Journal of Classification 24:123-143 (2007)
DOI: 10.1007/s00357-007-0002-1

Asymmetric Agglomerative Hierarchical Clustering Algorithms and Their Evaluations

Akinobu Takeuchi

Jissen Women's University, Japan

Takayuki Saito

Tokyo Institute of Technology, Japan

Hiroshi Yadohisa

Doshisha University, Japan

Abstract: This paper presents asymmetric agglomerative hierarchical clustering algorithms in an extensive view point. First, we develop a new updating formula for these algorithms, proposing a general framework to incorporate many algorithms. Next we propose measures to evaluate the fit of asymmetric clustering results to data. Then we demonstrate numerical examples with real data, using the new updating formula and the indices of fit. Discussing empirical findings, through the demonstrative examples, we show new insights into the asymmetric clustering.

Key Words: Asymmetric data; Asymmetric dendrogram; Clustering analysis; Dissimilarity; Goodness of fit.

Authors' Addresses: Akinobu Takeuchi, Faculty of Humanities and Social Sciences, Jissen Women's University, 4-1-1 Osakaue, Hino, Tokyo 191-8510, Japan, e-mail: akitake@univ.jissen.ac.jp; Takayuki Saito, e-mail: gyo@valdes.titech.ac.jp; Hiroshi Yadohisa, e-mail: hyadohis@mail.doshisha.ac.jp

1. Introduction

1.1 Background

Let us suppose a situation in which one observes pairwise relationships among n objects. The entire set of data takes a form of $n \times n$ data matrix $\boldsymbol{D} = (d_{ij})$ where d_{ij} is a nonnegative numerical value, indicating the degree of relationship of an ordered pair $[i, j]$. There are often cases for which $\boldsymbol{D}$ is asymmetric, that is, $d_{ij} \neq d_{ji}$. For examples of such asymmetric measures, we mention brand switching, psychological (dis)similarity, social exchange, citation and so on (Zielman and Heiser 1996; Saito and Yadohisa 2005). For analysis of asymmetric data, researchers have developed models and methods by a variety of approaches, such as multidimensional scaling, clustering, network analysis, graphical representation techniques (Saito and Yadohisa 2005).

1.2 Purpose

We are concerned with clustering of asymmetric data. Referring to Saito and Yadohisa (2005), it is found that there have been developed many methods and algorithms to deal with the asymmetric clustering. It is noted that some of them have been given in terms of agglomerative hierarchical clustering algorithms (AHCA).

In this paper we are interested in extensions of the asymmetric AHCA and their evaluations. In Section 2, we provide asymmetric updating formulas to define asymmetric AHCA. Then we generalize asymmetric AHCA in a comprehensive framework, and present the representation of asymmetric dendrograms to show the results of asymmetric AHCA. In Section 3, we turn to measures of the goodness of fit for asymmetric AHCA in order to evaluate the clustering results. With some measures to indicate the degree of asymmetry of data, we suggest three measures for the evaluation. In Section 4, we present numerical examples for illustrative purposes of asymmetric AHCA. In Section 5, we summarize our contribution to asymmetric AHCA.

In passing extension of average linkage algorithms (Anderberg 1973) for symmetric case to asymmetric case is briefly described in Section 2. It requires some development of formulas which is provided in Appendix.

1.3 Notation

In what follows, we consider the situation in which the relationship between a pair of objects is given by an asymmetric dissimilarity measure. However, all the description of this paper can be applied to asymmetric similarity measures. Let d_{ij} denote the dissimilarity for an ordered pair $[i, j]$. Then, the asymmetry means that there exists at least one pair of objects i and j such that $d_{ij} \neq d_{ji}$. It is assumed that there exist no missing values. We will not deal with diagonal elements in the clustering algorithms. However, we let $d_{ii} = 0$ $(i = 1, 2, \cdots, n)$ to define some measures of goodness of fit for asymmetric AHCA in Section 3. We use a natural number I to indicate the I-th cluster, such as C_I. Let n_I be the number of objects which belong to C_I. Expression $i \in C_I$ states that object i belongs to C_I. Let d_{IJ} denote the dissimilarity for an ordered pair $[C_I, C_J]$. Note that $d_{IJ} \neq d_{JI}$ in general. Let h_I be the combined distance at the stage when C_I was generated in a clustering process (Yadohisa et al. 1999). When the result of the clustering is represented by a dendrogram, h_I indicates the height of C_I. When C_I is merged with C_J, the resultant cluster is written as $C_{IJ} = C_I \cup C_J$.

2. Asymmetric AHCA and Representation of Clustering Results

Researchers proposed so far some updating formulas for symmetric or asymmetric AHCA. The updating formula proposed by Lance and Williams (1966, 1967) for symmetric case is widely used since it implies many clustering algorithms by suitable selection of some parameters. Let us abbreviate it LWUF.

In this section, we describe briefly those symmetric and asymmetric updating formulas. Then we define a more general updating formula to represent asymmetric AHCA. Further, we give a representation of the asymmetric AHCA by using the new updating formula. Clustering results of the asymmetric AHCA are shown by asymmetric dendrograms.

2.1 Symmetric Updating Formulas

Let us state LWUF for dissimilarity $d_{(IJ)K}$ between cluster C_{IJ} and another cluster C_K as

$$d_{(IJ)K} = \alpha_I d_{IK} + \alpha_J d_{JK} + \beta d_{IJ} + \gamma \left| d_{IK} - d_{JK} \right|, \tag{1}$$

where parameters α_I, α_J, β and γ are either constants or functions of n_I, n_J and n_K. This formula implies many symmetric AHCA including single linkage algorithm (Sneath 1957), complete linkage algorithm (McQuitty 1960), group average algorithm and weighted average algorithm (McQuitty 1967), Ward's algorithm (Ward 1963), centroid algorithm (Gower 1967), median algorithm and flexible algorithm (Lance and Williams 1966).

Jambu (1978) extended LWUF for dissimilarity $d_{(IJ)K}$ between cluster C_{IJ} and another cluster C_K as

$$d_{(IJ)K} = \alpha_I d_{IK} + \alpha_J d_{JK} + \beta d_{IJ} + \gamma\,|d_{IK} - d_{JK}| + \delta_I h_I + \delta_J h_J + \varepsilon h_K, \quad (2)$$

where α_I, α_J, β and γ are the same in LWUF, and newly added δ_I, δ_J, ε are also parameters to define the algorithms.

By using this formula, several algorithms such as mean dissimilarity algorithm (Podani 1989) and sum of squares algorithm (Jambu 1978; Podani 1989) are stated, which cannot be represented by LWUF. See Jambu and Lebeaux (1983) and Gordon (1996) for details of symmetric AHCA. Note that the dissimilarities are assumed to be symmetric in the description of this subsection.

