Fifth International Workshop on Computational Intelligence & Applications
IEEE SMC Hiroshima Chapter, Hiroshima University, Japan, November 10, 11 & 12, 2009

Applying Cluster Ensemble to Adaptive Tree Structured Clustering

Takashi Yamaguchi[*1], Yuki Noguchi[*1], Takumi Ichimura[*2], and Kenneth J. Mackin[*1]
1) Tokyo University of Information Sciences, Department of Information Systems
4-1Onaridai, Wakaba-ku, Chiba, 265-8501 Japan
Email: mackin@rsch.tuis.ac.jp
2) Hiroshima City University, Graduate School of Information Sciences
3-4-1 Ozuka-Higashi, Asa-Minami-ku, Hiroshima, 731-3194, Japan
Email: ichimura@hiroshima-cu.ac.jp

Abstract — **Adaptive tree structured clustering (ATSC) is our proposed divisive hierarchical clustering method that recursively divides a data set into 2 subsets using self-organizing feature map (SOM). In each partition, the data set is quantized by SOM and the quantized data is divided using agglomerative hierarchical clustering. ATSC can divide data sets regardless of data size in feasible time. On the other hand clustering result stability of ATSC is equally unstable as other divisive hierarchical clustering and partitioned clustering methods.**

In this paper, we apply cluster ensemble for each data partition of ATSC in order to improve stability. Cluster ensemble is a framework for improving partitioned clustering stability. As a result of applying cluster ensemble, ATSC yields unique clustering results that could not be yielded by previous hierarchical clustering methods. This is because a different class distances function is used in each division in ATSC.

I. INTRODUCTION

Recently, huge data can be stored due to the progression of network technology and decreasing cost of mass storage devices. However, these large data has no merit if useful information or knowledge cannot be extracted. Methods for autonomous knowledge extraction have been researched in knowledge discovery in databases, or data mining. Clustering methods are one type of knowledge extraction method that divides data set into some groups based on feature of data without known categories.

Self-organizing feature map (SOM) proposed by Kohonen [1] is an effective clustering method because it can learn regardless of data size and can intuitively show clustering results visually using maps. On the other hand, the clustering result of SOM depends on visual human decision. The boundary of clusters is not clear. Ambiguity of clustering result limits the extensibility of SOM.

In the previous research [2][3], we proposed adaptive tree structured clustering (ATSC) in order to clarify clustering result of SOM. ATSC is divisive hierarchical clustering algorithm (DHCA) that recursively divides a data set into 2 subsets using SOM. In each partition, the data set is quantized by SOM and the quantized data is divided using agglomerative hierarchical clustering algorithm (AHCA). In the previous experiments using an iris data set [2] and a medical data set with large data size [3], we confirmed that ATSC can extract a tree structure that include potential hierarchical relationship without decreasing SOM classification performance, within feasible time. On the other hand clustering result stability of ATSC is equally unstable as other DHCA and partitioned clustering methods.

In this paper, we apply cluster ensembles for each data partition of ATSC in order to improve stability. Cluster ensemble is a framework for improving partitioned clustering stability [4][5][6]. As a result of applying cluster ensemble, ATSC yields unique clustering results that could not be yielded by previous hierarchical clustering methods. This is because a different class distances function is used in each division in ATSC.

II. ADAPTIVE TREE STRUCTURED CLUSTERING

ATSC is framework of DHCA with online processing. Figure 1 shows a model of recursive data division process. ATSC recursively divide the dataset $A = \{ \boldsymbol{x}_i \mid \boldsymbol{x}_i \in \mathbb{R}^n \}$, $i = 1, 2, \ldots, N$, that is given in an ATSC node into K disjoint clusters $\{ A_k \mid A_k \in A, A_k \neq \varnothing, \cup A_k = A, A_k \cap A_{k'} = \varnothing \}$, $k, k' = 1, 2, \ldots, K$, $k \neq k'$ where $K = 2$. The criterion to select the divided cluster is defined by the decrease of variance of A and A_k. When each cluster A does not satisfy the criterion to select the divided cluster, recursive procedure is terminated. These processes can be considered to the Kary tree generation process. In other word the proposed method start with only root node and recursively create K node. The node creation depends on the clustering result in each node. The each node has data subset of parent node data subset. As a result a number of cluster and a tree structure are obtained.

