Pergamon

Pattern Recognition, Vol. 31, No. 10, pp. 1573–1578, 1998

0031-3203/98 $19.00 + 0.00

PII: S0031-3203(98)00002-8

CHARACTERIZATIONS OF NEAREST AND FARTHEST NEIGHBOR ALGORITHMS BY CLUSTERING ADMISSIBILITY CONDITIONS

ZHENMIN CHEN*,† and JOHN VAN NESS‡

† Department of Statistics, Florida International University, Miami, FL 33 199, U.S.A.
‡ Programs in Mathematical Sciences, University of Texas at Dallas, Richardson, TX 75083, U.S.A.

(*Received* 15 *March* 1996; *in revised form* 24 *December* 1997)

Abstract—Monotone admissibility for clustering algorithms was introduced in Fisher and Van Ness [*Biometrika* **58**, 91–104 (1971)]. The present paper discusses monotone admissibility for a broad class of clustering algorithms called the Lance and Williams algorithms. Necessary and sufficient conditions for Lance and Williams algorithms to be monotone admissible are discussed here. It is shown that the only such algorithms which are monotone admissible are nearest neighbor and farthest neighbor. © 1998 Published by Elsevier Ltd on behalf of the Pattern Recognition Society. All rights reserved

Clustering Clustering admissibility Monotone clustering admissibility
Nearest-neighbor clustering Farthest-neighbor clustering Lance and Williams algorithms

1. MONOTONE ADMISSIBLE CLUSTERING ALGORITHMS

Monotone admissibility for clustering algorithms was introduced in Fisher and Van Ness[1] and Van Ness[2] (see Definition 2 below). The concept of monotone admissibility is based on the following idea: in many clustering situations, the relative scale of the variables is not particularly meaningful and a strictly monotone transformation of the distances between data points should, therefore, not affect the resulting clustering. In the present paper, necessary and sufficient conditions for a broad class of algorithms, called the Lance and Williams algorithms (see below), to be monotone admissible are discussed. The family of Lance and Williams algorithms is important since it includes most of the commonly used agglomerative clustering algorithms as special cases (see Table 1).

The discussion will be limited to agglomerative clustering algorithms. An agglomerative clustering algorithm is a hierarchical algorithm (or sequence of clusterings or tree of clusterings) which starts, at stage 1, with each data point as its own cluster. It is assumed that a distance function, d_1, is defined which specifies the distance between any two data points. At stage $t+1$ the algorithm combines the two closest stage t clusters (if there are ties for the closest clusters, an arbitrary pair of closest clusters can be combined). After two clusters are combined, the distance, d_{t+1}, between the newly combined cluster and each of the remaining clusters must be defined. The distances among the clusters which were not combined remain the same; that is, if $C_i^{(t)} = C_i^{(t+1)}$ and $C_j^{(t)} = C_j^{(t+1)}$ then $d_{t+1}(C_i^{(t+1)}, C_j^{(t+1)}) = d_t(C_i^{(t)}, C_j^{(t)})$. One agglomerative algorithm is distinguished from another by the way in which it defines this new distance, d_{t+1}, at each stage.

Lance and Williams[3] proposed a general parametric formula to define the new distances at each stage. Suppose that $C_i^{(t)}$ and $C_j^{(t)}$ are the clusters which are combined at stage $t+1$ and that $C_k^{(t)}$ is any other cluster at stage t. The Lance and Williams distance of order $r > 0$ defines

$$\begin{aligned} d_{t+1}^r(C_i^{(t)} \cup C_j^{(t)}, C_k^{(t)}) &= \alpha(m_i) d_t^r(C_i^{(t)}, C_k^{(t)}) \\ &+ \alpha(m_j) d_t^r(C_j^{(t)}, C_k^{(t)}) + \beta(m_i, m_j)\, d_t^r(C_i^{(t)}, C_j^{(t)}) \\ &+ \gamma |d_t^r(C_i^{(t)}, C_k^{(t)}) - d_t^r(C_j^{(t)}, C_k^{(t)})|, \end{aligned} \tag{1}$$

where $m_i = \|C_i^{(t)}\| / \|C_i^{(t)} \cup C_j^{(t)}\|$, $m_j = \|C_j^{(t)}\| / \|C_i^{(t)} \cup C_j^{(t)}\|$, $\|C\|$ is the number of points in cluster C, α is a bounded positive function on the rationals in the interval (0, 1), β is a bounded function on the rationals in (0, 1), γ is a constant, and r just an ordinary positive exponent (usually $r = 1$ or 2).