2.2 Asymmetric Updating Formulas

For asymmetric case, an extended updating formula was proposed by Yadohisa (2002). Incorporating the asymmetry into the formulation gives rise to two types of updating formulas. The updating formula for dissimilarity $d_{(IJ)K}$ for $[C_{IJ}, C_K]$ is written as

$$\begin{aligned} d_{(IJ)K} \;=\; & \alpha_I^{(1)} f^{(1)}(d_{IK}, d_{KI}) + \alpha_J^{(1)} f^{(1)}(d_{JK}, d_{KJ}) + \beta^{(1)} g^{(1)}(d_{IJ}, d_{JI}) \\ & + \gamma^{(1)} |f^{(1)}(d_{IK}, d_{KI}) - f^{(1)}(d_{JK}, d_{KJ})|. \end{aligned} \quad (3)$$

The updating formula for dissimilarity $d_{K(IJ)}$ for $[C_K, C_{IJ}]$ is written as

$$\begin{aligned} d_{K(IJ)} \;=\; & \alpha_I^{(2)} f^{(2)}(d_{IK}, d_{KI}) + \alpha_J^{(2)} f^{(2)}(d_{JK}, d_{KJ}) + \beta^{(2)} g^{(2)}(d_{IJ}, d_{JI}) \\ & + \gamma^{(2)} |f^{(2)}(d_{IK}, d_{KI}) - f^{(2)}(d_{JK}, d_{KJ})|. \end{aligned} \quad (4)$$

Here $\alpha_I^{(1)}$, $\alpha_J^{(1)}$, $\alpha_I^{(2)}$, $\alpha_J^{(2)}$, $\beta^{(1)}$, $\beta^{(2)}$, $\gamma^{(1)}$, $\gamma^{(2)}$ are constants or functions in terms of n_I, n_J and n_K, which are specified prior to analysis, and $f^{(1)}$, $f^{(2)}, g^{(1)}$, $g^{(2)}$ are functions of two dissimilarities determined prior to analysis. This pair of formulas is called extended updating formula (EUF).

In the same way as the symmetric case, we update the dissimilarities by using EUF, calculating the dissimilarities between clusters at stage $m+1$ on the basis of the dissimilarities at stage m. Unlike the symmetric case, these formulas update $d_{(IJ)K}$ and $d_{K(IJ)}$ separately. In the same way as LWUF, the parameters such as $\alpha_I^{(1)}$ control the combination of the clusters. In addition, functions $f^{(1)}$, $f^{(2)}$, $g^{(1)}$ and $g^{(2)}$ control some weighting of the asymmetric relationship between a pair of clusters. The EUF implies a large number of asymmetric AHCA (see Saito and Yadohisa (2005) for details).

As is shown later, EUF cannot represent some of the asymmetric AHCA dealt with in this paper. Then let us propose a more general asymmetric updating formula to state comprehensive treatments of the asymmetric AHCA.

Definition 1 (Asymmetric updating formula):

The updating formula for dissimilarity $d_{(IJ)K}$ for $[C_{IJ}, C_K]$ is written as

$$\begin{aligned} d_{(IJ)K} &= \alpha_I^{(1)} f^{(1)}(d_{IK}, d_{KI}) + \alpha_J^{(1)} f^{(1)}(d_{JK}, d_{KJ}) + \beta^{(1)} g^{(1)}(d_{IJ}, d_{JI}) \\ &\quad + \gamma^{(1)} |f^{(1)}(d_{IK}, d_{KI}) - f^{(1)}(d_{JK}, d_{KJ})| \\ &\quad + \delta_I^{(1)} h_I + \delta_J^{(1)} h_J + \varepsilon^{(1)} h_K + \zeta^{(1)}. \end{aligned} \tag{5}$$

The updating formula for dissimilarity $d_{K(IJ)}$ for $[C_K, C_{IJ}]$ is written as

$$\begin{aligned} d_{K(IJ)} &= \alpha_I^{(2)} f^{(2)}(d_{IK}, d_{KI}) + \alpha_J^{(2)} f^{(2)}(d_{JK}, d_{KJ}) + \beta^{(2)} g^{(2)}(d_{IJ}, d_{JI}) \\ &\quad + \gamma^{(2)} |f^{(2)}(d_{IK}, d_{KI}) - f^{(2)}(d_{JK}, d_{KJ})| \\ &\quad + \delta_I^{(2)} h_I + \delta_J^{(2)} h_J + \varepsilon^{(2)} h_K + \zeta^{(2)}. \end{aligned} \tag{6}$$

Here $\alpha_I^{(1)}$, $\alpha_J^{(1)}$, $\alpha_I^{(2)}$, $\alpha_J^{(2)}$, $\beta^{(1)}$, $\beta^{(2)}$, $\gamma^{(1)}$, $\gamma^{(2)}$, $\delta_I^{(1)}$, $\delta_I^{(2)}$, $\delta_J^{(1)}$, $\delta_J^{(2)}$, $\varepsilon^{(1)}$, $\varepsilon^{(2)}$, $\zeta^{(1)}$, and $\zeta^{(2)}$ are constants or functions of n_I, n_J and n_K, which are specified prior to analysis, and $f^{(1)}$, $f^{(2)}$, $g^{(1)}$ and $g^{(2)}$ are functions of two dissimilarities determined prior to analysis. We call these pairs of formulas asymmetric updating formula (AUF) hereafter. Figure 1 shows asymmetric dissimilarities between clusters.

2.3 Asymmetric AHCA

Let us formulate asymmetric AHCA by using AUF. Let C_S and C_R be arbitrary clusters, and d_{SR} and d_{RS} denote asymmetric dissimilarities between them. Consider C_{IJ} and C_K, particular clusters to be combined at a stage (see Figure 1). Without loss of generality, it is assumed that cluster C_K may be a singleton, or was formed before cluster C_{IJ} has been formed. To include many algorithms in this formulation, we introduce function W. The function W defines a combining criterion of two asymmetric dissimilarities between clusters (see formula (7)). For example it may be $\max$, $\min$, mean, and so on.

The algorithms consist of recursion of two steps; the first step for selection of the objects for combining and the second step for updating the asymmetric dissimilarity between clusters.

Definition 2 (Asymmetric AHCA):

The clustering algorithms determined by recursion of the following two steps are called asymmetric AHCA. Suppose that an asymmetric dissimilarity matrix is given.

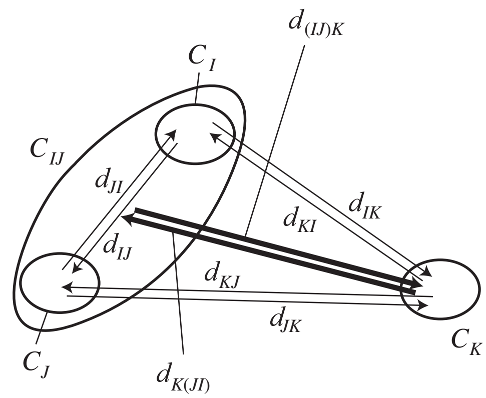

Figure 1. Asymmetric dissimilarities between clusters

Step 1 : For a given asymmetric dissimilarity matrix, find $[C_I, C_J]$ such that it attains the minimum of W values over all the pairs of clusters. For clarity, writing

$$v_{IJ} = \min_{S<R} W(d_{SR}, d_{RS}), \tag{7}$$

we determine $[C_I, C_J]$ which satisfies $v_{IJ} = W(d_{IJ}, d_{JI})$.

Step 2 : Join the two clusters which are selected in Step 1 and update $d_{(IJ)K}$ and $d_{K(IJ)}$ by AUF.

The entire process repeats Step 1 and Step 2 until all the objects form a cluster. The v_{IJ} value in (7) is called combined distance at the stage.

By using AUF (5) and (6), we can extend several symmetric AHCA to asymmetric ones, for example, average linkage algorithms (Anderberg 1973), mean dissimilarity algorithm (Podani 1989), and sum of squares algorithm (Jambu 1978). It should be noted that these asymmetric AHCA cannot be represented by EUF. The average linkage between the merged groups algorithm and the average linkage within the new group algorithm in Anderberg (1973) are called group average algorithm and total average algorithm, respectively, in this paper.

