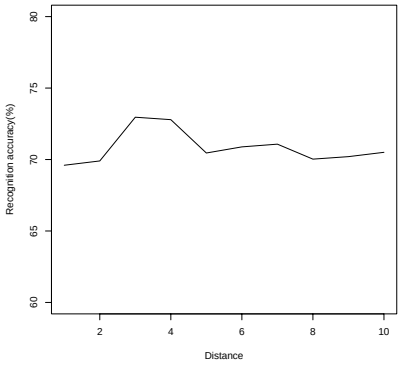

Fig. 1. Changes in the recognition rate on wine data by difference in n

It is noted from Figures 1 and 2 that recognition accuracy reaches the highest level at $n = 3$ but does not improve as n becomes larger with little changes shown at $n = 4$ or larger. For this reason, this study uses the value of $n = 3$, at which the highest recognition accuracy for both wine data and vowel

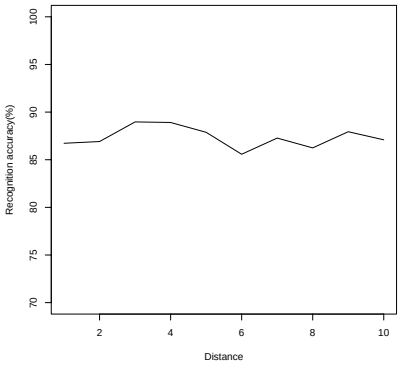

Fig. 2. Changes in the recognition rate on vowel data by difference in n

data is obtained. Thus, the pheromone update function τ is expressed by equation (19).

$$\tau_{i,j} \leftarrow (1-\rho)\tau_{i,j} + \sum_{k=1}^{m} \Delta\tau_{i,j}^{k} + \frac{1}{d^3} \tag{19}$$

$\Delta\tau$ and d are computed by equations (17) and (18). Using equation (19), setting initial values of reference vectors near the decision boundary becomes simple. Since initial values of reference vectors are set from learning data in the proposed method, problems such as those found with the distribution of initial values of reference vectors not fitting into the entire learning data or reference vectors that belong to a different class not appearing in the clustering center of learning data, when compared to SOM-based or k-means-based methods, are solved. Moreover, unlike some other methods in which initial values of reference vectors are set from learning data, but only at random, the proposed method sets initial values of reference vectors primarily from the data derived from near the decision boundary where recognition accuracy may be affected. This allows the proposed method to achieve higher recognition accuracy than other methods, which set initial values of reference vectors at random.

V. Evaluation of the Proposed Method and its Experimental Results

This chapter explains the experimental data used in this study and discusses the methodology used in evaluation of the proposed method and its experimental results.

A. Experimental data

In this study, we used the wine data and English vowel data obtained from the UCI Machine Learning Repository [18].

Wine data contains 13 attributes (1. Alcohol class, 2. Malic acid class, . . . , 13. Proline), and three types of wine are used. There are a total of 178 instances in wine data: 59 for wine 1, 71 for wine 2 and 48 for wine 3.

Vowel data is classification data for English vowels; it classifies 11 types of vowels using 10 attributes. Vowel data only uses data measured as learning data and contains a total of 528 instances with 48 sets for each class data.

B. Experimental method

The experiment was conducted for each data, which is described in V-A, by the K-fold cross-validation method [17]. In this study, the partition number k is 10.

The following show ACO parameters of wine data and vowel data.

- Wine data
  - Number of agents : 150
  - Search count : 10
  - Initial pheromone value : 0.0005
  - Pheromone evaporation rate : 0.8
- Vowel data
  - Number of agents : 100
  - Search count : 50
  - Initial pheromone value : 0.0005
  - Pheromone evaporation rate : 0.8

The following show LVQ parameters of wine data and vowel data.

- Wine data
  - Number of reference vectors : 10 for each class
  - Learning count : 20000
  - Learning rate : 0.01
  - Width of the window : 0.4
- Vowel data
  - Number of reference vectors : 30 for each class
  - Learning count : 20000
  - Learning rate : 0.8
  - Width of the window : 0.5

C. Experimental results of the proposed method

Recognition accuracy of each class obtained by applying the proposed method to wine data is shown in Table I.

