Self-Organizing Neural Grove and Its Application to Incremental Learning

Hirotaka Inoue

Department of Electrical Engineering and Information Science,
Kure National College of Technology,
2-2-11 Agaminami, Kure, Hiroshima, 737-8506 Japan
hiro@kure-nct.ac.jp

Abstract. Recently, multiple classifier systems have been used for practical applications to improve classification accuracy. Self-generating neural networks (SGNN) are one of the most suitable base-classifiers for multiple classifier systems because of their simple settings and fast learning ability. However, the computation cost of the multiple classifier system based on SGNN increases in proportion to the numbers of SGNN. In this paper, we propose a novel pruning method for efficient classification and we call this model a self-organizing neural grove (SONG). Experiments have been conducted to compare the SONG with bagging and the SONG with boosting, the multiple classifier system based on C4.5, and support vector machine (SVM). The results show that the SONG can improve its classification accuracy as well as reducing the computation cost. Additionally, we investigate SONG's incremental learning performance.

1 Introduction

Classifiers need to find hidden information within a large amount of given data effectively and classify unknown data as accurately as possible [1]. Recently, to improve the classification accuracy, multiple classifier systems such as neural network ensembles, bagging, and boosting have been used for practical data mining applications [2]. In general, base classifiers of multiple classifier systems use traditional models such as neural networks (backpropagation network and radial basis function network) [3] and decision trees (CART and C4.5) [4].

Neural networks have great advantages such as adaptability, flexibility, and universal nonlinear input-output mapping capability. However, to apply these neural networks, it is necessary for the network structure and some parameters to be determined by human experts, and it is quite difficult to choose the right network structure suitable for a particular application at hand. Moreover, they require a long training time to learn the input-output relation of the given data. These drawbacks prevent neural networks from being the base classifier of multiple classifier systems for practical applications.

Self-generating neural networks (SGNN) [5] have a simple network design and high speed learning. SGNN are an extension of the self-organizing maps (SOM) of Kohonen [6] and utilize the competitive learning which is implemented as a self-generating neural tree (SGNT). The abilities of SGNN make it suitable for the base classifier of multiple classifier systems. In order to improve in the accuracy of SGNN, we proposed

Z.-H. Zhou, F. Roli, and J. Kittler (Eds.): MCS 2013, LNCS 7872, pp. 109–120, 2013.
© Springer-Verlag Berlin Heidelberg 2013

ensemble self-generating neural networks (ESGNN) for classification [7] as one of multiple classifier systems. Although the accuracy of ESGNN improves by using various SGNN, the computation cost, that is the computation time and the memory capacity increases in proportion to the increase in numbers of SGNN in multiple classifier systems.

In an earlier paper [8], we proposed a pruning method for the structure of the SGNN in multiple classifier systems to reduce the computation cost. In this paper, we propose a novel pruning method for more effective processing and we call this model a self-organizing neural grove (SONG). This pruning method is constructed in two stages. In the first stage, we introduce an on-line pruning algorithm to reduce the computation cost by using class labels in learning. In the second stage, we optimize the structure of the SGNT in multiple classifier systems to improve the generalization capability by pruning the redundant leaves after learning. In the optimization stage, we introduce a threshold value as a pruning parameter to decide which subtree's leaves to prune and estimate with 10-fold cross-validation [9]. After the optimization, the SONG improve its classification accuracy as well as reducing the computation cost. We use bagging [10] and boosting [11] as a resampling technique for the SONG.

We investigate the improvement performance of the SONG by comparing it with a multiple classifier system based on C4.5 [12] using ten problems in the UCI machine learning repository [13]. Moreover, we compare the SONG with support vector machine (SVM) [14] to investigate the computational cost and the classification accuracy.

The rest of the paper is organized as follows. The next section shows how to construct the SONG. Section 3 shows the experimental results. Then section 4 is devoted to some experiments to investigate the incremental learning performance of SONG. Finally we present some conclusions, and outline plans for future work.