We characterize some of asymmetric AHCA in Table 1 with specified parameters in AUF. In Table 1, algorithm names are abbreviated as single linkage algorithm (SL), complete linkage algorithm (CL), weighted average algorithm

Table 1. Asymmetric AHCA by AUF parameters

Algorithm	$\alpha_I^{(1)}$ $(=\alpha_I^{(2)})$	$\alpha_J^{(1)}$ $(=\alpha_J^{(2)})$	$\beta^{(1)}$ $(=\beta^{(2)})$	$\gamma^{(1)}$ $(=\gamma^{(2)})$	$\delta_I^{(1)}$ $(=\delta_I^{(2)})$	$\delta_J^{(1)}$ $(=\delta_J^{(2)})$	$\varepsilon^{(1)}$ $(=\varepsilon^{(2)})$	$\zeta^{(1)}$ $(=\zeta^{(2)})$
ASL$_W$	$\frac{1}{2}$	$\frac{1}{2}$	0	$-\frac{1}{2}$	0	0	0	0
ACL$_W$	$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{2}$	0	0	0	0
AWA$_W$	$\frac{1}{2}$	$\frac{1}{2}$	0	0	0	0	0	0
ACE$_W$	$\frac{n_I}{n_{IJ}}$	$\frac{n_I}{n_{IJ}}$	$-\frac{n_I n_J}{n_{IJ}^2}$	0	0	0	0	0
AMD$_W$	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{4}$	0	0	0	0	0
AWD$_W$	$\frac{n_{IK}}{n_{IJK}}$	$\frac{n_{JK}}{n_{IJK}}$	$-\frac{n_K}{n_{IJK}}$	0	0	0	0	0
AFX$_W$	$\frac{1-\beta}{2}$	$\frac{1-\beta}{2}$	$\beta\ (<1)$	0	0	0	0	0
AGA$_W$	$\frac{n_I}{n_{IJ}}$	$\frac{n_J}{n_{IJ}}$	0	0	0	0	0	0
ATA$_W$	$\frac{\binom{n_{IK}}{2}}{\binom{n_{IJK}}{2}}$	$\frac{\binom{n_{JK}}{2}}{\binom{n_{IJK}}{2}}$	$\frac{\binom{n_{IJ}}{2}}{\binom{n_{IJK}}{2}}$	0	0	0	0	$\frac{-s^*}{\binom{n_{IJK}}{2}}$
ASS$_W$	$\frac{n_{IK}}{n_{IJK}}$	$\frac{n_{JK}}{n_{IJK}}$	$\frac{n_{IJ}}{n_{IJK}}$	0	$\frac{-n_I}{n_{IJK}}$	$\frac{-n_J}{n_{IJK}}$	$\frac{-n_K}{n_{IJK}}$	0
ADI$_W$	$\frac{\binom{n_{IK}}{2}}{\binom{n_{IJK}}{2}}$	$\frac{\binom{n_{JK}}{2}}{\binom{n_{IJK}}{2}}$	$\frac{\binom{n_{IJ}}{2}}{\binom{n_{IJK}}{2}}$	0	$\frac{-\binom{n_I}{2}}{\binom{n_{IJK}}{2}}$	$\frac{-\binom{n_J}{2}}{\binom{n_{IJK}}{2}}$	$\frac{-\binom{n_K}{2}}{\binom{n_{IJK}}{2}}$	0

In the table, definitions are given as $f^{(1)}(x,y) = x$, $f^{(2)}(x,y) = y$, $g^{(1)}(x,y) = g^{(2)}(x,y) = W(x,y)$. Here W stands for a function such as $\max, \min$, mean, etc. The symbol $\binom{}{}$ means a binomial coefficient. $n_{IJ} = n_I + n_J$, $n_{IJK} = n_I + n_J + n_K$, $s^* = s_I + s_J + s_K$. For definitions of s_I, s_J and s_K, refer to Appendix.

(WA), centroid algorithm (CE), median algorithm (MD), flexible algorithm (FX), group average algorithm (GA), total average algorithm (TA), sum of squares algorithm (SS), and mean dissimilarity algorithm (DI). For each of them, we code the corresponding asymmetric clustering algorithm by attaching "A", for example, ACL for asymmetric CL and so on. One may realize the advantage of AUF to EUF, that ATA, ASS and ADI are incorporated only by using AUF. Regarding asymmetric TA, in particular, we provide the formulation and its updating formula in Appendix.

2.4 Asymmetric Dendrogram

To illustrate the result of asymmetric AHCA, we may consider several extensions of the customary dendrogram. We make use of a general dendrogram (Saito and Yadohisa 2005) in the following sections. In the dendrogram, some distance information is provided to customary dendrogram (see Figure 2). The height of continuous bars in the figure shows combined distance v_{IJ}. In addition, d_{IJ} and d_{JI} are also displayed by dotted bars at each combining stage. These three values at each stage are used for evaluations of the clustering results in what follows. The left dendrogram shows the case for $d_{IJ} > d_{JI}$, and the right one shows the case for $d_{IJ} < d_{JI}$. Note that the continuous bar might coincide with one of the dotted bars when v_{IJ} is set as $\max(d_{IJ}, d_{JI})$ or $\min(d_{IJ}, d_{JI})$. For details about the general dendrogram, see Saito and Yadohisa (2005).

3. Measures of the Goodness of Fit for Asymmetric AHCA

In order to evaluate clustering results derived by asymmetric AHCA, let us consider some measures. At first we like to use measures of the degree of asymmetry in the data matrix $\boldsymbol{D} = (d_{ij})$. Define a symmetric $\boldsymbol{S} = (s_{jk})$ and a skew-symmetric $\boldsymbol{A} = (a_{jk})$ as

$$\boldsymbol{S} = (\boldsymbol{D} + \boldsymbol{D}')/2, \tag{8}$$
$$\boldsymbol{A} = (\boldsymbol{D} - \boldsymbol{D}')/2, \tag{9}$$

where $\boldsymbol{D}'$ is the transposed matrix of $\boldsymbol{D}$. Then we see that

$$\boldsymbol{D} = \boldsymbol{S} + \boldsymbol{A}, \tag{10}$$
$$\|\boldsymbol{D}\|^2 = \|\boldsymbol{S}\|^2 + \|\boldsymbol{A}\|^2, \tag{11}$$

and

$$\text{Var}(\{d_{jk}\}) = \text{Var}(\{s_{jk}\}) + \text{Var}(\{a_{jk}\}), \tag{12}$$

where $||\cdot||^2$ and Var$(\{\cdot\})$ are concerned with n^2 elements of each matrix. For measures of asymmetry of data, we choose to use the following indices:

$$\theta = \|\boldsymbol{A}\|^2/\|\boldsymbol{D}\|^2, \tag{13}$$
$$\phi = \text{Var}(\{a_{jk}\})/\text{Var}(\{d_{jk}\}), \tag{14}$$

which were suggested by Saito and Yadohisa (2005). The values of these indices are bounded as