TABLE I
Recognition accuracy of each class when applying the proposed method to wine data

	Class 1 (%)	Class 2 (%)	Class 3 (%)
Class 1	81.67	5.00	13.33
Class 2	5.71	63.21	31.07
Class 3	4.00	22.00	74.00

Next, recognition accuracy of each class obtained by applying the proposed method to vowel data is shown in Table II.

VI. Comparison with Conventional Methods

This chapter compares the proposed method with conventional algorithm techniques and discusses its experimental results.

A. Comparison between the proposed method and conventional algorithm techniques

In this section, the following three comparisons of algorithms are discussed: the proposed method with a technique in which initial values of reference vectors are randomly set; the proposed method with an SOM-used method for setting initial values of reference vectors; and the proposed method with a k-means-used method for setting initial values of reference vectors.

1) Random setting method for initial values of reference vectors: The method for randomly setting initial values of reference vectors is a technique in which several sets of data are randomly selected from learning data and then appointed as initial values of reference vectors.

Figure 3 shows a comparison of algorithms between the proposed method and the random setting method for initial values of reference vectors.

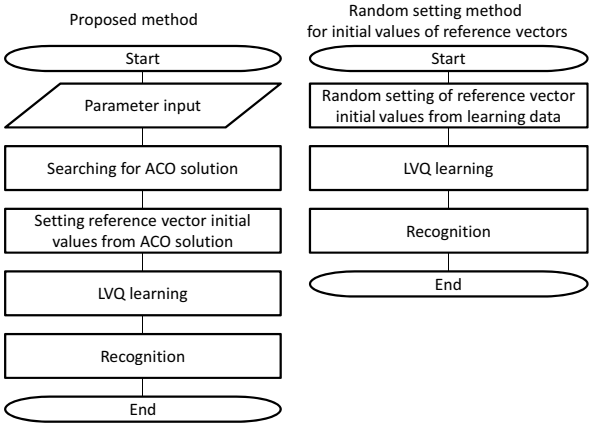

Fig. 3. Difference between the proposed method and random setting method for reference vector initial values

2) SOM-used method for setting initial values of reference vectors: The SOM-used method for setting initial values of reference vectors is a technique in which learning data is trained by SOM and its results are appointed as initial values of reference vectors. There are two major learning methods for SOM. One method is that sets of learning data are labeled by class and each class of learning data is trained by SOM. The other is that all sets of learning data are trained by SOM without consideration to classes. In this study, we adopted the first method in which learning data sets are labeled by class and each class of data is trained by SOM.

Figure 4 shows a comparison of algorithms between the proposed method and the SOM-based method for setting initial values of reference vectors.

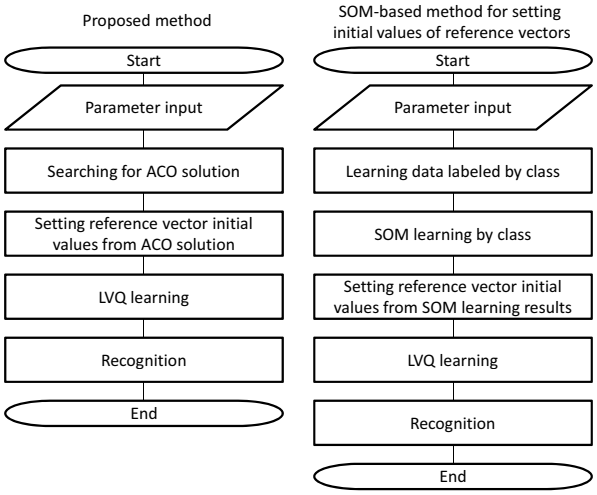

Fig. 4. Difference between the proposed method and SOM-based method for setting reference vector initial values

3) k-means-used method for setting initial values of reference vectors: The k-means-used method for setting initial values of reference vectors is a technique in which learning data is trained by k-means and its results are appointed as initial values of reference vectors. Like SOM, there are two major learning methods for k-means: one method is to label sets of learning data by class and train each class of learning data by k-means and the other is to train all sets of learning data by k-means without consideration to classes. In this study, we adopted the first method in which learning data sets are labeled by class and each class of data is trained by k-means.