2 Constructing Self-Organizing Neural Grove

In this section, we describe how to prune redundant leaves in the SONG. First, we mention the on-line pruning method in the learning of SGNT. Second, we show the optimization method in constructing the SONG. Finally, we show a simple example of the pruning method for a two dimensional classification problem.

2.1 On-Line Pruning of Self-Generating Neural Tree

SGNN are based on SOM and are implemented as an SGNT architecture. The SGNT can be constructed directly from the given training data without any intervening human effort. The SGNT algorithm is defined as a tree construction problem of how to construct a tree structure from the given data which consists of multiple attributes under the condition that the final leaves correspond to the given data.

Before we describe the SGNT algorithm, we denote some notations.

- input data vector: $\boldsymbol{e}_i \in \mathbb{R}^m$.
- root, leaf, and node in the SGNT: n_j.
- weight vector of n_j: $\boldsymbol{w}_j \in \mathbb{R}^m$.

Table 1. Sub procedures of the SGNT algorithm

Sub procedure	Specification
$copy(n_j, \boldsymbol{e}_i/\boldsymbol{w}_{win})$	Create n_j, copy $\boldsymbol{e}_i/\boldsymbol{w}_{win}$ as $\boldsymbol{w}_j$ in n_j.
$choose(\boldsymbol{e}_i, n_1)$	Decide n_{win} for $\boldsymbol{e}_i$.
$leaf(n_{win})$	Check n_{win} whether n_{win} is a leaf or not.
$connect(n_j, n_{win})$	Connect n_j as a child leaf of n_{win}.
$prune(n_{win})$	Prune leaves if the leaves have the same class.

- the number of the leaves in n_j: c_j.
- distance measure: $d(\boldsymbol{e}_i, \boldsymbol{w}_j)$.
- winner leaf for $\boldsymbol{e}_i$ in the SGNT: n_{win}.

The SGNT algorithm is a hierarchical clustering algorithm. The pseudo C code of the SGNT algorithm is given as follows:

Algorithm (SGNT Generation)

```
Input:
  A set of training examples E = {e_i},
 i = 1, ... , N.
  A distance measure d(e_i,w_j).
Program Code:
  copy(n_1,e_1);
  for (i = 2, j = 2; i <= N; i++) {
    n_win = choose(e_i, n_1);
    if (leaf(n_win)) {
      copy(n_j, w_win);
      connect(n_j, n_win);
      j++;
    }
    copy(n_j, e_i);
    connect(n_j, n_win);
    j++;
    prune(n_win);
  }
Output:
  Constructed SGNT by E.
```

In the above algorithm, several sub procedures are used. Table 1 shows the sub procedures of the SGNT algorithm and their specifications.

In order to decide the winner leaf n_{win} in the sub procedure `choose(e_i,n_1)`, competitive learning is used. This sub procedure is recursively used from the root to the leaves of the SGNT. If an n_j includes the n_{win} as its descendant in the SGNT, the weight w_{jk} $(k = 1, 2, \ldots, m)$ of the n_j is updated as follows:

$$w_{jk} \leftarrow w_{jk} + \frac{1}{c_j} \cdot (e_{ik} - w_{jk}), \quad 1 \leq k \leq m. \tag{1}$$

In the SGNT, the input vector $\boldsymbol{x}_i$ corresponds to $\boldsymbol{e}_i$, and the desired output y_i corresponds to the network output o_i which is stored in one of the leaf neurons, for $(\boldsymbol{x}_i, y_i) \in D$. Here, D is the training data set which consists of data $\{\boldsymbol{x}_i, y_i | i = 1, \ldots, N\}$, $\boldsymbol{x}_i \in \mathbb{R}^m$ is the input and y_i is the desired output. After all training data are inserted into the SGNT as the leaves, the leaves each have a class label as the outputs and the weights of each node are the averages of the corresponding weights of all its leaves. The whole network of the SGNT reflects the given feature space by its topology.

