The Sixteenth International Symposium on Artificial Life and Robotics 2011 (AROB 16th '11),
B-Con Plaza, Beppu,Oita, Japan, January 27-29, 2011

Application of Neural Network Swarm Optimization for Paddy Field Classification from Remote Sensing Data

Kazuma Mori*, Takashi Yamaguchi*, Jong Geol Park** and Kenneth J. Mackin*

** Department of Information Systems, Tokyo University of Information Sciences,*
4-1Onaridai, Wakaba-ku, Chiba, 265-8501 Japan
*** Department of Environmental Information, Tokyo University of Information Science,*
4-1Onaridai, Wakaba-ku, Chiba, 265-8501 Japan
(Tel:81-43-236-1329; Fax:81-43-236-1329)
(s07122km@edu.tuis.ac.jp)

Abstract: Monitoring changes in paddy area is important for economic and environment research since rice is staple food in Asia, and paddy agriculture is a major cropping system in Asia. Recently, remote sensing is used actively to observe the change of paddy area. However, monitoring paddy area by remote sensing is difficult due to the temporal changes of paddy and difference of spatiotemporal characteristics of paddy agriculture between countries or regions. In our previous research using MLP and spatiotemporal satellite sensor data, the proposed classifier yielded 90.8% correct classification rate. In this paper, we proposed a cooperative learning method using PSO as the global search method and MLP as the local search method in order to improve the classification accuracy for practical use.

Keywords: multi-layered perceptron, particle swarm optimization, cooperative learning, classification, remote sensing.

I. INTRODUCTION

Monitoring changes in paddy area is important for economic and environment research since rice is staple food in Asia, and paddy agriculture is thus a major cropping system in Asia. Recently, remote sensing is actively used to observe the change of paddy area. However, monitoring paddy area by remote sensing is difficult due to the temporal changes of paddy and difference of spatiotemporal characteristics of paddy agriculture between countries or regions. To solve this problem, we proposed creating a paddy field classifier by machine learning. Our aim is to automatically create localized classifiers for targeted countries and regions.

In our previous research using multi-layered perceptron (MLP) and spatiotemporal satellite sensor data, the proposed classifier yielded 90.8% correct classification rate [1]. However, it is necessary to further improve accuracy for practical use. One of the known weaknesses of MLP training is the probability of falling into the local optimum causing decline of accuracy. For solving this problem, we proposed a cooperative learning method using particle swarm optimization (PSO) for introducing perturbation to the local search of multiple MLPs in order to improve the accuracy.

In this paper, we applied the proposed artificial neural networks to paddy field classification using moderate resolution sensor data that includes spatiotemporal information. In previous research [2], the teaching signal was either paddy or non-paddy for the input, making it difficult to assess the accuracy of training. On the other hand, promising results for paddy estimation had been reported using paddy area ratio [3]. Thus the teaching signal in this research is modified to using paddy area ratio. Through computer simulation, we investigated the effectiveness of the proposed cooperative learning method and application to paddy field classification using moderate resolution sensor data.

II. METHOD

1. Neural Network Swarm Optimization

In this paper, we proposed a cooperative learning method in multiple MLPs. The proposed method repeats local search and global search in order to find a global optimum in a serialized weight vector space of MLP.

Behavior of the proposed method is shown in Figure 1. First, each MLP (acting as a particle in PSO) searches the neighborhood from its initial position by back-propagation. The search space is a serialized weight vector space of MLP (A in Figure 1). After the back-propagation process, weight values are changed by global search mechanism of PSO (B in Figure 1). The search is repeated from the neighborhood search using the new initial position (updating by the previous PSO search) for each MLP particle (C in Figure 1).

Figure 2 shows a flow chart of the proposed method. Training of MLP is started after the initialization of weight values of each MLP. Training of each MLP is continued to time step τ where τ is a parameter defining the training time step of MLP. Training of PSO is started and continued to time step σ in training process of MLP where σ is a parameter that decide training time step of PSO. Training of MLP and PSO are repeated till the final termination condition is satisfied.