$$0 \leq \theta \leq 1, \tag{15}$$
$$0 \leq \phi \leq 1. \tag{16}$$

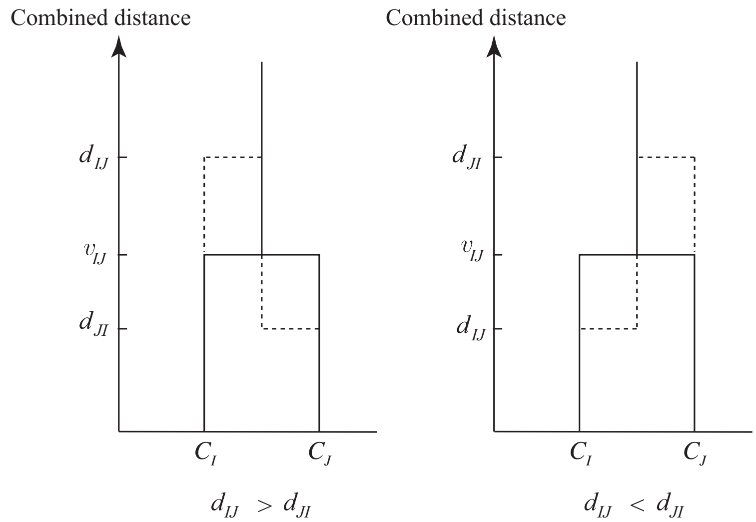

Figure 2. Asymmetric dendrograms

It is noticed that formulas from (8) to (14) are valid even for data matrices including nonzero diagonal elements. Other measures of asymmetry of data might be used (Gower 1977).

Next we consider measures of the fit of an asymmetric dendrogram to the data to which a certain asymmetric AHCA has been applied. Let us explain basic ideas by referring to Figure 3, which illustrates the example of an asymmetric dendrogram. By v_{IJ}, we denote the combined distance at each merger of two clusters, C_I and C_J. Then we have the resultant symmetric dissimilarity matrix $\boldsymbol{V}$, such as

$$\boldsymbol{V} = \begin{pmatrix} 0 & v_{AB} & v_{(AB)(CD)} & v_{(AB)(CD)} \\ v_{AB} & 0 & v_{(AB)(CD)} & v_{(AB)(CD)} \\ v_{(AB)(CD)} & v_{(AB)(CD)} & 0 & v_{CD} \\ v_{(AB)(CD)} & v_{(AB)(CD)} & v_{CD} & 0 \end{pmatrix}. \tag{17}$$

To state the process of clustering, we have another matrix of the resultant asymmetric dissimilarity matrix such as

$$\boldsymbol{U} = \begin{pmatrix} 0 & u_{AB} & u_{(AB)(CD)} & u_{(AB)(CD)} \\ u_{BA} & 0 & u_{(AB)(CD)} & u_{(AB)(CD)} \\ u_{(CD)(AB)} & u_{(CD)(AB)} & 0 & u_{CD} \\ u_{(CD)(AB)} & u_{(CD)(AB)} & u_{DC} & 0 \end{pmatrix}. \tag{18}$$

With $\boldsymbol{V}$ and $\boldsymbol{U}$, we like to consider some measures to indicate the fit.

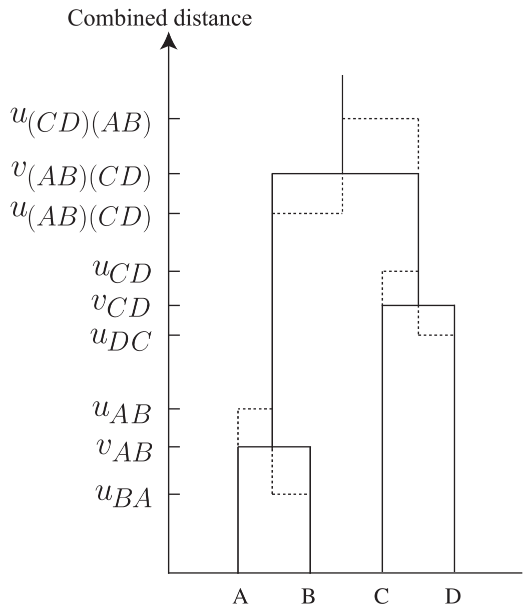

Figure 3. Example of an asymmetric dendrogram

Now we think of the general case of a result of AHCA, and let $\boldsymbol{U} = (u_{ij})$ be the resultant asymmetric dissimilarity where u_{ij} elements are read from the asymmetric dendrogram. We retain the original dissimilarity matrix $\boldsymbol{D} = (d_{ij})$.

As extension of Cophenetic correlation coefficient (Sokal and Rohlf 1962) in the case of symmetric clustering to the case of asymmetric clustering, we set

$$r_* = \frac{\text{Cov}(\{d_{ij}\}, \{v_{ij}\})}{(\text{Var}(\{d_{ij}\})\text{Var}(\{v_{ij}\}))^{\frac{1}{2}}}\,, \tag{19}$$

where variance and covariance terms are concerned with only $n(n-1)$ off-diagonal elements. This convention is applied to formulas including those terms later. We note the following relations:

$$\text{Cov}(\{d_{ij}\}, \{v_{ij}\}) = \text{Cov}(\{s_{ij}\}, \{v_{ij}\}) + \text{Cov}(\{a_{ij}\}, \{v_{ij}\}), \tag{20}$$

$$\text{Cov}(\{a_{ij}\}, \{v_{ij}\}) = \sum_{i \neq j} a_{ij} v_{ij} - \frac{1}{n(n-1)} \sum_{i \neq j} a_{ij} \sum_{i \neq j} v_{ij} = 0, \tag{21}$$

$$\sum_{i \neq j} a_{ij} v_{ij} = 0\,, \quad \sum_{i \neq j} a_{ij} = 0. \tag{22}$$

Then we have

$$\mathrm{Cov}(\{d_{ij}\}, \{v_{ij}\}) = \mathrm{Cov}(\{s_{ij}\}, \{v_{ij}\}). \tag{23}$$

Hence the cophenetic correlation coefficient is stated as

$$r_* = \frac{\mathrm{Cov}(\{s_{ij}\}, \{v_{ij}\})}{(\mathrm{Var}(\{d_{ij}\})\mathrm{Var}(\{v_{ij}\}))^{\frac{1}{2}}}. \tag{24}$$

As another measure of fit, let us propose the sum of squared errors ratio ψ_* (Hartigan 1967). Through simple manipulation, we have

$$\|\boldsymbol{D} - \boldsymbol{V}\|^2 = \|\boldsymbol{S} - \boldsymbol{V}\|^2 + \|\boldsymbol{A}\|^2, \tag{25}$$

because

$$\mathrm{tr}[(\boldsymbol{S} - \boldsymbol{V})\boldsymbol{A}] = 0. \tag{26}$$

Then we present ψ_* as

$$\psi_* = \frac{\|\boldsymbol{D} - \boldsymbol{V}\|^2}{\|\boldsymbol{D}\|^2}. \tag{27}$$

It turns out that both (24) and (27) indicate the fit of $\boldsymbol{V}$ to $\boldsymbol{S}$ for a dendrogram. Therefore it is realized that we should use $\boldsymbol{U}$, asymmetric information given by the dendrogram. This consideration gives rise to a suggestion of two measures such as

$$r = \frac{\mathrm{Cov}(\{d_{ij}\}, \{u_{ij}\})}{(\mathrm{Var}(\{d_{ij}\})\mathrm{Var}(\{u_{ij}\}))^{\frac{1}{2}}} \tag{28}$$

and

$$\psi = \frac{\|\boldsymbol{D} - \boldsymbol{U}\|^2}{\|\boldsymbol{D}\|^2}. \tag{29}$$

Let us transform ψ to ω by

$$\omega = \frac{1}{1+\psi}, \tag{30}$$

which indicates the degree of goodness of fit.

