HIERARCHICAL CLUSTER ANALYSIS by arboART NEURAL NETWORKS and its APPLICATION to KANSEI EVALUATION DATA ANALYSIS

Shigekazu Ishihara[1], Keiko Ishihara[2], and Mitsuo Nagamachi[1]

Department of Kansei Ergonomics[1]
Department of Language and Communication[2]
College of Human and Social Environment, Hiroshima International University
555-36, Kurose-cho, Kamo-gun, Hiroshima, 724-0695 Japan

ABSTRACT
ART (Adaptive Resonance Theory [1]) neural network and its variations perform non-hierarchical clustering by unsupervised learning. We propose a scheme "arboART" for hierarchical clustering by using several ART1.5-SSS networks. It classifies multidimensional vectors as a cluster tree, and finds features of clusters. The Basic idea of arboART is to use the prototype formed in an ART network as an input to other ART network that has looser distance criteria (Ishihara, et al., [2,3]). By sending prototype vectors made by ART to one after another, many small categories are combined into larger and more generalized categories. We can draw a dendrogram using classification records of sample and categories. We have confirmed its ability using standard test data commonly used in pattern recognition community. The clustering result is better than traditional computing methods, on separation of outliers, smaller error (diameter) of clusters and causes no chaining. This methodology is applied to Kansei evaluation experiment data analysis.

Kansei Engineering; Multivariate Analysis, Hierarchical Clustering; Neural Network, Adaptive Resonance Theory

1. Introduction

The aim of this work is to develop ART based hierarchical clustering scheme and apply it to analyzing multivariate data obtained from Kansei engineering experiment.

The Japanese word Kansei has a meaning of feeling, sensitivity and some kind of emotion. Kansei Engineering is defined as "a translation system for images or feelings into real design components". It is used as a method to convert a customer's ambiguous image of product into a detailed design. It supports product designers by providing relation among customers' feeling and corresponding design. It also assists consumer to select a product that fits his/her feeling, among a variety of products. The standard procedure of Kansei engineering consists of an assessment and statistical analyses, these are (1) selection of adjective words, (2) evaluation of the design components using semantic differential questionnaire and (3) multivariate analysis of evaluated data. Multivariate analysis is used for disclose the implicit relations among adjectives and products or physical attributes (i.e., color, size) of each design component. Then finally (4) conversion of obtained relations among a component's design, features and structure and the adjectives into rules that a computer uses for reasoning. Our Kansei engineering expert systems (Kansei ES) can draw designs that correspond to user's Kansei expressed by adjectives. We have developed various Kansei ES such as one for house interior design, car interior design, garment coordination, construction machine and other product designs [4,5]. Recently, to support designers by on-the-site analysis of Kansei experiment and automatic extraction of inference rules and automatic generation of ES, we are working on neural network based analyzer and ES builder [6-8]. The work described in this paper develops ART based hierarchical clustering scheme and feature selection.

2. arboART

ART-type neural networks are based on the Adaptive Resonance Theory by Grossberg and Carpenter [1]. These networks perform non-hierarchical clustering by unsupervised learning. On the other hand, hierarchical clustering is the most popular technique for explicit classification by measured multivariate vectors.

In this paper, we propose a scheme for hierarchical clustering by using several ART1.5-SSS networks. We call this method as arboART. It can classify multidimensional vectors as a cluster tree, and can find features of clusters. In the following section, we describe the scheme of ART-based hierarchical clustering and rule extracting. Finally, analysis on dishwasher design is discussed.

We use ART1.5-SSS as a building block of arboART. ART1.5-SSS is our modified version of ART1.5 [9]. ART1.5-SSS has two layers of units and a reset mechanism. By adding reset mechanism to simple competitive learning network, ART can perform stable clustering. Each unit of the F2 represents a category. The F2 units are interconnected by inhibitory links, and unit of them competes each other. The two layers of units are linked by bottom-up (F1 -> F2) and top-down (F2->F1) adaptive connections.