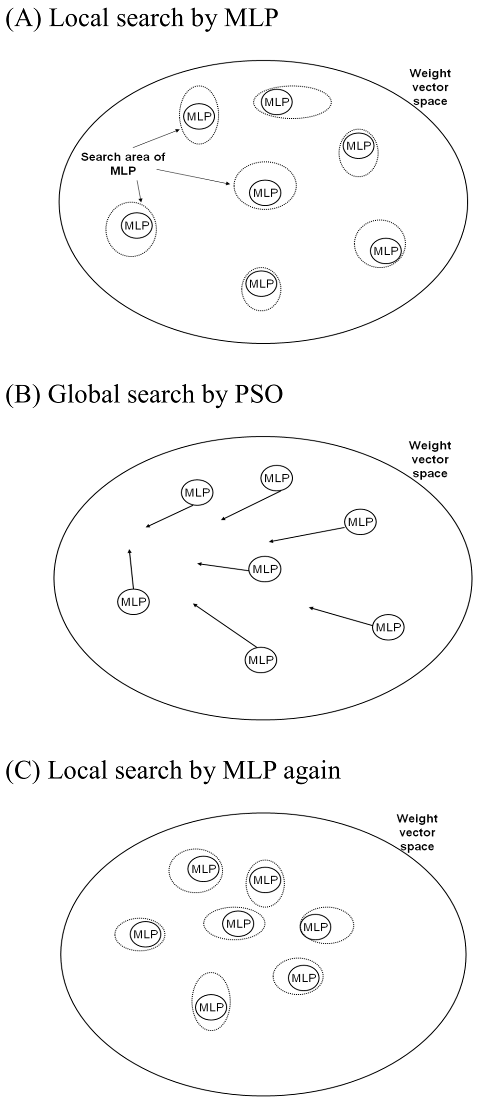

Figure 1. Outline of the proposed method

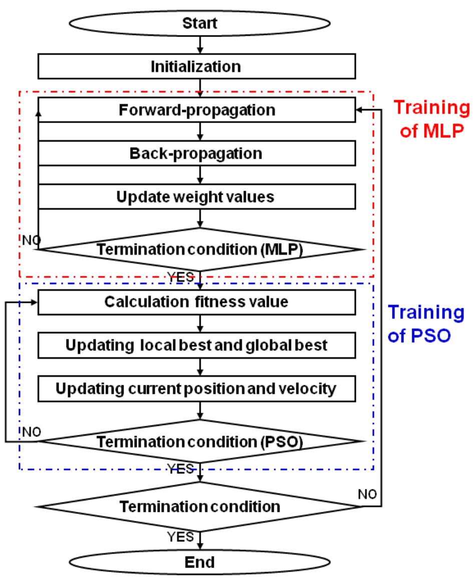

Figure 2. Flow chart of the proposed method

Finally, the MLP classifier with the global best weight values is selected as the final training result.

The detailed algorithm is described below.

2. Multi-Layered Perceptron

MLP is a type of artificial neural network that can approximate complex functions by machine leaning [4].

In this research a standard three-layered MLP is used. The structure of three-layered MLP is shown in Figure 3.

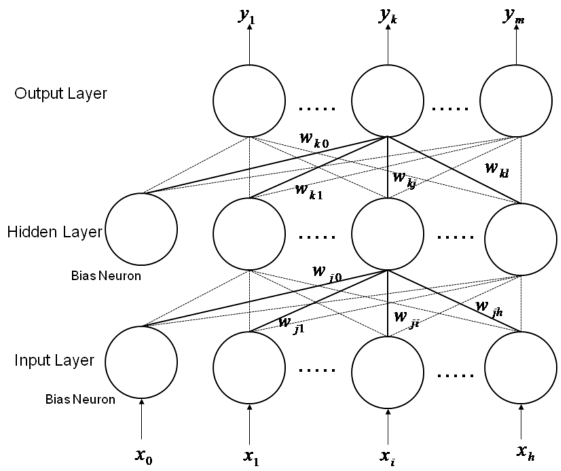

Figure 3. Structure of three-layered MLP

MLP consists of an input layer, a hidden layer, and an output layer. The input layer has h neurons and a bias neuron. The hidden layer has l neurons and a bias neuron. The output layer has m neurons. Each neuron is connected with every neuron in the next layer and each connection has a weight value.