As a third measure of fit, it is suggested to compute rank correlation coefficient between $\{u_{ij}\}$ and $\{d_{ij}\}$, using Spearman's ρ and Kendall's τ. As shown in (18), many ties occur in the values of $\boldsymbol{U}$ matrix. In view of this property, we incorporate the correction factors of Siegel (1956) for the computation of ρ and τ.

Let us provide numerical illustrations for what have been suggested so far. Given an asymmetric dissimilarity matrix

$$\boldsymbol{D} = \begin{pmatrix} 0 & 30 & 80 & 90 \\ 10 & 0 & 70 & 100 \\ 110 & 90 & 0 & 80 \\ 120 & 110 & 60 & 0 \end{pmatrix}, \tag{31}$$

we have its symmetric part $\boldsymbol{S}$ and skew-symmetric part $\boldsymbol{A}$. From $\boldsymbol{D}$ and $\boldsymbol{A}$, the degrees of asymmetry of the data are indicated by $\theta = 0.018$ and $\phi = 0.050$. When we set the arithmetic mean for function W, for example, we can perform the asymmetric complete linkage algorithm (abbrev. ACL_{mean}) to the data. Then the result is represented by the asymmetric dendrogram in Figure 4. From the dendrogram, we derive

$$\boldsymbol{V} = \begin{pmatrix} 0 & 20 & 110 & 110 \\ 20 & 0 & 110 & 110 \\ 110 & 110 & 0 & 70 \\ 110 & 110 & 70 & 0 \end{pmatrix} \text{ and } \boldsymbol{U} = \begin{pmatrix} 0 & 30 & 100 & 100 \\ 10 & 0 & 100 & 100 \\ 120 & 120 & 0 & 80 \\ 120 & 120 & 60 & 0 \end{pmatrix}, \quad (32)$$

according to (17) and (18). Given $\boldsymbol{D}$ and $\boldsymbol{U}$, the goodness of fit for ACL_{mean} is evaluated by $r = 0.951$, $\omega = 0.972$, $\rho = 0.898$ and $\tau = 0.823$.

4. Numerical Examples

Let us give numerical examples by using some of asymmetric AHCA and measures of the goodness of fit for clustering results.

We take up the data of brand switching of soft drinks between two periods (DeSarbo 1982), which are reported in Table 2. The brands are eight soft drinks (A: Coke, B: 7-up, C: Tab, D: Like, E: Pepsi, F: Sprite, G: Diet Pepsi, H: Fresca). The data matrix in the table shows the brand switching in terms of proportions with the rowwise sum being unity. As is seen, the matrix is asymmetric and each entry means the similarity between a pair of brands. To apply asymmetric AHCA to the data $\{o_{ij}\}$, it is required to transform them into dissimilarity data $\{d_{ij}\}$ by a monotone function,

$$d_{ij} = 1000 \times (\max_{i \neq j}(o_{ij}) - o_{ij}) + 1. \quad (33)$$

We set $d_{ii} = 0$ $(i = 1, \cdots, 8)$. The results are given in Table 3. The degree of asymmetry of the transformed data is shown by $\theta = 0.032$ and $\phi = 0.180$. Thus the degree of skew-symmetry is small in terms of the norm, while it is intermediate in terms of the variance.

For purposes of demonstration, we selected four typical algorithms (ASL, ACL, AGA and ATA) among asymmetric AHCA, which are derived from single linkage, complete linkage and average linkage, respectively. For investigating the effect of W, we take up two most distinct functions, $W = \min$ and $W = \max$. Table 4 summarizes clustering processes of ASL, ACL, AGA and ATA. The values in the column v_{IJ} indicate the combined distance at each stage. In the table, C_I and C_J show merged clusters. We illustrate the dendrograms of clustering results for each algorithm in Figure 5.

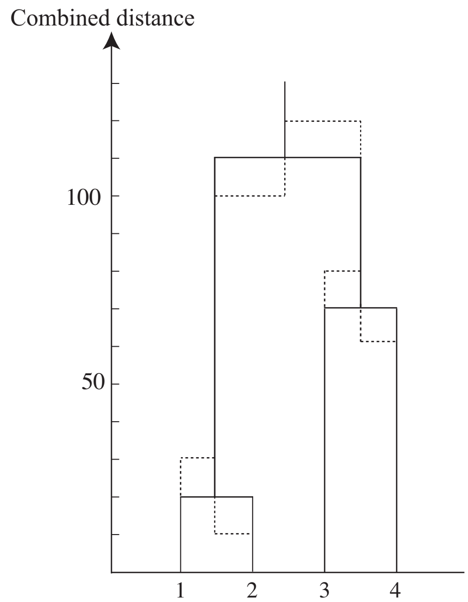

Figure 4. Numerical example of an asymmetric dendrogram

Let us mention some findings from Table 4 and Figure 5. Firstly, it is stressed that the monotonicity for the combined distance is not satisfied for asymmetric algorithms ASL$_{\max}$, AGA$_{\max}$ and ATA$_{\max}$, for which the monotonicity is known to hold in the case of symmetric AHCA. Secondly, the clustering results are influenced by the function type of W more than by the linkage type. Thirdly, the present results of ASL$_{\min}$ and ASL$_{\max}$ do not show the chain effect in the sense of Lance and Williams (1967) that is frequently observed in the case of symmetric SL. Fourthly, examining the values in terms of r, ω and ρ indices, we notice the following points. For the case $W = \min$, AGA$_{\min}$ attains the best fit in terms of r and ω, and ACL$_{\min}$ does the best in terms of ρ. For the case $W = \max$, AGA$_{\max}$ gives the best in terms of r and ω, while ATA$_{\max}$ does the best in terms of ρ. Across cases of $W = \min$ and $W = \max$, AGA$_{\min}$ shows the best fit in terms of r and ω and ATA$_{\max}$ does the best fit in terms of ρ.

To investigate the points mentioned above, we performed another study, using other algorithms of asymmetric AHCA. Changing the values of parameters in AUF, we can implement various clustering algorithms, and then obtain the corresponding results. To find the best result in these clusterings, we will refer to the goodness of fit.

Table 2. Soft drink brand switching data

	Period $[t+1]$							
Period $[t]$	A	B	C	D	E	F	G	H
A : Coke	.612	.107	.010	.033	.134	.055	.013	.036
B : 7-Up	.186	.448	.005	.064	.140	.099	.012	.046
C : Tab	.080	.120	.160	.360	.080	.040	.080	.080
D : Like	.087	.152	.087	.152	.239	.043	.131	.109
E : Pepsi	.177	.132	.008	.030	.515	.076	.026	.037
F : Sprite	.114	.185	.029	.071	.157	.329	.029	.086
G : Diet Pepsi	.093	.047	.186	.093	.116	.093	.256	.116
H : Fresca	.226	.093	.053	.107	.147	.107	.067	.200

Table 3. Transformed data of dissimilarities ($\theta = 0.032, \phi = 0.180$)

	Period $[t+1]$							
Period $[t]$	A	B	C	D	E	F	G	H
A : Coke	0	254	351	328	227	306	348	325
B : 7-Up	175	0	356	297	221	262	349	315
C : Tab	281	241	0	1	281	321	281	281
D : Like	274	209	274	0	122	318	230	252
E : Pepsi	184	229	353	331	0	285	335	324
F : Sprite	247	176	332	290	204	0	332	275
G : Diet Pepsi	268	314	175	268	245	268	0	245
H : Fresca	135	268	308	254	214	254	294	0