We explain the SGNT generation algorithm using an simple example. In this example, m is one and the four training data (x_i, y_i) is (1,1), (2,2), (3,3), and (4,4). Hence, $e_{11} = 1, e_{21} = 2, e_{31} = 3$, and $e_{41} = 4$. Fig. 1 shows an example of the SGNT generation. First, e_{11} is just copied to a neuron n_1 as the root, and e_{11} is substituted to w_{11} (Fig. 1 (a)). In Fig. 1, the circle is the neuron, the integer in the circle is the number of neuron j, the integer of left-upper of the circle is c_j, and the integer of under the circle is w_{j1}. Next, n_2 and n_3 are generated as the children of n_1 with $w_{21} = 1, w_{31} = 2$. w_{11} is updated by e_{21} to $1 + 1/2(2 - 1) = 1.5$ (Fig. 1 (b)). Next, the winner in $\{n_1, n_2, n_3\}$ is n_3 since $d(\boldsymbol{e}_3, \boldsymbol{w}_1) = 1.5$, $d(\boldsymbol{e}_3, \boldsymbol{w}_2) = 2$, and $d(\boldsymbol{e}_3, \boldsymbol{w}_3) = 1$; and thus, n_4 and n_5 are generated as the children of n_3 with $w_{41} = 2, w_{51} = 3$. w_{31} is updated by e_{31} to $2 + 1/2(3 - 2) = 2.5$ and w_{11} is updated by e_{31} to $1.5 + 1/3(3 - 1.5) = 2$ (Fig. 1 (c)). Finally, n_6 and n_7 are generated as the children of n_5 with $w_{61} = 3$, $w_{71} = 4$. w_{51} is updated by e_{41} to $3 + 1/2(4 - 3) = 3.5$, w_{31} is updated by e_{41} to $2.5 + 1/3(4 - 2.5) = 3$, and w_{11} is updated by e_{41} to $2 + 1/4(4 - 2) = 2.5$ (Fig. 1 (d)).

Note, to optimize the structure of the SGNT effectively, we remove the threshold value of the original SGNT algorithm in [5] to control the number of leaves based on the distance because of the trade-off between the memory capacity and the classification accuracy. In order to avoid the above problem, we introduce a new pruning method in the sub procedure `prune(n_win)`. We use the class label to prune leaves. For leaves that have the n_{win}'s parent node, if all leaves belong to the same class, then these leaves are pruned and the parent node is given to the class.

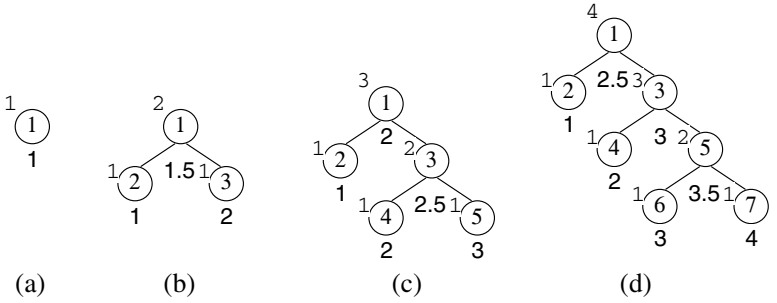

Fig. 1. An example of the SGNT generation

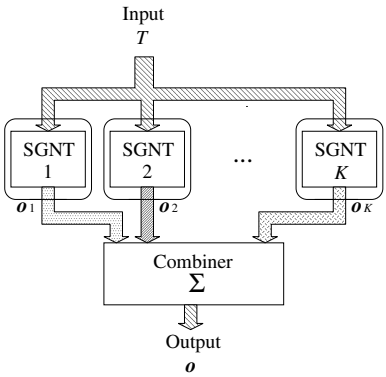

Fig. 2. The SONG which is constructed from K SGNTs. The test dataset T is entered at each SGNT, the output $\boldsymbol{o}_i$ is computed as the output of the winner leaf for the input data, and the SONG's output is decided by voting outputs of K SGNTs.

```
1 begin    initialize j = the height of the SGNT
2   do for each subtree's leaves in the height j
3     if the ratio of the most class ≥ α,
4     then merge all leaves to parent node
5     if all subtrees are traversed in the height j,
6     then j ← j − 1
7   until j = 0
8 end.
```

Fig. 3. The merge phase

2.2 Optimization of the SONG

The SGNT has the capability of high speed processing. However, the accuracy of the SGNT is inferior to the conventional approaches, such as nearest neighbor, because the SGNT has no guarantee to reach the nearest leaf for unknown data. Hence, we construct the SONG by taking the majority of multiple SGNT's outputs to improve the accuracy (Fig. 2).