Let $x^d = \{x^d_1, x^d_2, \ldots, x^d_h\}$ be the dth input vector and $t^d = \{t^d_1, t^d_2, \ldots, t^d_m\}$ be the dth teaching signal vector where $d = 1,2,\ldots,n$ and n is the number of samples in training data set. Let $S = \{s^d\}$ be the training set where $s^d = \{x^d, t^d\}$ for the dth sample.

When input signal x_i is given, the output of jth hidden layer neuron z_j is calculated by x_i and w_{ji} where w_{ji} is the weight value between ith input layer neuron and jth hidden layer neuron by equation (1)

$$z_j = f\left(\sum_{i=0}^{h} w_{ji} x_i\right) \tag{1}$$

where $i = 0,1,\ldots,h$; $j = 0,1,\ldots,l$; f is the activation function and x_0 is the output of bias neuron in the input layer and always outputs 1.

The output of kth output layer neuron y_k is calculated by z_j and w_{kj} where w_{kj} is the weight value between kth output layer neuron and jth hidden layer neuron by equation (2)

$$y_k = f\left(\sum_{j=0}^{l} w_{kj} z_j\right) \tag{2}$$

where $k = 1,2,\ldots,m$ and z_0 is the output of bias neuron in the hidden layer and always outputs 1.

For the activation function, a sigmoid function is commonly used. The sigmoid function is defined by equation (3)

$$f(X) = \frac{1}{1+\exp(-aX)} \quad (3)$$

where a is a user defined parameter.

In the proposed method, back-propagation is used for the training method [4]. Back-propagation is based on a steepest descent method. Updating weight values by back-propagation is shown in equation (4) and equation (5).

Let $\Delta w_{ji}(s)$ be the weight modification of ith input layer neuron and jth hidden layer neuron at time step s, and $\Delta w_{kj}(s)$ be weight modification of kth output layer neuron and jth hidden layer neuron at time step s

$$\Delta w_{ji}(s) = -\lambda \cdot \frac{\partial E}{\partial w_{ji}} + \mu \cdot \Delta w_{ji}(s-1) \quad (4)$$

$$\Delta w_{kj}(s) = -\lambda \cdot \frac{\partial E}{\partial w_{kj}} + \mu \cdot \Delta w_{kj}(s-1) \quad (5)$$

where λ is leaning rate, μ is momentum rate and E is the mean square error.

E is derived by the error between output y_k and teaching signal t^d_k as shown in equation (6)

$$E = \frac{1}{2n}\sum_{d=1}^{n}\sum_{k=1}^{m}(t_k^d - y_k^d)^2 \quad (6)$$

where t^d_k is the teaching signal for the output of kth output layer neuron in dth training data and y^d_k is the output of kth output layer neuron in dth training data.

3. Cooperative Learning Method using Particle Swarm Optimization

For the global search method, PSO, a type of population-based search algorithm in evolutionary computation [5], was used in the proposed method. PSO searches for an optimum value in a vector space using multiple particles, similar to other population based search algorithm. For this research, PSO was selected in favor of other population based search methods such as GA, with the expectation that the fast conversion rate of PSO would be advantageous in combination with MLP since the learning in multiple MLPs requires long calculation time.