Figure 2 shows data division process of ATSC node. Left model is abstracted ATSC model and right model is SOM+AHCA model discussed on this paper. In each node of the tree structure, there are following 3 steps: (1) Quantization, (2) Clustering, (3) Node Generation. In this paper, we discuss on a SOM+AHCA model that SOM for quantization method and AHCA for clustering method are used.

The previous ATSC model has re-clustering step. The re-clustering step was necessity to verify incorrect classified instance that occurred by dependant on the weight initialization, the order of input vector, and etc. However, previous re-clustering method is cross-sectional processes to current ATSC node and 2 child nodes [2][3]. Considering extensibility to distribution computing in the future and a possibility that these dependants are reduced by applying ensembles, we do not use re-clustering processes in ensemble ATSC model. Similarly, previous SOM training termination method that is cross-sectional processes is modified.

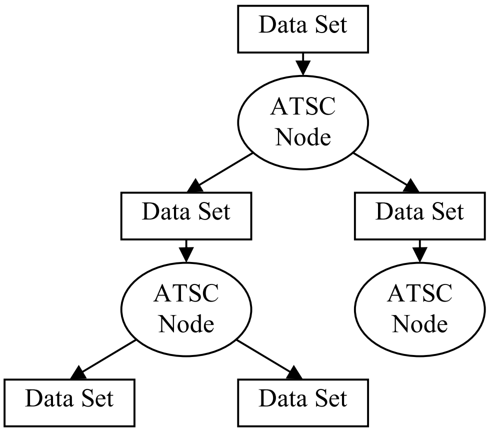

Figure 1. Model of ATSC Tree Structure

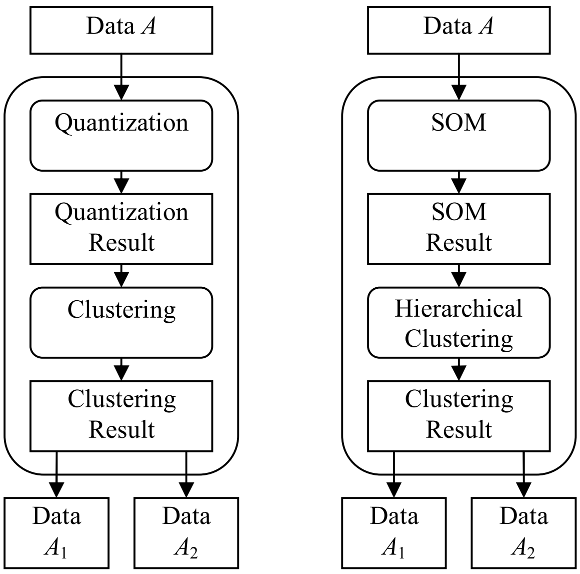

Figure 2. Abstracted Model (Left) and SOM+ HCA Model (Right) of ATSC Node

A. Step of Quantization (SOM Training)

Let the input data set be A, and the weights of the competitive layer where the units are arranged into a 2 dimensional lattice be the set of n dimensional real vectors $W = \{\boldsymbol{w}_j \mid \boldsymbol{w}_j \in \mathbb{R}^n\}$, $j = 1,\ldots, M$. In the step of Quantization, given input data set A is quantized to M codebook vector by using basic online SOM. SOM approximates set of input vectors A by set of weight vectors W, and visualizes the relation between the vectors in the input A through the neighborhood learning.

The value of weight vectors $\boldsymbol{w}_j$ is initialized using by random values. In the SOM training, while repeating the steps of determining the winner unit and updating the weight vector for the selected input vector, the weight vector values converges towards the input vector values.