The F1 layer receives the input signal and sends it to every unit of F2 layer.

x_i is the activity of i th F1 unit, and T_j is activity of j th F2 unit. z_{ji} is the bottom-up synapse weight from i th F1 unit to the j th F2 unit.

$$T_j = \sum_i z_{ji} x_i \tag{1}$$

Through a competition process, the F2 unit that receives maximum input signals multiplied by bottom up synapse weights is activated. It is called winner unit.

$$T_J = \begin{cases} T_j & : if\ T_j = \max_j \{T_j\} \\ 0 & : \text{otherwise} \end{cases} \tag{2}$$

The reset mechanism has distance criteria, and it sends a reset signal to the F2 layer when the input vector and a prototype stored in the top-down weight vector from the winner unit are not similar. Then, the next to the maximum unit is tested again. **x** is the vector of F1 activities, and $\mathbf{z}_j$ is the vector of top-down weights from the jth F2 node. When the following equation is satisfied, a match occurs. Otherwise, a reset signal is sent, and a search occurs. r is a constant of angle threshold.

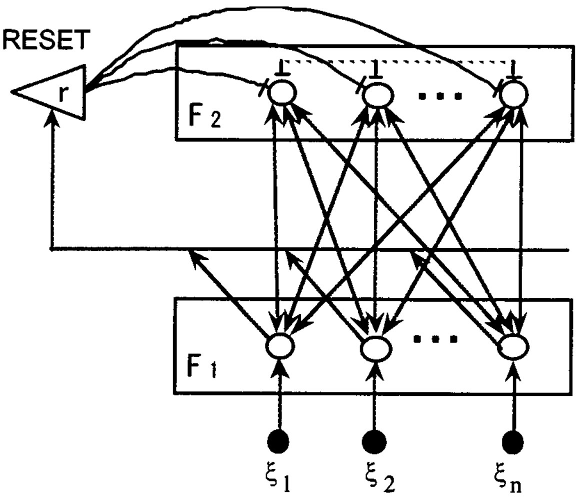
$$\frac{(\mathbf{x} \cdot \mathbf{z}_J)}{|\mathbf{x}|\,|\mathbf{z}_J|} > r \tag{3}$$

If all F2 units that are used for category failed the test, an unused unit is chosen for the new category. Connections between the chosen unit in F2 and the units in F1 are modified so that they become slightly similar to the input vector. It is called learning. Each top-down and bottom-up synapse weights vector of a category (each F2 unit) can be then interpreted as prototype vector of input vectors that belong to the category. Our modification of ART1.5-SSS is in its learning rule. "SSS" is stand for small sample size. We introduced decay factor to learning rate.

$$\frac{d}{dt} z_{Ji} = \frac{1}{q_J}(x_i - z_{Ji})$$
$$\frac{d}{dt} z_{iJ} = \frac{1}{q_J}(x_i - z_{iJ}) \tag{4}$$

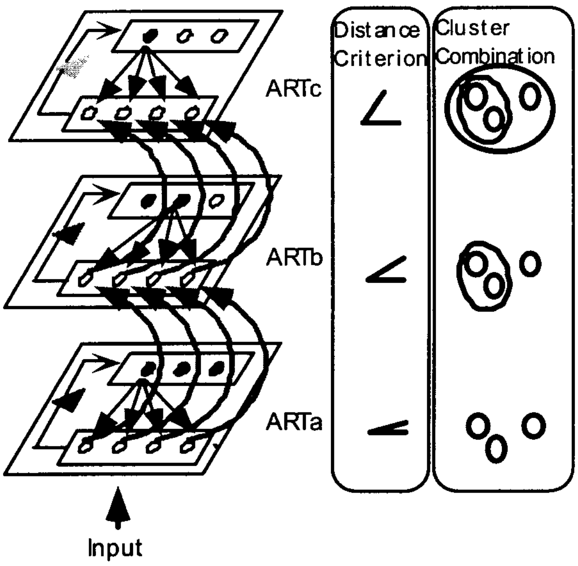
(q is the number of Jth F2 unit was chosen)

In Ishihara et al. [7,8], we have shown this modification enables right classification.