Although the accuracy of the SONG is superior or comparable to the accuracy of conventional approaches, the computational cost increases in proportion to the increase in the number of SGNTs in the SONG. In particular, the huge memory requirement prevents the use of SONG for large datasets even with the latest computers.

In order to improve the classification accuracy, we propose an optimization method of the SONG for classification. This method has two parts, the merge phase and the evaluation phase. The merge phase is performed as a pruning algorithm to reduce dense leaves (Fig. 3).

This phase uses the class information and a threshold value α to decide which subtree's leaves to prune or not. For leaves that have the same parent node, if the proportion

```
1 begin initialize α = 0.5
2   do for each α
3     evaluate the merge phase with 10-fold CV
4     if the best classification accuracy is obtained,
5     then record the α as the optimal value
6     α ← α + 0.05
7   until α = 1
8 end.
```

Fig. 4. The evaluation phase

of the most common class is greater than or equal to the threshold value α, then these leaves are pruned and the parent node is given the most common class.

The optimum threshold values α of the given problems are different from each other. The evaluation phase is performed to choose the best threshold value by introducing 10-fold cross validation (Fig. 4).

2.3 An Example of the Pruning Method for the SONG

We show an example of the pruning method for the SONG in Fig. 5. This is a two-dimensional classification problem with two equal circular Gaussian distributions that have an overlap. The shaded plane is the decision region of class 0 and the other plane is the decision region of class 1 by the SGNT. The dotted line is the ideal decision boundary. The number of training samples is 200 (class0: 100, class1: 100) (Fig. 5(a)).

The unpruned SGNT is given in Fig. 5(b). In this case, 200 leaves and 120 nodes are automatically generated by the SGNT algorithm. In this unpruned SGNT, the height is 7 and the number of units is 320. In this, we define the unit to count the sum of the root, nodes, and leaves of the SGNT. The root is the node which is of height 0. The unit is used as a measure of the memory requirement in the next section. Fig. 5(c) shows the pruned SGNT after the optimization stage in $\alpha = 1$. In this case, 159 leaves and 107 nodes are pruned away and 48 units remain. The decision boundary is the same as the unpruned SGNT. Fig. 5(d) shows the pruned SGNT after the optimization stage in $\alpha = 0.6$. In this case, 182 leaves and 115 nodes are pruned away and only 21 units remain. Moreover, the decision boundary is improved more than the unpruned SGNT because this case can reduce the effect of the overlapping class by pruning the SGNT.

In the above example, we use all training data to construct the SGNT. The structure of the SGNT is changed by the order of the training data. Hence, we can construct the SONG from the same training data by changing the input order. We investigate the pruning method for more complex problems in the next section.

3 Experimental Results

We investigate the computational cost (the memory capacity and the computation time) and the classification accuracy of the SONG with bagging for ten benchmark problems in the UCI machine learning repository [13]. We evaluate how the SONG is pruned