Figure 5 shows a comparison of algorithms between the proposed method and the k-means-based method for setting initial values of reference vectors.

TABLE II
RECOGNITION ACCURACY OF EACH CLASS WHEN APPLYING THE PROPOSED METHOD TO VOWEL DATA

	Class 1 (%)	Class 2 (%)	Class 3 (%)	Class 4 (%)	Class 5 (%)	Class 6 (%)	Class 7 (%)	Class 8 (%)	Class 9 (%)	Class 10 (%)	Class 11(%)
Class 1	88.00	12.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Class 2	0.00	98.00	2.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Class 3	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Class 4	0.00	0.00	0.00	98.00	0.00	2.00	0.00	0.00	0.00	0.00	0.00
Class 5	0.00	0.00	0.00	0.00	88.00	6.00	6.00	0.00	0.00	0.00	0.00
Class 6	0.00	0.00	0.00	14.00	4.00	72.67	0.00	0.00	0.00	0.00	9.33
Class 7	0.00	0.00	0.00	0.00	10.00	2.00	76.00	10.00	0.00	2.00	0.00
Class 8	0.00	0.00	0.00	0.00	0.00	0.00	10.00	82.00	6.00	2.00	0.00
Class 9	0.00	0.00	0.00	0.00	2.00	0.00	6.00	2.00	84.00	0.00	6.00
Class 10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	8.00	92.00	0.00
Class 11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00

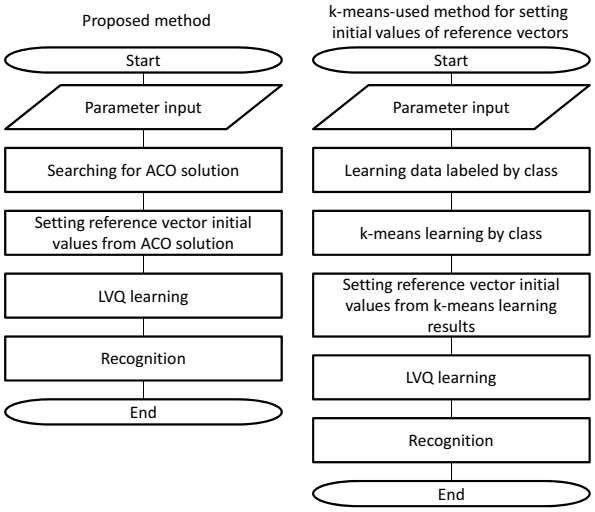

Fig. 5. Difference between the proposed method and k-means-used method for setting reference vector initial values

B. Conventional experimental method

As with the proposed method, the experiment was conducted for each method using the K-fold cross-validation method. The partition number k is 10.

The following show SOM parameters of wine data and vowel data.

- Wine data
  - Learning count : 10000
  - Learning rate : 0.02
  - Radius : 10
  - Map size : 10×10
- Vowel data
  - Learning count : 50000
  - Learning rate : 0.01
  - Radius : 15
  - Map size : 15×15

The following show k-means parameters of wine data and vowel data.

- Learning count : 10
- Number of clusters : Number of classes
- Selection of initial center : Random

Tables III and IV show LVQ parameters of wine data and vowel data.

TABLE III
LVQ PARAMETERS OF WINE DATA

	Random	SOM	k-means	Proposed method
Number of reference vectors	10 for each class	10 for each class	10 for each class	10 for each class
Learning count	20000	20000	20000	20000
Learning rate	0.8	0.01	0.05	0.01
Width of window	0.5	0.3	0.3	0.4

TABLE IV
LVQ PARAMETERS OF VOWEL DATA

	Random	SOM	k-means	Proposed method
Number of reference vectors	30 for each class	30 for each class	30 for each class	30 for each class
Learning count	20000	20000	20000	20000
Learning rate	0.8	0.01	0.6	0.8
Width in window	0.5	0.3	0.3	0.5

C. Evaluation of the experimental results and discussion

Firstly, the experiment was conducted using wine data. Recognition accuracy of each method is shown by class in Tables V, VI and VII.