The basic idea of arboART is using the prototype formed in an ART network as an input to other ART network that has looser distance criteria [2,3]. By sending prototype vectors made by ART to one after another, many small categories are combined into larger and more generalized categories. Fig.2 shows an example of arboART of 3 ART1.5-SSS.

We can draw dendrogram using classification records of sample and categories. Our arboART has several advantages over conventional methods. On the computation side, conventional methods have to retain a similarity matrix of samples, which size is at least as ($\text{sample\_size}^2/2$). arboART uses several matrices those sizes are number of input vector dimension. Huge memory is not required. On the analytic side, cluster prototype of ART1.5-SSS provides features of clusters. On Kansei evaluation data, it provides specific relations between product and Kansei. In the last section of this paper, we show analyzing result by arboART which exceeds conventional methods.



3. ANALYSIS RESULTS

3.1 Kansei experiment conditions

We conducted an experiment for evaluating imported beer can samples on a 5-point SD scale. Eighty-five cans were collected, and are used for evaluation experiment using questionnaire.

Eight subjects are participated. All of subjects were women students of college, and at age 19 or 20. They have less or no regular drinking habit, thus, imported beer is unfamiliar for them. Ninety-two adjective words (Kansei words) were used for evaluation with the SD scale. The words proper to beer are chosen from frequently appeared words on beer guidebooks, brewing books and web pages of brewery. Other words are our standard Kansei words for doing experiments.

Averages of evaluations between subjects were used as an input vector on a sample. Each element of the input vector corresponds to an evaluation of each adjective word, so the input vector has 92 elements, and there are 85 input vectors.

3.2 Clustering beer cans

Figure 3 shows the nesting of clusters made by 1st to 7th ART1.5-SSS. At 7th stage, 4 clusters and 3 singleton clusters (one-member cluster) are created. Large valued elements of prototype vector of each cluster are written.

Cluster1 has larger values on *light taste*, *crisp*, *lite* and *refreshing taste*. Relations between design elements and Kansei are also recognized. Many members of can cluster 1 are painted on white, metallic silver or pale colors. Brand names are written in horizontal line. Blue metallic and silver metallic cans form small subgroup and corresponding K-words are *chill* and *cool*. White subgroup is corresponding to *drinkable* and *refreshing*. Members of can cluster 2 is dark color and/or traditional oval shape label design and using Serif font for brand name. They correspond to *aromatic*, *calm* and *bitter*. All members of can cluster 3 has animal (bear, ox, wolf and cobra) or man illustration on it. They correspond to *masculine*, *hard*, *individual* and *full bodied*. Can cluster 4 has sub-cluster that contains metallic red or gold cans, and traditionally designed. *High-grade*, *aromatic*, *affected* and *showy* are high on this cluster.

Design elements of color, label and brand name, and illustration are seems to be related with cluster discrimination.

4. Discussion on Clustering Ability of the ArboART

4.1 Required ability for the clustering analysis

Following 4 abilities are generally required for the cluster analysis:

1. Separating multidimensional outliers: For Kansei data, individual design samples often get different evaluation from others on many Kansei words. Those samples have to be separated as another clusters.
2. Small errors within a cluster: More closely similar samples have to be classified into a group for making the same number of groups.
3. Suppression of making large meaningless clusters called "chaining cluster".
4. Making correct clusters for small sample size case: Most of Kansei evaluation data have large dimension, from 40 to 80. Relatively, sample sizes are small for the cases. This is the considerable issue especially in incremental non-hierarchical clustering.

4.2 Analysis on Kansei evaluation data for beer cans

The Kansei evaluation data for imported or foreign beer cans are analysed here. We discuss how our arboART and conventional clustering methods represent the structure of multidimensional data.

At first, we examine separation of multivariate outlier. We used Kaufman & Rousseeuw method [10]. K&R method calculates distance of between samples and finds the nearest neighbor. If a sample has enormous distance from nearest neighbor, it regarded as outlier.

The most outstanding one is Budweiser and the next is Hite. These cans have distance far more than 2SD of distance distribution. These should be classified as each singleton clusters. More than 1SD samples should not be grouped at early stage of cluster combining.