At the each SOM training time step $t = 1,\ldots, tsmax$, the winner unit c that minimizes the distance between input vector $\boldsymbol{x}_i(ts)$ and jth weight vector $\boldsymbol{w}_j(ts)$ is selected. When Euclidean distance is used, the winner unit c is determined by equation (2).

$$c = \arg\min_j \left\| \boldsymbol{x}_i(ts) - \boldsymbol{w}_j(ts) \right\| \tag{1}$$

For the updating of the weight vectors, the weight vectors of the winner unit and its neighbors on the competitive layer are updated. The weight modification defined as follows:

$$\Delta \boldsymbol{w}_j(ts+1) = h_{cj}(ts) \cdot \left(\boldsymbol{x}_i(ts) - \boldsymbol{w}_j(ts) \right) \tag{2}$$

where $h_{cj}(ts)$ is the neighborhood function. The Gaussian type neighborhood function is defined as follows:

$$h_{cj} = A(ts) \cdot \exp\left(-\frac{\left\| r_c - r_j \right\|^2}{2\sigma^2(ts)} \right) \tag{3}$$

where $A(ts)$ is learning-rate factor, $\| r_j - r_c \|$ is distance between winner unit c and unit j in coordinates of the competitive layer, $\sigma^2(ts)$ is a parameter that define the width of updating. $A(ts)$ and $\sigma^2(ts)$ are monotonic decreasing parameters for training step ts.

In ATSC, detailed learning is carried over to child node SOMs. Thus SOM training is terminated at early training time step by using learning error. At the training step ts, let e be the error between input vector $\boldsymbol{x}_i$ and weight vector $\boldsymbol{w}_{ci}$ of the corresponding winner unit ci. The learning error e is show as equation (4).

$$e(ts) = \sum_{i=1}^{N} \left\| \boldsymbol{x}_i - \boldsymbol{w}_{ci} \right\| \tag{4}$$

The average changes in learning error between training step ts to $ts + \tau$ calculate by equation (5):

$$\Delta e(ts) = \int_{ts}^{ts+\tau} \lim_{\Delta ts \to 0} \frac{e(ts - \Delta ts) - e(ts)}{\Delta ts} dt \quad (5)$$

where Δts is the length of training steps to sample the learning error $e(ts)$, and τ defines the length of training steps to calculate average. For the average change in learning error Δe, we considered approximate equation (6) because of the calculation cost of e and Δe is very high.

$$\Delta e(ts) = \int_{ts}^{ts+\tau} \frac{e(ts - \Delta ts) - e(ts)}{\Delta ts} dt, (\Delta ts < \tau) \quad (6)$$

The training of SOM is terminated when expression (7) is satisfied:

$$\Delta e < \varphi \quad (7)$$

where φ is thresholds of the SOM training termination criterion that is a monotonic decreasing function.

B. Step of Clustering (AHCA with SOM Result)

After the SOM training, dataset A is divided into 2 subsets A_k based on the SOM training result. When the SOM training converges, winner unit c and units that have close weight vector values with weight vector values of the winner unit c, forms a Voronoi cell on the map. The relative location of the winner units on the map shows the relationship between the input vectors. Therefore clustering using SOM can be decided by the weight vector values and neighbor information of each winner unit.

When the set of winner units is $C = \{ \boldsymbol{w}_{ci} \}$, $ci = 1,\ldots, cimax$ and the set of disjoint subset of C is $B = \{B_k \mid B_k \in C, B_k \neq \varnothing, \cup B = C, B_k \cap B_{k'} = \varnothing,\}$, $k, k' = 1,\ldots, |B|$, $k \neq k'$; the clusters B_k is recursively merged using AHCA until $|B| = K$ start with each element as separate cluster. Final 2 subset A_k is obtained from equation (1) and B_k where $k = K$. In the process of merging winner units, when the set of winner units in the neighbor of winner unit ci is N_{ci}, the cpth and cqth merged winner unit satisfies expression (8)

$$cq \in N_{cp}, cp \neq cq \quad (8)$$

where $cp, cq = 1,\ldots, cimax$, $cp \neq cq$.