The particle searches for an optimum by updating current position according to particle velocity. In generation u, updating pth particle velocity $\mathbf{v}^u_p$ and current position $\mathbf{r}^u_p$ is shown as equation (7) and equation (8)

$$\mathbf{v}_p^{u+1} \leftarrow \omega \cdot \mathbf{v}_p^u + \mathbf{u}(0,\phi_1) \otimes (\mathbf{pl}_p^u - \mathbf{r}_p^u) + \mathbf{u}(0,\phi_2) \otimes (\mathbf{pl}_{gb}^u - \mathbf{r}_p^u) \quad (7)$$

$$\mathbf{r}_p^{u+1} \leftarrow \mathbf{r}_p^u + \mathbf{v}_p^{u+1} \quad (8)$$

where $p = 1,2,\ldots,q$; q is the number of particles, $\mathbf{pl}^u_p$ is the position of best fitness for the pth particle (local best) vector in generation u, $\mathbf{pl}^u_{gb}$ is position of best fitness in all particles (global best) vector in generation u, $\mathbf{u}$ is a random number vector from 0 to φ_1 or φ_2, ω is momentum rate and $\otimes$ is Hadamard product.

Let index set of particle be $P = \{1,2,\ldots,q\}$. In updating local best step, if fitness value of pth particle in generation u is better than fitness value of pth particle local best in generation u-1, then $\mathbf{pl}^u_p$ is set to $\mathbf{r}^u_p$, else $\mathbf{pl}^u_p$ is set to $\mathbf{pl}^{u-1}_p$. In updating global best step, index for gb □ P is update to index that fitness value of local best is best in all particle in generation u.

In addition, search space for PSO is a weight vector space of MLP where the weight vector is serialized weight values in MLP. Thus current position $\mathbf{r}$ is defined as follow in the proposed method.

$$\mathbf{r} = \{w_{10},\ldots,w_{ji},\ldots,w_{lh},w_{10},\ldots,w_{kj},\ldots,w_{ml}\} \quad (9)$$

The best position in search space is the position where mean square error between the output and the actual answer for each data is smallest. In other words the fitness function of PSO is error function in MLP.

III. EXPERIMENT

1. Application to MODIS Data

In this research, we used MODIS (Moderate Resolution Imaging Spectoradiometer) satellite data collected at Tokyo University of Information Sciences, Japan, for remote sensing data. MODIS data can be widely used for the remote sensing in land, sea and atmosphere research since MODIS data contains 36 bands. In addition, MODIS observes the same area in high frequency because MODIS satellites have a fast orbit compared to high-resolution satellites. Therefore MODIS is effective for monitoring annual cycle of a paddy area.

In this research, spatiotemporal MODIS band 1 (red), band 2 (infra red) and band 6 (short wave infra red) data were used similar to previous research [2].

In previous research [2], input signal consisting of monthly band data from January to December (3 bands x 12 months inputs) was used. The band data of each month was derived from 1-month composite MODIS sensor data of 500m resolution. For the teaching signal, we constructed a land-truth data derived from 1km resolution mesh land-use data from digital national land information provided by the Japanese Ministry of Land, Infrastructure, Transport and Tourism (JMLIT). This value is paddy or non-paddy {1,0}.

In this research, the input signal consists of 4 pixels refraction values of 500m resolution at the corresponding 1km square (4 pixels x 3 bands x 12 months inputs). The value of input is normalized to [0,1]. 4 Neighboring 500m MODIS pixels are merged to a 1km square. The 1km square merged teaching signal is the estimated paddy area ratio. The value of teaching signal is [0,1]. Therefore, it is difficult to approximate for an unknown function in comparison with past research data, and a difference of the capability for classification appeared.

In addition, the answer value of the test data and the output of classifier are divided into paddy or non-paddy by a threshold θ for classification. The threshold θ is defined by equation (10)

$$g(o) = \begin{cases} 1 & if \quad o > \theta \\ 0 & otherwize \end{cases} \tag{10}$$

where o is y^b that is the final output in bth test data or t^b is the answer value for the output in bth test data, b = 1,2,…,c and c is the number of samples in test data set.

In this experiment, the parameters were defined as follows. Leaning rate of MLP λ is set to 0.1, momentum rate of MLP μ is set to 0.05, input size h is set to 144, hidden size l is set to 144, output size m is set to 1, particle size q is set to 50, momentum rate of PSO ω is set to 0.9, upper limit of random number φ_1 and φ_2 is set to 2.0, training time step of MLP τ is set to the number of sample x 10, training time step of PSO σ is set to 10, threshold θ is set to 0.4. In addition, the training of proposed method ends when training loop of MLP and PSO are repeated 50 times. The training step and the test step were repeated 10 times, each time changing the initial values and training data and test data, and the average accuracy for the 10 trials were used to evaluate the calcification accuracy.