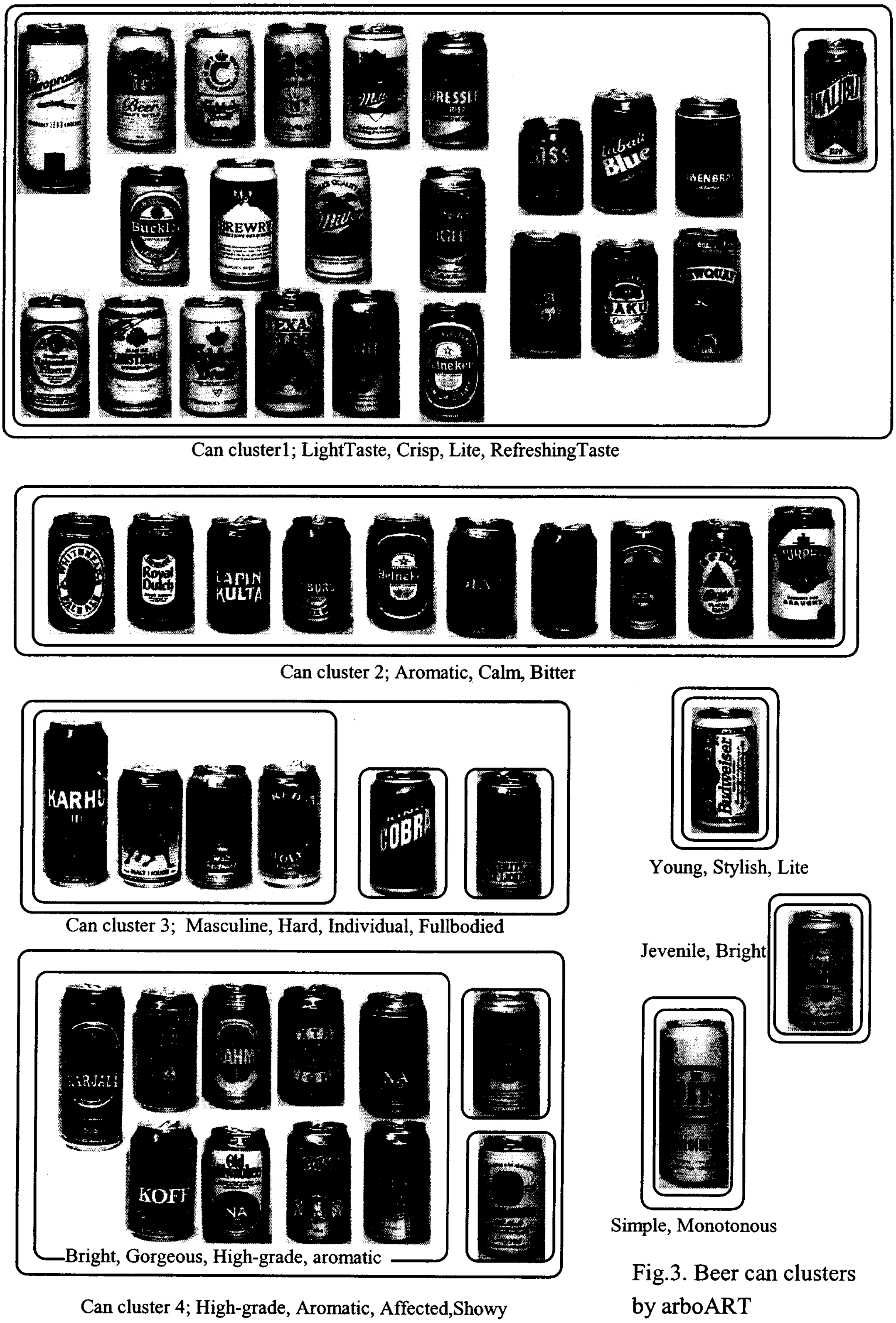


Fig.3. Beer can clusters by arboART

Finally we decided outliers with regarding Multidimensional Scaling (MDS) result. Budweiser, Hite, Malibu, Young's London Lagar and Cobra are identified outliers to be considered.