For distance function $d(B_p, B_q)$, we use following equations based on single linkage method(9), complete linkage method(10), group average method(11), and Ward's method(12).

$$d(B_p, B_q) = \min_{\boldsymbol{w}_{cp} \in B_p, \boldsymbol{w}_{cq} \in B_q} \|\boldsymbol{w}_{cp} - \boldsymbol{w}_{cq}\| \quad (9)$$

$$d(B_p, B_q) = \max_{\boldsymbol{w}_{cp} \in B_p, \boldsymbol{w}_{cq} \in B_q} \|\boldsymbol{w}_{cp} - \boldsymbol{w}_{cq}\| \quad (10)$$

$$d(B_p, B_q) = \frac{1}{|B_p||B_q|} \sum_{\boldsymbol{w}_{cp} \in B_p} \sum_{\boldsymbol{w}_{cq} \in B_q} \|\boldsymbol{w}_{cp} - \boldsymbol{w}_{cq}\| \quad (11)$$

$$d(B_p, B_q) = QE(B_p \cup B_q) - QE(B_p) - QE(B_q) \quad (12)$$

$$QE(B_k) = \sum_{\boldsymbol{w}_{cp} \in B_k} \|\overline{\boldsymbol{w}} - \boldsymbol{w}_{cp}\| \quad (13)$$

where $\boldsymbol{w}$ is centroid of B_k. For updating distance $d(B_p, B_q)$ when merged cluster is merged, Lance-Williams update formula [7] is used. Lance-Williams update formula is defined as follow:

$$d(B_k, B_p \cup B_q) = \alpha_1 \cdot d(B_k, B_p) + \alpha_2 \cdot d(B_k, B_q) + \beta \cdot d(B_p, B_q) + \gamma \cdot (d(B_k, B_p) - d(B_k, B_q)) \quad (14)$$

where $k \neq p, q$; and α_1, α_2, β and γ are coefficients of each AHCA method defined as Table I.

TABLE I. PARAMETER OF LANCE WILLIAMS UPDATE FUNCTION

method	α_1	α_1	β	Γ
Single linkage method	0.5	0.5	0	-0.5
Complete linkage method	0.5	0.5	0	0.5
Group average method	$\frac{\lvert B_p\rvert}{\lvert B_p\rvert+\lvert B_q\rvert}$	$\frac{\lvert B_p\rvert}{\lvert B_p\rvert+\lvert B_q\rvert}$	0	0
Ward's method	$\frac{\lvert B_k\rvert+\lvert B_p\rvert}{\lvert B_k\rvert+\lvert B_p\rvert+\lvert B_q\rvert}$	$\frac{\lvert B_k\rvert+\lvert B_q\rvert}{\lvert B_k\rvert+\lvert B_p\rvert+\lvert B_q\rvert}$	$-\frac{\lvert B_k\rvert}{\lvert B_k\rvert+\lvert B_p\rvert+\lvert B_q\rvert}$	0

C. Step of Node Generation

When the decreasing error ΔE is larger than threshold θ, 2 new child nodes are created.

$$\Delta E(A) > \theta \cdot E(S) \quad (15)$$

Where S is dataset input into root ATSC node and A is dataset input into current ATSC node. For the error of cluster E, the quantization error is used. Let the quantization error of cluster before division be $E(A_1 \cup A_2)$, and the quantization error of clusters after division be $E(A_1)$ and $E(A_2)$. ΔE and the quantization error are defined as follows:

$$\Delta E = E(A_1 \cup A_2) - E(A_1) - E(A_2) \quad (16)$$

$$E(A) = \sum_{\boldsymbol{w}_{cp} \in A} \|\overline{\boldsymbol{w}} - \boldsymbol{w}_{cp}\| \quad (17)$$

where $\overline{\boldsymbol{w}}$ is centroid of A.