2. Experiment Result

In this experiment, we evaluated classification accuracy by using the proposed paddy classifier. For evaluating classification accuracy, data set was divided into 2 disjoint subsets, training data set and test data set, by using random sampling from the north region of Chiba, Japan. The number of test data set was 10% of the number of training data set.

Table 1 shows classification accuracy of a) proposed paddy classifier with input signal of 1km square (proposed NNSO), b) previous method with input signal consisting of monthly band data from January to December (previous NNSO), c) and MLP using bagging with input signal of 1km square (MLP-Bagging).

Experiment result showed improved an accuracy of proposed NNSO compared with previous NNSO. In addition, total accuracy of proposed NNSO improved by about 0.3% compared to total accuracy of the MLP-bagging. Notably, as much as 9% improvement over paddy accuracy was seen. In this experiment, the problem is more difficult since the teaching signal is a continuous value. The experimental results showed that the proposed method has better function approximation capability for continuous values.

Table 1. The comparison of classification accuracy in proposed paddy classifier.

	Total	Paddy	Non-Paddy
Proposed NNSO	0.889	0.751	0.946
Previous NNSO	0.867	0.752	0.907
MLP-Bagging	0.886	0.659	0.932

IV. CONCLUSION

In this paper, we proposed applying a cooperative learning method where multiple MLP are used as PSO particles, and weight values of MLP is affected with training of PSO for improving the classifier accuracy.

Experiment result showed the accuracy of proposed NNSO improved in comparison with previous NNSO. However, the dimensions of the weight values were increased. The number of the neuron in the input layer and the hidden layer were fixed in the experiment, and unnecessary weight training may be exists in the process of training. We plan to investigate the structure which can optimize the number of the dimensions automatically.

In addition, the proposed NNSO yields higher accuracy than the MLP-bagging. The proposed method creates a classifier using only 1 MLP with weight values set to the global best, and is advantageous in that the proposed method requires less memory in comparison with commonly used bagging method.

In this research, the classifier output was a binary value of paddy or non-paddy, but the accuracy was improved by using a continuous value of paddy area ration for classifier training. The experiment results showed an advantage of the proposed method in approximation capability. For future works, we plan to investigate the methods to apply the proposed method to estimate paddy area ratio for a given input area.

ACKNOWLEDGMENTS

This research was supported by the Research project of Tokyo University of Information Sciences for the sustainable development of economic and social structure dependent on the environment in eastern Asia.

REFERENCES

[1] T.Yamaguchi, K.Kishida, E.Nunohiro, J.G.Park, K.J.Mackin, K.Hara, K.Matsushita and I.Harada(2010), Artificial Neural Networks Paddy Field Classifier using Spatiotemporal Remote Sensing Data. Artificial Life and Robotics, vol.15, pp.221-224

[2] K.Mori, T.Yamaguchi, J.G.Park and K.J. Mackin(2010), Cooperative Learning Artificial Neural Networks using Particle Swarm Optimization for Paddy Field Classification from Remote Sensing Data (in Japanese). Proceedings of the Fuzzy System Symposium, vol.26, pp.642-645

[3] W.Takeuchi and Y.Yasuoka(2006), Mapping of fractional coverage of paddy fields over China using MODIS 250m data (in Japanese). Journal of The Remote Sensing Society of Japan, vol.26, pp.213-223

[4] D.E.Rumelhart, G.E.Hinton and R.J.Williams(1986), Learning representations by backpropagating errors. Nature, vol.323, pp.533-536

[5] R.Poli, J.Kennedy and T.Blackwell(2007), Particle swarm optimization An overview. Swarm Intelligence, vol.1, no.1, pp.33-57, Springer