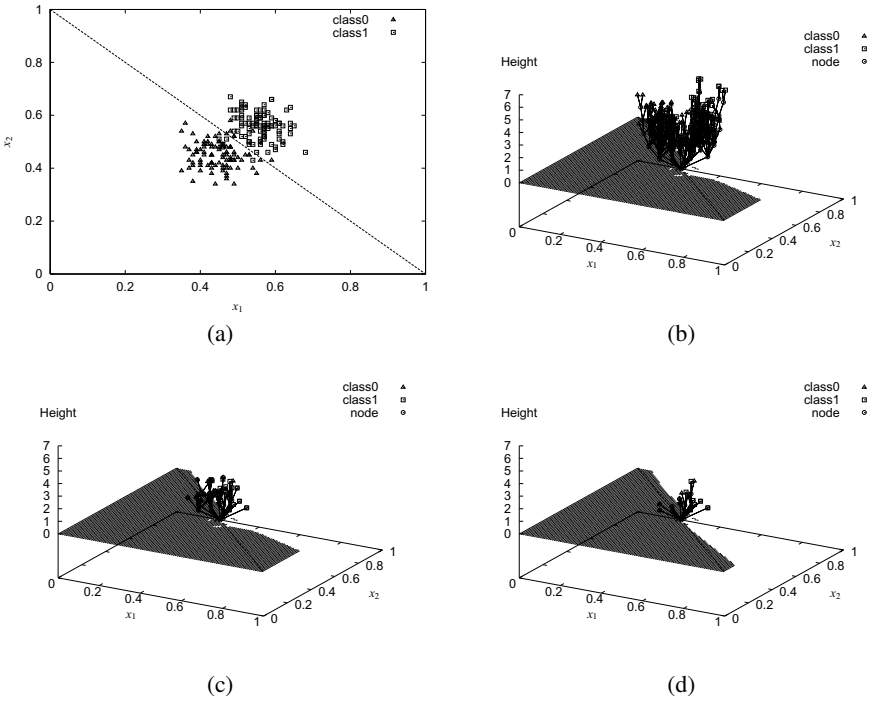

Fig. 5. An example of the SONG's pruning algorithm, (a) a two dimensional classification problem with two equal circular Gaussian distribution, (b) the structure of the unpruned SGNT, (c) the structure of the pruned SGNT ($\alpha = 1$), and (d) the structure of the pruned SGNT ($\alpha = 0.6$). The shaded plane is the decision region of class 0 by the SGNT and the dotted line shows the ideal decision boundary.

using 10-fold cross-validation for the ten benchmark problems. In this experiment, we use a modified Euclidean distance measure for the SONG. Since the performance of the SONG is not sensitive to the threshold value α, we set the different threshold values α to vary from 0.5 to 1; $\alpha = [0.5, 0.55, 0.6, \ldots, 1]$. We set the number of SGNT K in the SONG as 25 and execute 100 trials by changing the sampling order of each training set. All experiments in this section were performed on an UltraSPARC workstation with a 900MHz CPU, 1GB RAM, and Solaris 8.

Table 2 shows the average memory requirement and classification accuracy of 100 trials for the SONG. As the memory requirement, we count the number of units which is the sum of the root, nodes, and leaves of the SGNT. The average memory requirement is reduced from 65% to 96.6% and the classification accuracy is improved 0.1% to 2.9% by optimizing the SONG. This supports that the SONG can be effectively used for all datasets with regard to both the computation cost and the classification accuracy.

Table 3 shows the average classification accuracy of 10 trials for the SONG with bagging and boosting. On boosting, we implement AdaBoost [11] to the SONG. Since

Table 2. The average memory requirement and classification accuracy of 100 trials for the bagged SGNT in the SONG. The standard deviation is given inside the bracket on classification accuracy ($\times 10^{-3}$).

	memory requirement			classification accuracy		
Dataset	pruned	unpruned	ratio	pruned	unpruned	ratio
balance-scale	107.68	861.18	12.5	0.866(6.36)	0.837(7.83)	+2.9
breast-cancer-w	30.88	897.37	3.4	0.97(2.41)	0.966(2.71)	+0.4
glass	104.33	297.75	35	0.714(13.01)	0.709(14.86)	+0.5
ionosphere	50.75	472.39	10.7	0.891(6.75)	0.862(7.33)	+2.9
iris	15.64	208.56	7.4	0.962(6.04)	0.955(5.45)	+0.7
letter	6197.5	27028.56	22.9	0.956(0.77)	0.955(0.72)	+0.1
liver-disorders	163.12	471.6	34.5	0.648(12.89)	0.636(13.36)	+1.2
new-thyroid	49.45	298.21	16.5	0.958(7.5)	0.957(7.49)	+0.1
pima-diabetes	204.4	1045.03	19.5	0.749(7.05)	0.728(7.83)	+2.1
wine	15	238.95	6.2	0.976(4.41)	0.972(5.57)	+0.4
Average	693.88	3181.96	16.9	0.869	0.858	+1.1

Table 3. The average classification accuracy of 10 trials for the SONG with bagging and boosting. The standard deviation is given inside the bracket ($\times 10^{-3}$).