Let r, ω, ρ be index values of clustering results by using the nine algorithms of asymmetric AHCA (ASL, ACL, AWA, ACE, AMD, AWD, AFX, AGA, ATA), respectively. Examining the fit values in Table 5, let us summarize findings as follows. For the case $W = \min$, the best r is attained by AFX$_{\min}$, the best ω by both AWA$_{\min}$ and AGA$_{\min}$, and the best ρ by AFX$_{\min}$. For the case $W = \max$, the best r is attained by AGA$_{\max}$, the best ω by AGA$_{\max}$, and the best ρ by ATA$_{\max}$. Combining both cases for each fit index, we find that the best r is given by AFX$_{\min}$, the best ω is by both AWA$_{\min}$ and AGA$_{\min}$, and the best ρ by ATA$_{\max}$. Furthermore, there is a general tendency that ACE, AMD and AWD give lower degrees of fit than the other algorithms.

As we stand to a basic view that symmetrization of asymmetric data discards some important information involved in data, from which the study of this paper started. Then we are not concerned with discussing possible advantages of asymmetric AHCA to symmetric AHCA, but with proposing asymmetric AHCA and their evaluations. However, it may be interesting to compare those results of asymmetric AHCA with performance of symmetric AHCA using the same set of data. Then we symmetrized the data of Table 3, and applied to them

Table 4. Clustering results of soft drink data by using ASL, ACL, AGA and ATA

$\mathrm{ASL}_{\min}$	v_{IJ}	C_I	C_J	d_{IJ}	d_{JI}	$\mathrm{ASL}_{\max}$	v_{IJ}	C_I	C_J	d_{IJ}	d_{JI}
1	1.0	C	D	1.0	274.0	1	227.0	A	E	227.0	184.0
2	122.0	C	E	122.0	331.0	2	229.0	A	B	229.0	175.0
3	135.0	A	H	325.0	135.0	3	254.0	D	H	252.0	254.0
4	175.0	B	H	175.0	254.0	4	245.0	G	H	245.0	230.0
5	175.0	C	G	230.0	175.0	5	175.0	C	G	1.0	175.0
6	176.0	B	F	254.0	176.0	6	262.0	A	F	262.0	176.0
7	184.0	F	G	204.0	184.0	7	275.0	A	G	275.0	122.0
r: 0.597, ω: 0.913, ρ: 0.444						r: 0.521, ω: 0.881, ρ: 0.570					
$\mathrm{ACL}_{\min}$	v_{IJ}	C_I	C_J	d_{IJ}	d_{JI}	$\mathrm{ACL}_{\max}$	v_{IJ}	C_I	C_J	d_{IJ}	d_{JI}
1	1.0	C	D	1.0	274.0	1	227.0	A	E	227.0	184.0
2	135.0	A	H	325.0	135.0	2	254.0	A	B	254.0	221.0
3	176.0	B	F	262.0	176.0	3	254.0	D	H	252.0	254.0
4	221.0	E	F	285.0	221.0	4	281.0	C	G	281.0	175.0
5	268.0	C	G	281.0	268.0	5	306.0	A	F	306.0	247.0
6	281.0	G	H	281.0	351.0	6	308.0	C	H	281.0	308.0
7	321.0	F	G	356.0	321.0	7	356.0	A	H	356.0	321.0
r: 0.720, ω: 0.949, ρ: 0.580						r: 0.602, ω: 0.944, ρ: 0.732					
$\mathrm{AGA}_{\min}$	v_{IJ}	C_I	C_J	d_{IJ}	d_{JI}	$\mathrm{AGA}_{\max}$	v_{IJ}	C_I	C_J	d_{IJ}	d_{JI}
1	1.0	C	D	1.0	274.0	1	227.0	A	E	227.0	184.0
2	135.0	A	H	325.0	135.0	2	241.5	A	B	241.5	198.0
3	176.0	B	F	262.0	176.0	3	254.0	D	H	252.0	254.0
4	201.5	C	E	201.5	342.0	4	262.0	G	H	256.5	262.0
5	229.3	C	G	282.0	229.3	5	252.3	C	H	187.7	252.3
6	253.0	F	H	253.0	270.5	6	284.3	A	F	284.3	209.0
7	268.4	F	G	294.1	268.4	7	327.6	A	H	327.6	250.8
r: 0.755, ω: 0.976, ρ: 0.543						r: 0.686, ω: 0.971, ρ: 0.723					
$\mathrm{ATA}_{\min}$	v_{IJ}	C_I	C_J	d_{IJ}	d_{JI}	$\mathrm{ATA}_{\max}$	v_{IJ}	C_I	C_J	d_{IJ}	d_{JI}
1	1.0	C	D	1.0	274.0	1	227.0	A	E	227.0	184.0
2	134.7	C	E	134.7	228.3	2	236.7	A	B	236.7	207.7
3	135.0	A	H	325.0	135.0	3	254.0	D	H	252.0	254.0
4	176.0	B	F	262.0	176.0	4	259.3	G	H	255.7	259.3
5	182.0	C	G	208.3	182.0	5	255.8	C	H	223.5	255.8
6	220.5	F	H	220.5	232.2	6	260.5	A	F	260.5	222.8
7	239.6	F	G	254.3	239.6	7	297.8	A	H	297.8	254.0
r: 0.676, ω: 0.952, ρ: 0.367						r: 0.670, ω: 0.966, ρ: 0.743					

eight algorithms of symmetric AHCA. For eight dendrograms derived by those algorithms, we computed the goodness of fit in terms of the three indices. Table 6 summarizes the results. It is found that the asymmetric AHCA approach gives generally higher fit than the symmetric AHCA approach for indices r and ρ but they do not exhibit clear differences for ω. It is stressed that this comparison in terms of the three indices is based on a single set of data.