III. CLUSTER ENSEMBLE

Cluster Ensemble is a framework for building a robust clustering from combining different clustering results given by individual clustering algorithms. Figure 3 shows a basic model of cluster ensemble framework proposed by A. Strehl et al [4]. In the basic cluster ensemble, input data set $X = \{x_1, x_2, \ldots, x_N\}$ is partitioned into r sets of k clusters $C = \{ C_l^s \}$, $l = 1, 2, \ldots, k$; $s = 1, 2, \ldots, r$; from individual clustering algorithms $\Phi = \{ \Phi_s \}$ where r is number of ensembles. As a result, for the clustering result, r label vectors $\{ \lambda^s \mid \lambda^s \in \mathbb{N}^N \}$ are yielded. Finally, the clustering results $\{ \lambda^s \}$ is combined into a single clustering result λ using a consensus function Γ.

Clustering Ensemble is a very loose or abstract framework. There are various methods to build a cluster ensemble. Typical cluster ensemble methods are follows; feature-distributed clustering that use different subsets of features, object-distributed clustering that use different subsets of input data set, heterogeneous ensembles that use different clustering algorithms in each ensemble, homogeneous ensembles that use same clustering algorithms and different parameters. Proposed cluster ensemble method is extended method of heterogeneous ensembles for application to ATSC.

Cluster Ensemble bring stability to dependency of input data set and parameters for clustering methods, robustness to noise and outlier, novelty to final clustering result and independency of cluster combining process on distribution computing.

Most important problem of ATSC or other DHCA is that the errors of earlier clustering results occur bad influence to partitions on subsets. The Appling cluster ensemble to ATSC is expected to improve stability of earlier clustering results. As a result, the stability of total clustering result and a generated tree structure is expected to be improved.

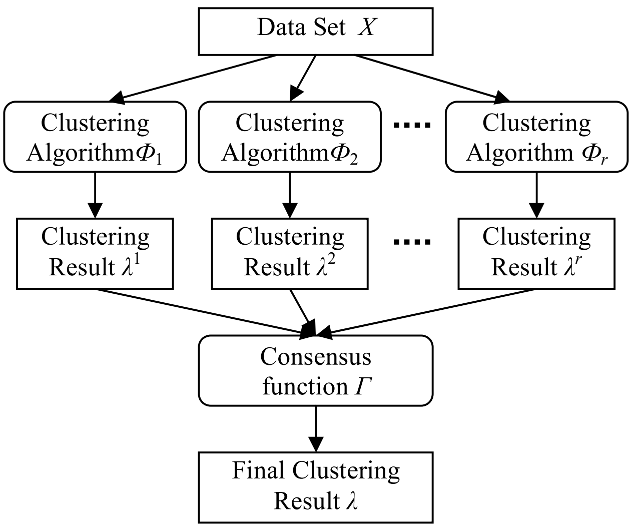

Figure 3. Basic Cluster Ensemble Framework

IV. APRICATION OF CLUSTER ENSEMBLE

In this paper, we applied cluster ensemble on each ATSC node that shown in Figure 4. Each ATSC has r ensembles. Each ensemble has different SOM initialization and different data input order such as homogeneous cluster ensemble. In addition, Each ensemble has different class distance functions $d(B_p, B_q)$ on AHC such as heterogeneous cluster ensemble. r ensembles are consisted from ht distance function types of hm ensembles each such as Figure 5. ht and hm are parameter that define number of ensemble.

For consensus function, decrease of the quantization error is used. When ΔE^s is decrease of the quantization error of sth ensemble, bth ensemble which is satisfy expression (18) is selected.

$$\Gamma(C) = C^b \tag{18}$$

$$b = \arg\max_s \Delta E^s \tag{19}$$

Finally, clustering result of bth ensemble is used for final clustering result of ATSC node. In general clustering ensemble framework, similarity between input objects is not used for consensus function in order to certify independency of input data set and clustering result when input data set is deferent. However, in this ensemble ATSC node model, it is not necessary to consider because same input data set is used in each ensemble.

The Applying cluster ensemble to ATSC is expected to improve stability of clustering result and a generated tree structure. In addition, ensemble ATSC yields original clustering result and a generated tree structure. In non-ensemble ATSC, a generated tree structure is depends on class distance function $d(B_p, B_q)$ [2]. In ensemble ATSC, each ATSC node yields different class distance based partition. It is important advantage compared with non-ensemble ATSC and AHCA.