	SONG with bagging			SONG with boosting		
Dataset	SGNT	SONG	ratio	SGNT	SONG	ratio
breast-cancer-w	0.96(4.74)	0.975(2.86)	+1.5	0.96(6.47)	0.957(4.13)	-0.3
ionosphere	0.847(19.3)	0.89(8.23)	+4.3	0.854(18.26)	0.773(17.4)	-8.1
liver-disorders	0.571(21.4)	0.636(11.0)	+6.5	0.588(17.0)	0.572(24.3)	-1.6
pima-diabetes	0.705(9.8)	0.754(4.96)	+4.9	0.696(12.2)	0.722(6.82)	+2.6
Average	0.771	0.814	+4.3	0.775	0.756	-1.9

original AdaBoost algorithm have been proposed for binary classification problems, we use four binary classification problems in Table 3. In comparison with boosting, bagging is superior to boosting on all of the 4 datasets. In short, bagging is better than boosting in terms of the classification accuracy.

To evaluate the SONG's performance, we compare the SONG with a multiple classifier system based on C4.5. We set the number of classifiers K in the multiple classifier system as 25 and we construct both multiple classifier systems by bagging. Table 4 shows the improved performance of the SONG and the multiple classifier system based on C4.5. The results of the SGNT and the SONG are the average of 100 trials. The SONG has a better performance than the multiple classifier system based on C4.5 for 6 of the 10 datasets. Although the multiple classifier system based on C4.5 degrades the classification accuracy for iris, the SONG can improve the classification accuracy for all problems. Therefore, the SONG is an efficient multiple classifier system on the basis of both the scalability for large scale datasets and the robustly improved generalization capability for the noisy datasets comparable to the multiple classifier system with C4.5.

Table 4. The improved performance of the SONG based on pruned SGNT and the multiple classifier system (MCS) based on C4.5 with bagging

	SONG based on SGNT			MCS based on C4.5		
Dataset	SGNT	SONG	ratio	C4.5	MCS	ratio
balance-scale	0.779	**0.866**	+8.7	0.795	0.827	+3.2
breast-cancer-w	0.956	**0.97**	+1.4	0.946	0.963	+1.7
glass	0.642	0.714	+7.2	0.664	**0.757**	+9.3
ionosphere	0.852	0.891	+3.9	0.897	**0.92**	+2.3
iris	0.943	**0.962**	+1.9	0.953	0.947	−0.6
letter	0.879	**0.956**	+7.7	0.880	0.938	+5.8
liver-disorders	0.59	0.648	+5.8	0.635	**0.736**	+10.1
new-thyroid	0.939	**0.958**	+1.9	0.93	0.94	+1
pima-diabetes	0.695	0.749	+5.4	0.749	**0.767**	+1.8
wine	0.955	**0.976**	+2.1	0.927	0.949	+2.2
Average	0.823	0.869	+4.6	0.837	**0.874**	+3

To show the advantages of the SONG, we compare it with SVM on the same problems. In the SONG, we choose the best classification accuracy of 100 trials with bagging. In SVM, we use C-SVM in libsvm [14] with radial basis function kernel. We select the parameters of SVM, the cost parameters C and the kernel parameters γ, from $15 \times 15 = 225$ combinations by 10-fold cross validation; $C = [2^{12}, 2^{11}, 2^{10}, \ldots, 2^{-2}]$ and $\gamma = [2^4, 2^3, 2^2, \ldots, 2^{-10}]$. We normalize the input data from 0 to 1 for all problems in k-nearest neighbor and SVM. All methods are compiled by using gcc with the optimization level `-O2` on the same workstation.