To examine, we cut dendrogram at the 14 clusters (1/8 the number of samples) and 7 clusters (1/4 of sample number). Budweiser, Hite and Young's London Lagar are singleton cluster at 7 clusters solution. At 14 clusters, Malibu is singleton. Malibu is included in other cluster at 8 clusters solution. ArboART can separate outliers well. Traditional computation methods are examined with the same procedure. Weighted Centroid and Centroid method are failed to separate even in 7 clusters solution. WPGMA (Weighted Group Average), UPGMA (Unweighted Group Average), CLINK (Complete Linkage) and SLINK (Single Linkage) can separate outliers as well as arboART.

We also compare clustering ability using within cluster error. Within cluster error is a sum of each sample's distance from the center of the cluster. The center of the cluster is defined as an average of cluster members in *n* dimension space. As shown in table 1, arboART is next to smallest error. This result agrees with verification with standard test data "Heart", which is defined StatLog project [3].

Although SLINK has smallest within cluster error, it has the serious deficit of "Chaining cluster". Figure 4 shows 2 dimensional mapping by MDS and 7 and 14 clusters (shown in circles). In same number of clusters, arboART made meaningful clusters. SLINK made huge inclusive cluster, then its result is meaningless.

From above verifications, we conclude that arboART most well satisfied 4 conditions for clustering.

Table 1. Comparison of clustering preciseness

	Outlier separation	sum of "within cluster error"	No Chaining
arboART	Good	2.9845	Good
WPGMA	Good	3.4621	Good
UPGMA	Good	3.3132	Good
CLINK	Good	3.4666	Good
W-centroid	NG	5.8173	Good
Centroid	NG	5.8173	Good
SLINK	Good	1.8918	NG

REFERENCES

[1] Carpenter, G.A. & Grossberg, S. (1987). ART2: Self-organization of stable category recognition codes for analog input patterns, Applied Optics, 26, 4919-4930.

[2] Ishihara, S., Nagamachi, M. & Ishihara, K. (1995). arboART: ART Based Hierarchical Clustering and Its Application to Questionnaire Data Analysis, Proc. of 1995 IEEE Int. Conf. on Neural Networks, Perth, pp. 532-537.

[3] Ishihara, S. Ishihara, K, & Nagamachi, M. (1999). Analyzing method for Kansei engineering data by hierarchical clustering using self-organizing neural networks, The Transactions of the institute of Electronics, Information and Communication Engineers A, vol. J82-A, no.1, 179-189. (in Japanese)

[4] Nagamachi, M. (1991). An Image Technology Expert System and Its Application to Design Consultation, Int. J. of Human-Computer Interaction, vol. 3, pp. 267-279.

[5] Nagamachi, M. (1995). Kansei engineering: a new ergonomic consumer-oriented technology for product development, Int. J. of Industrial Ergonomics, vol.15, no.1, pp. 3-11.

[6] Ishihara, S., Hatamoto, K., Nagamachi, M. and Matsubara, Y. (1993). ART 1.5-SSS for Kansei engineering expert system, Proc. of 1993 International Joint Conference on Neural Networks-Nagoya, pp.2512-2515.

[7] Ishihara, S., Ishihara, K., Nagamachi, M. and Matsubara, Y. (1995). An automatic builder for a Kansei engineering expert system using self-organizing neural networks, Int. J. of Industrial Ergonomics, vol.15, no.1, pp. 13-24.

[8] Ishihara, S., Ishihara, K., Nagamachi, M. and Matsubara, Y. (1997). An Analysis of Kansei Structure on Shoes Using Self-Organizing Neural Networks, Int. J. of Industrial Ergonomics, vol.19, no.2, pp.93-104.

[9] Levine, D.S. & Penz, P.A. (1990). ART 1.5 – A simplified adaptive resonance network for classifying low-dimensional analog data, *Proc. of 1990 Int. Joint Conf. on Neural Networks-WASH DC*, Vol.2, 639-642.