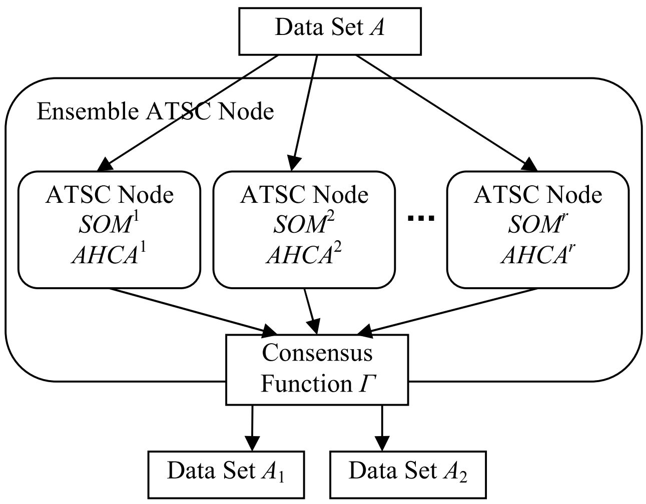

Figure 4. Heterogeneous Ensembles Model on ATSC Node (Ensemble 2)

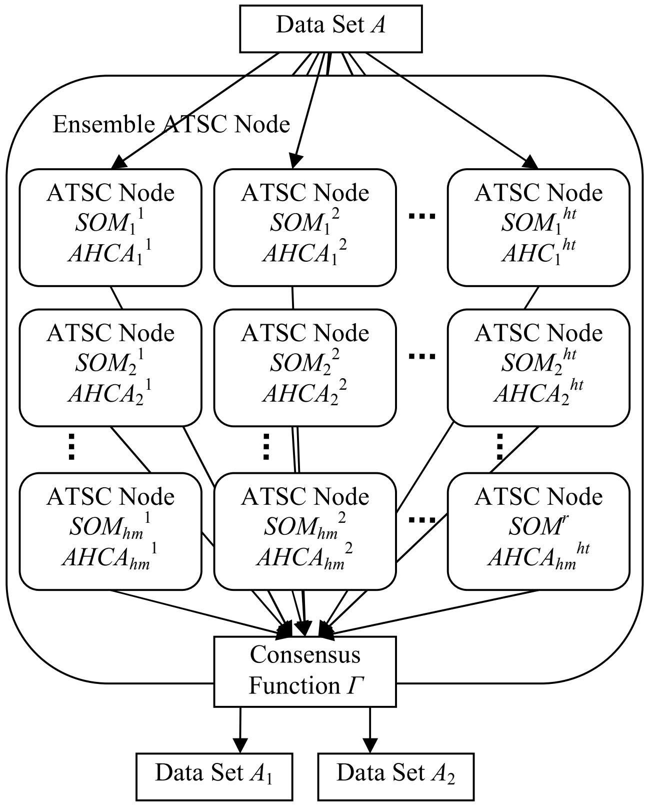

Figure 5. Homogenious and Heterogeneous Ensembles Model on ATSC Node (Ensemble 1)

V. EXPERIMENT

For evaluating the classification performance, we applied the proposed ensemble ATSC to following 3 data sets; iris, wine and coronary heart disease database (CHD_DB). Iris and wine data set are most popular benchmark data set provided by UCI Machine Learning Repository [8]. In addition to these benchmark data set, CHD_DB developed by Suka et al [9] was used in order to evaluate the classification performance on real data set with large instances.

The descriptions of these data are shown in Table II. Iris data set contains 3 classes of 50 instances each. Wine data set contains 3 classes with different class ratio. CHD_DB has 4 training data sets and 1 testing data set. In this experiment, Train_A of training data set was used.

In all data set, each value was normalized so that each item has same average and variance.