TABLE V
RECOGNITION ACCURACY OF EACH CLASS WHEN REFERENCE VECTOR INITIAL VALUES ARE RANDOMLY SET (WINE DATA)

	Class 1 (%)	Class 2 (%)	Class 3 (%)
Class 1	84.24	4.92	10.85
Class 2	6.20	61.27	32.54
Class 3	9.79	31.46	58.75

TABLE VI
RECOGNITION ACCURACY OF EACH CLASS WHEN REFERENCE VECTOR INITIAL VALUES ARE SET BY SOM (WINE DATA)

	Class 1 (%)	Class 2 (%)	Class 3 (%)
Class 1	92.00	0.00	8.00
Class 2	7.14	61.61	31.25
Class 3	12.50	32.50	55.00

TABLE VII
RECOGNITION ACCURACY OF EACH CLASS WHEN REFERENCE VECTOR INITIAL VALUES ARE SET BY K-MEANS (WINE DATA)

	Class 1 (%)	Class 2 (%)	Class 3 (%)
Class 1	84.00	6.00	10.00
Class 2	5.71	64.46	29.82
Class 3	9.16	31.67	59.17

Average recognition accuracy of the proposed method and conventional methods when wine data was used is shown in Table VIII.

Then, the experiment was conducted using vowel data. Average recognition accuracy of the proposed method and conventional methods when vowel data was used is shown in Table IX.

It can be noted from Tables VIII and IX that the recognition rate measured by the proposed method is higher than that of the random setting method, SOM-based method or k-means-based method, all of which are for setting initial values of reference vectors. In addition, Tables I, V, VI and VII illustrate that although recognition accuracy for Class 3 is low with the conventional methods, it was improved significantly with the proposed method. We assume that this is because initial values of reference vectors that affect the improvement of recognition accuracy are successfully set. A contributing factor in generating this improvement is that by adding the inverse $\frac{1}{d^3}$ of an exponential function, with the sum of the minimum

TABLE VIII
AVERAGE RECOGNITION ACCURACY OF THE PROPOSED METHOD AND CONVENTIONAL METHODS WHEN WINE DATA IS USED

	Recognition accuracy (%)
Random	68.08
k-means	69.21
SOM	69.33
Proposed method	72.96

TABLE IX
AVERAGE RECOGNITION ACCURACY OF THE PROPOSED METHOD AND CONVENTIONAL METHODS WHEN VOWEL DATA IS USED

	Recognition accuracy (%)
Random	79.43
k-means	87.88
SOM	51.59
Proposed method	88.97

Euclidean distances between a set of data and another set of data in a different class as the base, to the partial evaluation function, the value of the evaluation function becomes larger as the sum of the minimum Euclidean distances between a set of data and another set of data in a different class becomes smaller but does not grow when the sum of such distances becomes larger. If the sum of the minimum Euclidean distances is neither small nor large, i.e. a mean value, then the value of the evaluation function must, as a result, be made slightly larger. By using the sum of the minimum Euclidean distances between a data set and another data set in a different class, in the manner described above, the evaluation value for reference vectors near the decision boundary was successfully determined, which, we assume, has led to the improvement of recognition accuracy. It can be said from these results that the proposed method is effective than the conventional methods.

VII. CONCLUSION

This study proposed an ACO-used method that is effective in initializing reference vectors in LVQ. By adding the inverse of an exponential function, with the sum of the minimum Euclidean distances between a data set and another data set in a different class as the base, to the evaluation function of ACO, we have successfully set initial values of reference vectors close to the decision boundary where recognition accuracy is significantly affected. Furthermore, we have demonstrated the effectiveness of the proposed method by applying it to wine data and English vowel data and comparing its results with those of conventional methods.