Table 5 shows the classification accuracy, the memory requirement, and the computation time achieved by the SONG and SVM. Next, we show the results for each category.

First, in view point of the classification accuracy, the SONG superior to SVM 3 of the 10 datasets and degrade 1.7% in the average. Second, in terms of the memory requirement, even though the SONG includes the root and the nodes which are generated by the SGNT generation algorithm, this is less than SVM for 8 of the 10 datasets. Although the memory requirement of the SONG is totally used K times in Table 5, we release the memory of SGNT for each trial and reuse the memory for effective computation. Therefore, the memory requirement is suppressed by the size of the single SGNT. Finally, in view of the computation time, although the SONG consumes the cost of K times of the SGNT to construct the model and test for the unknown dataset, the average computation time is faster than SVM. The SONG is slower than SVM for small datasets such as glass, ionosphere, and iris. However, the SONG is faster than SVM for large datasets such as balance-scale, letter, and pima-diabetes. Especially, in letter, the computation time of the SONG is faster than SVM about 11 times. We need to repeat 10-fold cross validation many times to select the optimum parameter for α, k, C, and γ. This evaluation consumes much computation time for large datasets such as letter.

Table 5. The classification accuracy, the memory requirement, and the computation time of ten trials for the best pruned SONG and SVM

	classification acc.		memory requirement		computation time (s)	
Dataset	SONG	SVM	SONG	SVM	SONG	SVM
balance-scale	0.885	**0.992**	109.93	**60.6**	**0.82**	4.77
breast-cancer-w	**0.976**	0.973	**26.8**	79.6	**1.18**	**0.64**
glass	**0.758**	0.738	**91.33**	132.4	**0.36**	0.61
ionosphere	0.912	**0.954**	**51.38**	147.9	1.93	**1.25**
iris	**0.973**	0.96	**11.34**	51.3	0.13	**0.06**
letter	0.958	**0.977**	**6208.03**	7739.7	**208.52**	2359.39
liver-disorders	0.685	**0.73**	**134.17**	214.5	**0.54**	2.07
new-thyroid	0.972	**0.977**	45.74	**44.1**	0.23	**0.22**
pima-diabetes	0.764	**0.766**	**183.57**	363.5	**1.72**	5.63
wine	0.983	**0.989**	**11.8**	62.2	0.31	**0.15**
Average	0.887	**0.904**	**687.41**	889.58	**21.57**	236.88

Therefore, the SONG based on the fast and compact SGNT is useful and practical for large datasets. Moreover, the SONG has the ability of parallel computation because each classifier behaves independently. In conclusion, the SONG is a practical method for large-scale data mining compared with SVM.

4 Considerations

In this section, we investigate the performance of the incremental learning of the SONG. We use letter in this experiment since it contains large scale data (the number of input dimension: 16, the number of classes: 26, and the number of entries: 20000).

This experiment is performed as follows. First, we divide letter dataset into ten parts. Second, we select one of the ten parts as the testing data. Third, we enter one of the remaining nine parts into the SONG for training. Forth, we test the SONG using the testing data. Finally, we continue the training and the testing until all nine parts of the dataset are entered into the SONG.

Fig. 6 shows the relation between the number of training data and the classification accuracy. The more the number of training data increases, the more the classification accuracy improves for all the number of ensembles K. The width of the improvement is wide for small K and all values of N.

As the memory requirement, we count the number of units which is the sum of the root, nodes, and leaves of the SGNT. Fig. 7 shows the relation between the number of training data N and the number of units in $\alpha = 1$. Here, the total units are the number of all units without pruning and the remaining units are the number of all units with pruning. Both of them are the average of 25 SGNTs. The number of nodes increases linearly in proportion to the increase in the number of training data. The slope of the remaining units is smaller than the slope of the total nodes. This means that the SONG has the capability for good compression for large scale data. This supports that the SONG can be effectively used for large scale datasets.