TABLE II. DATA SETS

Dataset Name	Classes	Instances	Ratio	Values *(Continuous / Desecrate)*
Iris	3	150	1:1:1	4 / 0
Wine	3	177	59:71:48	13 / 0
CHD_DB Train A	2	13000	1:1	4 / 4

In this experiment, we compared the 2 proposed ensemble ATSC methods and 4 non-ensemble ATSC methods shown in Table III. Ensemble 1 and Ensemble 2 is proposed ensemble ATSC that has 4 different distance function ensembles. Single Linkage, Complete Linkage, Group Average, and Ward is used. non-ensemble ATSC that used corresponding single distance function. Ensemble 1 is model that shown in Figure 5. Ensemble 2 is model that shown in Figure 4.

For the Classification performance, classification accuracy and number of cluster were investigated. The classification accuracy was derived by correctly classification rate using class labeling. By the class labeling, each cluster was labeled by teaching class label that has maximum frequency.

TABLE III. COMPARED CONDITIONS OF ATSC

Condition Name	*r*	*hm*	*ht*	Distance Function
Ensemble 1	40	10	4	4 Distance Functions
Ensemble 2	4	1	4	4 Distance Functions
Single Linkage	1	1	1	Single Linkage
Complete Linkage	1	1	1	Complete Linkage
Group Average	1	1	1	Group Average
Ward	1	1	1	Ward

For ATSC parameters, following parameters were used. For θ = 0.01, φ = 0.01 was used. For the steps of SOM training the following SOM was used. A 9x6 competitive layer was initialized using random values. For the determining of the winner unit we use Euclidian distance. For the neighborhood function h_{cj} we use the Gaussian type. The following parameters were used, Δts=10, τ=100, $\Lambda(0)$=0.1, Λ'=0.9995, $\sigma(0)$=9, σ'=0.999 and H=1 was used where

$$\Lambda(ts) = \Lambda(0) \cdot \Lambda'^{ts}, \tag{21}$$

$$\sigma(ts) = H + (\sigma(0) - H) \cdot \sigma'^{ts}. \tag{22}$$

Table IV is result on iris data set and Table V is result on wine data set. These results show the comparison of classification accuracy and number of cluster. For evaluating the accuracy and stability, the average and variance are computed from the result of 10 times separate runs. The variance of number of cluster could be used for stability of given cluster and a tree structure.

Ensemble 1 was gave best result in average and variance on iris and wine data sets shown in Table IV and V. Specially, very low variance was yield compared with other conditions. On the average of classification accuracy, better result was yield than our expectation. This result shows proposed ensemble method was effective for improving stability and accuracy of ATSC.

Table VI shows result on CHD_DB that has larger instance than previous 2 data sets. This result shows that classification performance was improved by using proposed ensemble method same as previous result. However that improvement was felt that smaller than previous 2 result.

Ensemble 2 was evaluated for investigating the effectiveness of heterogeneous ensemble. However, it was not confirmed from the result of classification accuracy and number of cluster. This was caused by the instability of ATSC and the insufficient of runs. The examination of Ensemble 2 is necessary to reexamine with more large times of runs.

TABLE IV. CLASSIFICATION PERFORMANCE OF IRIS DATA SET

	Classification Accuracy			Number of Cluster	
	Ave.	*Var.*	*Best*	*Ave.*	*Var.*
Ensemble 1	0.934	0.000637	0.967	12.0	0.00
Ensemble 2	0.887	0.000948	0.933	11.0	1.11
Single Linkage	0.600	0.054983	0.853	4.8	15.07
Complete Linkage	0.872	0.001983	0.940	9.9	1.43
Group Average	0.877	0.002866	0.953	9.4	2.48
Ward	0.880	0.002331	0.940	10.8	1.96

TABLE V. CLASSIFICATION PERFORMANCE OF WINE DATA SET

	Classification Accuracy			Number of Cluster	
	Ave.	*Var.*	*Best*	*Ave.*	*Var.*
Ensemble 1	0.953	0.000321	0.977	13.8	2.18
Ensemble 2	0.941	0.001037	0.977	14.0	3.56
Single Linkage	0.583	0.046892	0.938	2.9	3.88
Complete Linkage	0.929	0.000704	0.960	14.6	2.71
Group Average	0.926	0.001043	0.960	14.4	4.27
Ward	0.908	0.000552	0.940	13.0	2.00