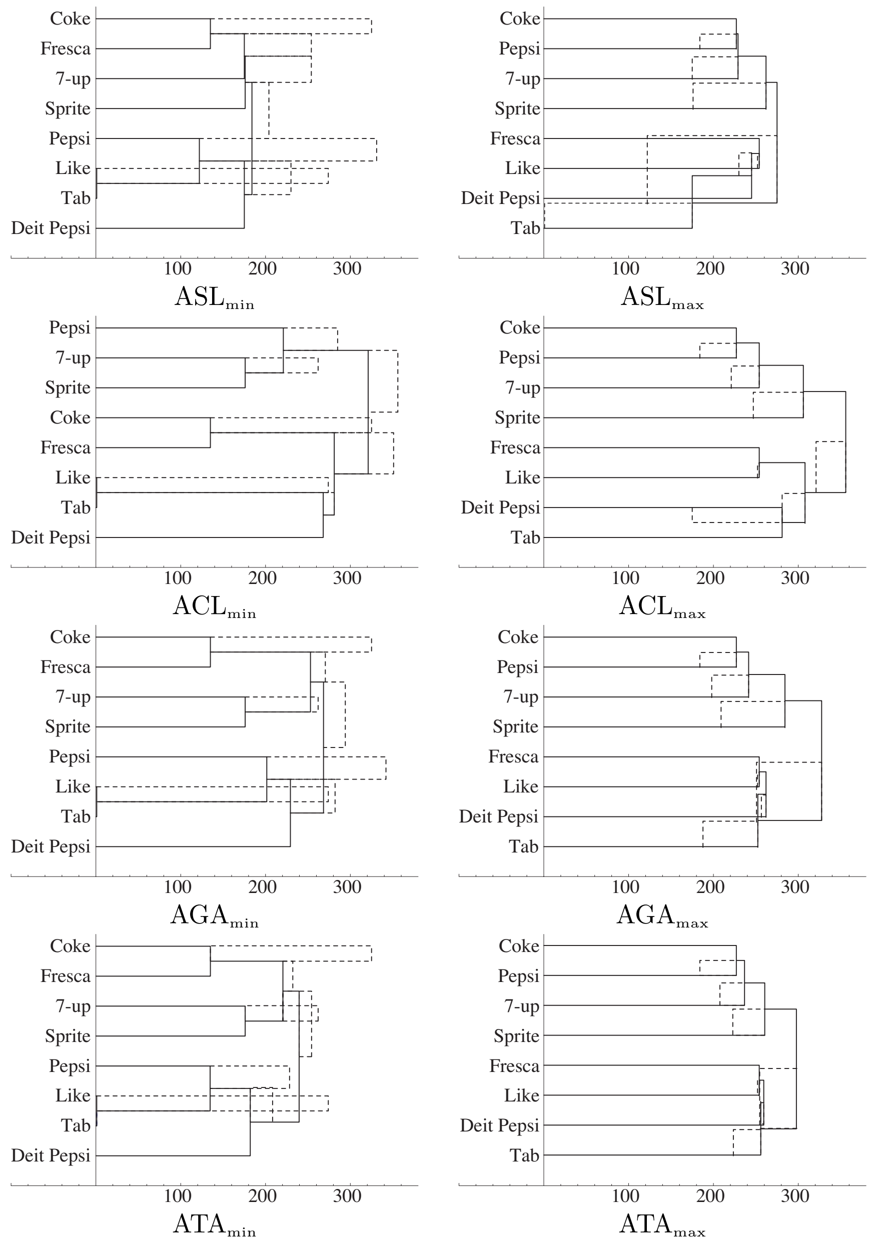

Figure 5. Dendrograms for soft drink data by using ASL, ACL, AGA and ATA

Table 5. Goodness of fit for soft drink data using asymmetric AHCA

Algorithm	r	ω	ρ	Algorithm	r	ω	ρ
$\mathrm{ASL_{min}}$	0.597	0.913	0.444	$\mathrm{ASL_{max}}$	0.521	0.881	0.570
$\mathrm{ACL_{min}}$	0.720	0.949	0.580	$\mathrm{ACL_{max}}$	0.602	0.944	0.732
$\mathrm{AWA_{min}}$	0.754	0.976	0.543	$\mathrm{AWA_{max}}$	0.654	0.967	0.737
$\mathrm{ACE_{min}}$	0.432	0.846	0.266	$\mathrm{ACE_{max}}$	0.433	0.863	0.405
$\mathrm{AMD_{min}}$	0.448	0.844	0.361	$\mathrm{AMD_{max}}$	0.165	0.846	0.082
$\mathrm{AWD_{min}}$	0.701	0.786	0.663	$\mathrm{AWD_{max}}$	0.652	0.766	0.694
$\mathrm{AFX_{min}}$	0.763	0.928	0.666	$\mathrm{AFX_{max}}$	0.640	0.906	0.691
$\mathrm{AGA_{min}}$	0.755	0.976	0.543	$\mathrm{AGA_{max}}$	0.686	0.971	0.723
$\mathrm{ATA_{min}}$	0.676	0.952	0.367	$\mathrm{ATA_{max}}$	0.670	0.966	0.743

Note: $\beta = -0.25$ in $\mathrm{AFX_{min}}$ and $\mathrm{AFX_{max}}$

Table 6. Goodness of fit for soft drink data using symmetric AHCA

Algorithm	r	ω	ρ
SL	0.455	0.932	0.224
CL	0.541	0.947	0.496
WA	0.555	0.962	0.523
CE	0.095	0.854	0.020
MD	0.103	0.844	0.059
WD	0.507	0.779	0.488
FX	0.542	0.931	0.524
GA	0.556	0.962	0.523
TA	0.532	0.955	0.513

Note: $\beta = -0.25$ in FX

5. Concluding Remarks

For asymmetric AHCA, this paper developed new updating formulas AUF ((5) and (6)), which are more general than EUF ((3) and (4)). Using the new formulas, we presented the general framework of asymmetric AHCA, giving nine algorithms explicitly (Table 1). Changing function W and/or parameters, we can incorporate many algorithms. For evaluation of clustering results of asymmetric AHCA, we suggested fit indices in terms of r, ω and ρ. For demonstration of our study, we presented numerical examples with real data using nine algorithms and evaluated them. As far as we are concerned with the soft drinks data, we observe that AGA gives the best result. It is not the goal of this paper to claim precedence of performance of some algorithms over the others. Hence, to make such a claim, one would need to investigate numerical study in more detail. There might be choices for updating formulas of asymmetric AHCA, functions of W, and definitions for measuring goodness of fit, which should be different from those suggested in this paper. It remains to perform a comprehensive study under possible choices of them.

Appendix

In Table 1, we have specified parameters for an algorithm, which is developed from the total average algorithm (TA) for the case of symmetric data. Here let us describe the derivation of those parameters. At first we review the definition of symmetric TA, and show the updating formula. Then we formulate asymmetric TA and present its updating formula.

To describe symmetric TA, consider a merger of clusters C_P and C_Q. Then the total average, which is called average linkage with the new group in Anderberg's formulation, is defined by

$$d_{PQ} = \sum_{i<j} \sum_{(i,j \in C_{PQ})} d_{ij} \Big/ \binom{n_{PQ}}{2}. \tag{34}$$

On the basis of this definition, we like to derive an updating formula for the total average algorithm. Denote the sum of dissimilarities within C_G by s_G and that between C_P and C_Q by t_{PQ}, such as

$$s_G = \sum_{i<j} \sum_{(i,j \in C_G)} d_{ij} \quad \text{and} \quad t_{PQ} = \sum_{i \in C_P} \sum_{j \in C_Q} d_{ij}. \tag{35}$$

Using these definitions, we derive the updating formula for merging C_{IJ} and C_K as follows.

$$\begin{aligned}
d_{(IJ)K} = \; & s_{(IJ)K} \Big/ \binom{n_{IJK}}{2} && (36)\\
= \; & (s_{IJ} + s_K + t_{(IJ)K}) \Big/ \binom{n_{IJK}}{2} \\
= \; & (s_I + s_J + t_{IJ} + s_K + t_{IK} + t_{JK}) \Big/ \binom{n_{IJK}}{2} \\
= \; & (s_I + s_K + t_{IK} + s_J + s_K + t_{JK} + s_I + s_J + t_{IJ}) \Big/ \binom{n_{IJK}}{2} \\
& -(s_I + s_J + s_K) \Big/ \binom{n_{IJK}}{2} \\
= \; & \left\{ \binom{n_{IK}}{2} d_{IK} + \binom{n_{JK}}{2} d_{JK} + \binom{n_{IJ}}{2} d_{IJ} - (s_I + s_J + s_K) \right\} \Big/ \binom{n_{IJK}}{2} \\
= \; & \left\{ \binom{n_{IK}}{2} \Big/ \binom{n_{IJK}}{2} \right\} d_{IK} + \left\{ \binom{n_{JK}}{2} \Big/ \binom{n_{IJK}}{2} \right\} d_{JK} && (37)\\
& + \left\{ \binom{n_{IJ}}{2} \Big/ \binom{n_{IJK}}{2} \right\} d_{IJ} - \left\{ 1 \Big/ \binom{n_{IJK}}{2} \right\} (s_I + s_J + s_K).
\end{aligned}$$