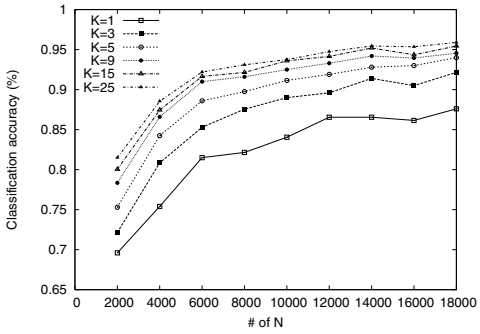

Fig. 6. The relation between the number of training data and the classification accuracy

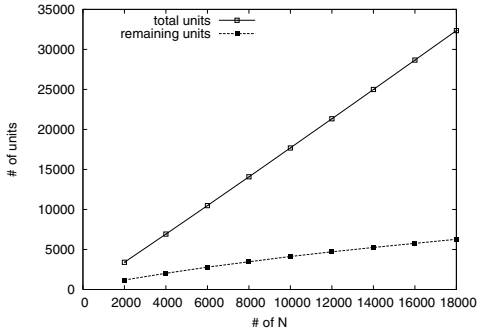

Fig. 7. The relation between the number of training data and the number of units

5 Conclusions

In this paper, we proposed a new pruning method for the multiple classifier system based on SGNT, which is called SONG, and evaluated the computation cost and the accuracy. We introduced an on-line and off-line pruning method and evaluated the SONG by 10-fold cross-validation. Experimental results showed that the memory requirement reduced remarkably, and the accuracy increased by using the pruned SGNT as the base classifier of the SONG. Additionaly, we investigated an incremental learning performance of the SONG. Experimental results showed that the SONG could be applicable to incremental learning. The SONG is a useful and practical multiple classifier system to classify large datasets. In future work, we will study a parallel and distributed processing of the SONG for large scale data mining.

Acknowledgment. The author would like to thank Kyoshiro Sugiyama for implementing AdaBoost algorithm on the SONG, Anthony Nepia and the anonymous referees for their helpful comments.

References

1. Han, J., Kamber, M.: Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, San Francisco (2000)
2. Quinlan, J.R.: Bagging, Boosting, and C4.5. In: Proceedings of the Thirteenth National Conference on Artificial Intelligence, Portland, OR, August 4-8, pp. 725–730. AAAI Press, The MIT Press (1996)
3. Bishop, C.M.: Neural Networks for Pattern Recognition. Oxford University Press, New York (1995)
4. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification, 2nd edn. John Wiley & Sons Inc., New York (2000)
5. Wen, W.X., Jennings, A., Liu, H.: Learning a neural tree. In: The International Joint Conference on Neural Networks, Beijing, China, November 3-6, vol. 2, pp. 751–756 (1992)
6. Kohonen, T.: Self-Organizing Maps. Springer, Berlin (1995)
7. Inoue, H., Narihisa, H.: Improving generalization ability of self-generating neural networks through ensemble averaging. In: Terano, T., Liu, H., Chen, A.L.P. (eds.) PAKDD 2000. LNCS (LNAI), vol. 1805, pp. 177–180. Springer, Heidelberg (2000)
8. Inoue, H., Narihisa, H.: Optimizing a multiple classifier system. In: Ishizuka, M., Sattar, A. (eds.) PRICAI 2002. LNCS (LNAI), vol. 2417, pp. 285–294. Springer, Heidelberg (2002)
9. Stone, M.: Cross-validation: A review. Math. Operationsforsch. Statist., Ser. Statistics 9(1), 127–139 (1978)
10. Breiman, L.: Bagging predictors. Machine Learning 24, 123–140 (1996)
11. Freund, Y., Schapire, R.E.: Boosting: Foundations and Algorithms. MIT Press, Cambridge (2012)
12. Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann, San Mateo (1993)
13. Frank, A., Asuncion, A.: UCI machine learning repository (2010), `http://archive.ics.uci.edu/ml`
14. Chang, C.-C., Lin, C.-J.: LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2, 27:1–27:27 (2011), Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`