TABLE VI. CLASSIFICATION PERFORMANCE OF CHD_DB

	Classification Accuracy			Number of Cluster	
	Ave.	*Var.*	*Best*	*Ave.*	*Var.*
Ensemble 1	0.654	0.000027	0.666	30.4	3.37
Ensemble 2	0.653	0.000144	0.667	29.6	6.04
Single Linkage	0.582	0.003159	0.667	9.7	46.67
Complete Linkage	0.65	0.000116	0.667	29.6	11.37
Group Average	0.65	0.000088	0.664	31.0	14.44
Ward	0.651	0.000102	0.667	28.0	16.89

VI. CONCLUSION

In this paper, we proposed an ensemble ATSC method in which cluster ensemble is applied to each ATSC node. For evaluating classification performance, proposed ensemble ATSC was applied to 3 different data sets. As a result, proposed ensemble ATSC method yielded best accuracy and stability in all data sets. This result shows proposed ensemble method was effective for improving stability and accuracy of ATSC. The bad influence of instability in earlier partition is fundamental problem of ATSC or other DHCA. In the case of other DHCA, Ensemble technique is expected to improve stability similar to the case of ATSC.

In the result of Iris data set shown in Table IV, Ensemble 1 divides data set into always 12 clusters. This result shows high effectiveness for stabilization of ATSC. On the other hand, it is difficult to discuss on appropriate cluster size for iris data set. In current ATSC, because of the cluster size is depends on threshold θ, it was not clear by which cluster size is determined. For giving a validity of cluster size, we plan to apply Akaike's information criterion (AIC) [10] to node generation criterion.

Our ensemble ATSC is expected to yields unique clustering result and a generated tree structure. This is an important advantage compared with non-ensemble ATSC or AHCA. These unique results are confirmed in the process of these experiments. However the quantitative evaluations on novelty and effectiveness were not examined. For the future works, we plan to investigate the evaluation methods for clustering result and a tree structure of ATSC.

REFERENCES

[1] T. Kohonen,Self-organizing maps, Berlin, Springer, 1995

[2] Takashi Yamaguchi, Takumi Ichimura, Kenneth J. Mackin, Adaptive Tree Structured Clustering Method using Self-Organizing Map, Joint 4th International Conference on Soft Computing and Intelligent Systems and 9th International Symposium on advanced Intelligent Systems, 2008

[3] Takashi Yamaguchi, Takumi Ichimura, Kenneth J. Mackin, Analysis using adaptive tree structured clustering method for medical data of patients with coronary heart disease, The 4th International Workshop on Computational Intelligence and Applications, 2008

[4] A. Strehl, J. Ghosh, Cluster ensembles - a knowledge reuse framework for combining multiple partitions, Journal of Machine Learning Research 3, 583-618, 2002

[5] A. Strehl, J. Ghosh, Cluster ensembles - a knowledge reuse framework for combining partitionings, in: Proc. of 11th National Conf. on Artificial Intelligence, NCAI, Edmonton, Alberta, Canada, 2002, pp. 93-98.

[6] X. Hu, I. Yoo, Cluster ensemble and its applications in gene expression analysis, in: Y.-P.P. Chen (Ed.), Proc. 2nd Asia Pacific Bioinformatics Conference, Dunedin, New Zealand, 297-302, 2004

[7] G.N.Lance and W.T.Williams, A general theory of classificatory sorting strategies, Computer journal, vol.9 p.373-380 1967

[8] UCI Machine Learning Repository, http://archive.ics.uci.edu/ml/

[9] Machi Suka, Takumi Ichimura, Katsumi Yoshida, Development of Coronary Heart Disease Databases, Proc. of the 8th International Conference on Knowledge-Based Intelligent Information & Engineering Systems, Vol.2, 1081-1088, 2004

[10] H. Akaike, A new look at the statistical model identification. IEEE Transactions on Automatic Control 19 (6) 716-723, 1974