REFERENCES

[1] Abhishek Bansal and G. N. Pillai, “High Impedance Fault Detection using LVQ Neural Networks”, International Journal of Computer, Information, and Systems Science, and Engineering 1;3 www.waset.org Summer 2007.
[2] Yuji Mizuno, Hiroshi Mabuchi, Goutam Chakraborty, Masafumi Matsuhara, “Clustering of EEG Data using Maximum Entropy Method and LVQ” INTERNATIONAL JOURNAL OF COMPUTERS, Issue 4, Volume 4, pp.193-200, 2010.
[3] T. Kohonen, “Self-Organizing Maps”, pringer Verlag, 1995.
[4] Atsushi SATO, Keiji YAMADA, “A Formulation of Learning Vector Quantization Using a New Misclassification Measure”, 14th International Conference on Pattern Recognition (ICPR’98) - Volume 1, pp.322-pp325, 1998.
[5] T. Kohonen, “The Self-Organizing Map”, PROCEEDINGS OF THE IEEE, VOL.78, NO.9, pp1464-1480, SEPTEMBER 1990.
[6] John S Baras and Anthony La Vigna, “Convergence of Kohonen’s Learning Vector Quantization”, IJCNN, Vol.3, pp.17-20, 1990.
[7] Kitajima N, “A new method for initializing reference vectors in LVQ”, Proceedings of IEEE International Conference on Neural Networks, Vol.5, pp.2775-2779, 1995.
[8] Marco Dorigo, Luca Maria Gambardella, “Ant Colony System: A Cooperative Learning Approach to the Traveling Salesman Problem”, IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, VOL. 1, NO. 1, pp.53-66, 1997.
[9] B. Bullnheimer, R. F. Hartl and C. Strauss, “An improved Ant System algorithm for the vehicle routing problem”, Technical Report POM10/97, Vienna, Austria: University of Vienna, Institute of Management Science. To appear in Dawid, Feichtinger, & Hartl (Eds.), Annals of Operations Research: Nonlinear Economic Dynamics and Control, 1997.
[10] B. Bullnheimer, R. F. Hartl and C. Strauss, “A new rank-based version of the ant system”, a computational study, Technical Report POM10/97, Institure of Management Science, University of Vienna, 1997.
[11] Thomas St*ü*tzle, Holger H. Hoos, “MAX-MIN Ant System”, Future Generation Computer Systems, Vol.16, No.8, pp.899-914, 2000.
[12] M. Dorigo and G. D. Caro, “Ant Algorithms for Discrete Optimization”, Artificial Life, vol.5, No.2, pp. 137-172, MIT Press(1999).
[13] Marco Dorigo, Vittorio Maniezzo, Alberto Colorni, “The Ant System: Optimization by a colony of cooperating agents”, IIEEE Transactions on Systems, Man, and Cybernetics.Part B, Vol.26, No.1, pp.1-13, 1996.
[14] V.Maniezzo, “Exact and approximate nondeterministic tree-search procedures for the quadratic assignment problem” Technical Report CSR 98-1, C. L. In scenze dell’Informazione, Universita di Bologna, sede di Cesena,1998.
[15] E.-G. Talbi, O. Roux, C. Fonlupt, D. Robillard, “Parallel Ant Colonies for the quadratic assignment problem”, Future Generation Computer Systems 17 pp.441-449, 2001.
[16] Shigeyoshi Tsutsui, “Cunning Ant System for Quadratic Assignment Problem with Local Search and Parallelization”, Proceedings of the Second International Conference on Pattern Recognition and Machine Intelligence, pp. 269-278, Springer-Velag , December 18-22, 2007, Indian Statistical Institute, Kolkata, 2007.12. Springer LNCS 4815, ISBN: 978-3-540-77045-9.
[17] Payam Refaeilzadeh, Lei Tang, Huan Liu, “Cross Validation”, In Encyclopedia of Database Systems, Editors: M. Tamer *Ö*zsu and Ling Liu. Springer, 2009.
[18] C. L. Blake, C. J. Merz,UCI Repositoy Machine Learning Databases, Irvine, University of California, http://archive.ics.uci.edu/ml/, 2010/12.