Let us turn to generalize the symmetric TA to an algorithm for asymmetric TA. Let C_G be a singleton or a merger of C_P and C_Q,

$$C_G = C_P \cup C_Q \quad (C_P \cap C_Q = \phi). \tag{38}$$

Considering asymmetric case, let

$$s_G = \begin{cases} 0, & (n_G = 1), \\ \binom{n_{PQ}}{2} \cdot A(d_{PQ}, d_{QP}), & (n_G > 1). \end{cases} \tag{39}$$

Here function A indicates a kind of average dissimilarities between C_P and C_Q. For two dissimilarities d and d', it is required to satisfy

$$A(d, d') \geq 0 \quad \text{and} \quad A(d, d) = d. \tag{40}$$

For asymmetric d_{ij}, let

$$t_{PQ} = \sum_{i \in C_P} \sum_{j \in C_Q} d_{ij}. \tag{41}$$

Using (39) and (41), we define asymmetric TA as follows.

$$\begin{aligned} d_{(IJ)K} &= (s_{IJ} + s_K + t_{(IJ)K}) \Big/ \binom{n_{IJK}}{2}, \\ d_{K(IJ)} &= (s_{IJ} + s_K + t_{K(IJ)}) \Big/ \binom{n_{IJK}}{2}. \end{aligned} \tag{42}$$

Referring to the case of symmetric d_{ij}, it is noted that (42) coincides with (34) for whatever choice of function A.

From the definition of asymmetric TA, we derive AUF as follows.

$$\begin{aligned} d_{(IJ)K} = & \quad (s_{IJ} + s_K + t_{(IJ)K}) \Big/ \binom{n_{IJK}}{2} \\ = & \quad (s_{IJ} + s_K + t_{IK} + t_{JK}) \Big/ \binom{n_{IJK}}{2} \\ = & \quad \{s_{IJ} + s_I + s_K + t_{IK} + s_J + s_K + t_{JK} - (s_I + s_J + s_K)\} \Big/ \binom{n_{IJK}}{2} \\ = & \quad \left\{ \binom{n_{IK}}{2} d_{IK} + \binom{n_{JK}}{2} d_{IJ} + \binom{n_{IJ}}{2} A(d_{IJ}, d_{JI}) - (s_I + s_J + s_K) \right\} \\ & \quad \Big/ \binom{n_{IJK}}{2} \end{aligned} \tag{43}$$

It is found that (43) is an expression of AUF by putting parameters as follows:

$$
\begin{aligned}
\alpha_I^{(1)} = \alpha_I^{(2)} &= \binom{n_{IK}}{2} \Big/ \binom{n_{IJK}}{2}, \\
\alpha_J^{(1)} = \alpha_J^{(2)} &= \binom{n_{JK}}{2} \Big/ \binom{n_{IJK}}{2}, \\
\beta^{(1)} = \beta^{(2)} &= \binom{n_{IJ}}{2} \Big/ \binom{n_{IJK}}{2}, \\
\gamma^{(1)} = \gamma^{(2)} = \delta_I^{(1)} = \delta_I^{(2)} &= \delta_J^{(1)} = \delta_J^{(2)} = \varepsilon^{(1)} = \varepsilon^{(2)} = 0, \\
\zeta^{(1)} = \zeta^{(2)} &= -(s_I + s_J + s_K) \Big/ \binom{n_{IJK}}{2}, \\
f^{(1)}(x,y) = x, f^{(2)}(x,y) = y,\ & g^{(1)}(x,y) = g^{(2)}(x,y) = A(x,y).
\end{aligned} \tag{44}
$$

References

ANDERBERG, M.R. (1973), *Cluster Analysis for Applications*, New York: Academic Press.

DESARBO, W.S. (1982), "GENNCLUS: New Models for General Nonhierarchical Clustering Analysis," *Psychometrika, 47*, 449–475.

GORDON, A.D. (1996), "Hierarchical Classification," in *Clustering and Classification*, Eds. P. Arabie, L.J. Hubert and G. De Soete, New Jersey: World Scientific Publication, 65–121.

GOWER, J.C. (1967), "A Comparison of Some Methods of Cluster Analysis," *Biometrics, 23*, 623–638.

GOWER, J.C. (1977), "The Analysis of Asymmetry and Orthogonality," in *Recent Developments in Statistics*, Eds. J.R. Barra, et al., Amsterdam: North-Holland, 109–123.

HARTIGAN, J.A. (1967), "Representation of Similarity Matrices by Trees," *Journal of the American Statistical Association, 62*, 1140–1158.

JAMBU, M. (1978), *Classification Automatique pour l'Analyse des Données*, Paris: Dunod.

JAMBU, M., and LEBEAUX, M-O. (1983), *Cluster Analysis and Data Analysis*, Amsterdam: North-Holland.

LANCE, G.N., and WILLIAMS, W.T. (1966), "A Generalized Sorting Strategy for Computer Classifications," *Nature, 212*, 218.

LANCE, G.N., and WILLIAMS, W.T. (1967), "A General Theory of Classificatory Sorting Strategies: 1. Hierarchical Systems," *The Computer Journal, 9*, 373–380.

MCQUITTY, L.L. (1960), "Hierarchical Linkage Analysis for the Isolation of Types," *Educational and Psychological Measurement, 20*, 55–67.

MCQUITTY, L.L. (1967), "Expansion of Similarity Analysis by Reciprocal Pairs for Discrete and Continuous Data," *Educational and Psychological Measurement, 27*, 253–255.

PODANI, J. (1989), "New Combinatorial SAHN Clustering Methods," *Vegetatio, 81*, 61–77.

SAITO, T., and YADOHISA, H. (2005), *Data Analysis of Asymmetric Structures*, New York: Marcel Dekker.

SIEGEL, S. (1956), *Nonparametric Statistics for the Behavioral Sciences*, New York: McGraw-Hill.

SNEATH, P.H.A. (1957), "The Application of Computers to Taxonomy," *Journal of General Microbiology, 17*, 201–226.

SOKAL, R.R., and ROHLF, F.J. (1962), "The Comparison of Dendrograms by Objective Methods," *Taxon, 11*, 33–40.

WARD, J.H., Jr. (1963), "Hierarchical Grouping to Optimize an Objective Function," *Journal of the American Statistical Association, 58*, 236–244.

YADOHISA, H. (2002), "Formulation of Asymmetric Agglomerative Hierarchical Clustering and Graphical Representation of its Result," *Bulletin of The Computational Statistics of Japan, 15*, 309–316 (in Japanese with English summary).

YADOHISA, H., TAKEUCHI, A., and INADA, K. (1999), "Developing Criteria for Measuring Space Distortion in Combinatorial Cluster Analysis and Methods for Controlling the Distortion," *Journal of Classification, 16*, 45–62.

ZIELMAN, B., and HEISER, W.J. (1996), "Models for Asymmetric Proximities," *British Journal of Mathematical and Statistical Psychology, 49*, 127–146.