[10] Kaufmann, L & Rousseeuw, P.J. (1990), Finding groups in data, Wiley Interscience.

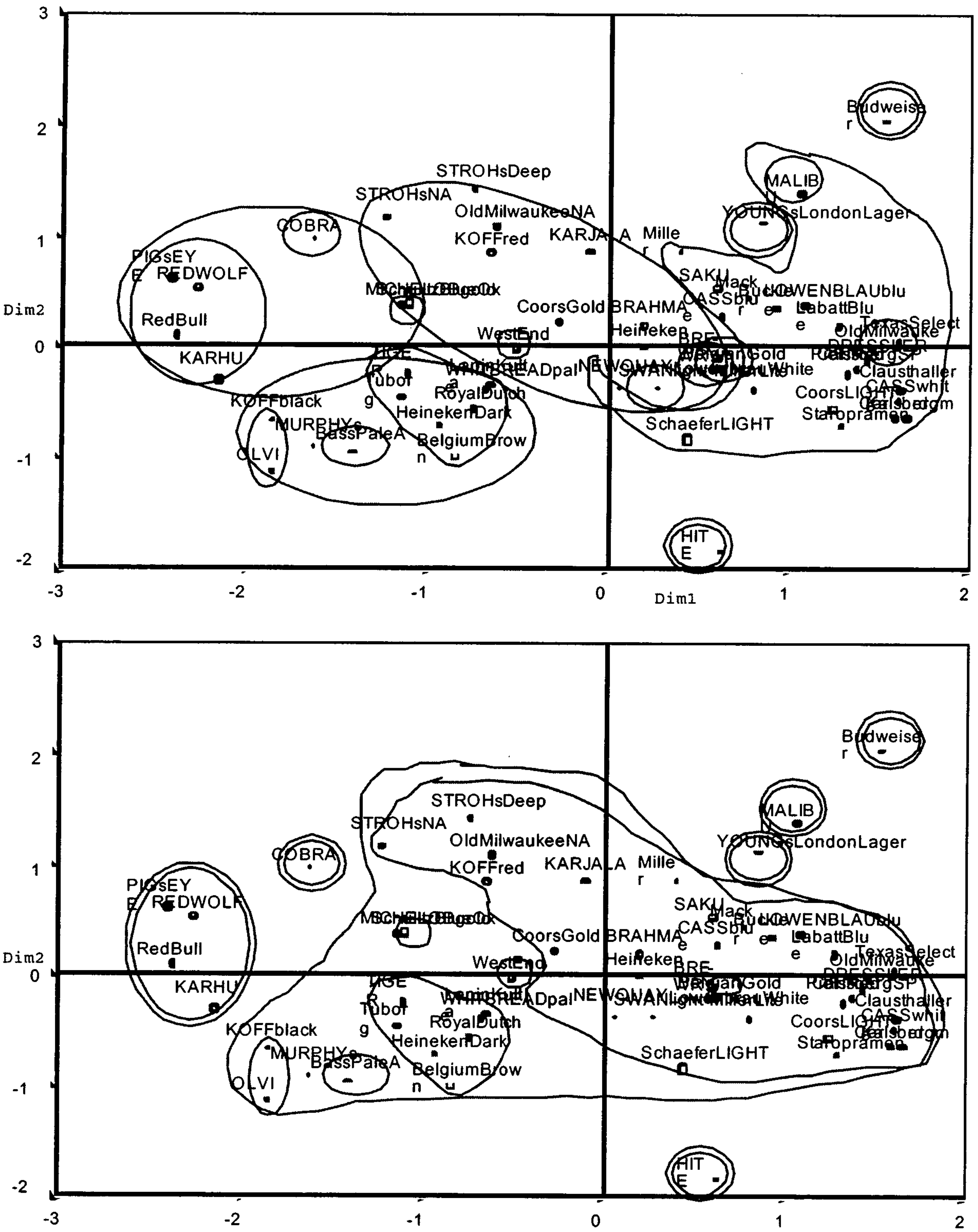


Figure 4. MDS map and Clusters (top: arboART, bottom: SLINK)