Definition 1. An agglomerative algorithm, using distance (1), is called an order r Lance and Williams algorithm and will be denoted LW ($\alpha, \beta, \gamma; r$).

It should be mentioned that the distance defined in equation (1) could be negative. Negative distances are usually unacceptable in practice, however, and some restrictions need to be placed on α, β, and γ to ensure that the distance defined in equation (1) is always nonnegative. Chen and Van Ness[4] showed that equation (1) is always nonnegative for any data set

* Author to whom correspondence should be addressed.

Table 1. Common Lance and Williams algorithms

Algorithm	$\alpha(u)$	$\beta(u,1-u)$	γ	Order
Nearest-neighbor	$\frac{1}{2}$	0	$-\frac{1}{2}$	$r \geq 1$
Farthest-neighbor	$\frac{1}{2}$	0	$\frac{1}{2}$	$r \geq 1$
Average linkage	u	0	0	1
Group average	u	0	0	2
Median	$\frac{1}{2}$	$-\frac{1}{4}$	0	2
Centroid	u	$-u(1-u)$	0	2
Flexible method	$(1-\beta)/2$	$\beta < 1$	0	2

with nonnegative original distances if and only if

(1) $\alpha(u) + \alpha(1-u) + \beta(u, 1-u) \geq 0$, and
(2) $\alpha(u) + \gamma \geq 0$

for any rational $0 < u < 1$. In this paper, it will be assumed that (1) and (2) hold.

Each combination of α, β, and γ in (1) represents a different agglomerative clustering algorithm. So here we have an infinite number of clustering algorithms, which will cluster any data set with a distance defined on it. Thus, we see how important it is to have criteria for selecting from the myriad of available clustering algorithms. The concept of clustering algorithm admissibility was invented as such a criterion. There are many clustering admissibility conditions [see reference (2)]. A user can restrict attention to algorithms satisfying those admissibility conditions that are important to his or her application and thereby eliminate many obviously "unreasonable" clustering algorithms for that application. Monotone admissibility is one of the proposed admissibility conditions [see reference (1)].

Definition 2. A clustering algorithm is said to be monotone admissible if it has the property that, for any data set with a distance function defined on it, a monotone transformation applied to the distance matrix at any stage does not change the resulting clustering or sequence of clusterings. Here a monotone transformation means a strictly increasing function g with $g(0) = 0$.

It might seem that this is a reasonable criterion for selecting clustering algorithms in some applications. It will be shown in this paper, however, that the family of monotone admissible LW algorithms is very limited, only containing the nearest-neighbor and the farthest-neighbor algorithms.

2. NECESSARY AND SUFFICIENT CONDITIONS FOR MONOTONE ADMISSIBILITY

It is assumed that the distance d_1 satisfies

(1) $d_1(x, x) = 0$ for all x;
(2) $d_1(x, y) = d_1(y, x) \geq 0$ for all x and y; and
(3) $d_1(x, y) = 0 \Rightarrow d_1(x, z) = d_1(y, z)$ for all x, y and z.

It is important to notice that in order for an algorithm to be monotone admissible it must be so for any data set and any initial distance function, d_1. Thus, in the proofs of the necessary conditions below, we can choose the data set to be anything we want. We will pick Euclidean data with the Euclidean metric.

Lemma 3. Suppose $C_i = \{x, x, \ldots, x\}$ is a cluster consisting of p identical points, $C_j = \{x, x, \ldots, x\}$ is a cluster consisting of q replications of the same points, and y is any point with distance d_0 from x. Suppose also that the LW algorithm is monotone admissible. Then for the appropriate t,

(1) $d_t(C_i, C_j) = 0$, and
(2) $d_t(C_i \cup C_j, \{y\}) = d_0$.

Proof. First, consider the case $p = 1$, $q = 1$. By equation (1) and above assumption (1) we immediately get that

$$d(C_i, C_j) = d(\{x\}, \{x\}) = 0,$$

where we have dropped the subscript, t, on d. We will prove

$$d(C_i \cup C_j, \{y\}) = d(\{x, x\}, \{y\}) = d_0.$$

This is the same as proving $2\alpha(\frac{1}{2}) = 1$. For contradiction, first assume $2\alpha(\frac{1}{2}) = \eta > 1$. Consider another data set consisting of four clusters on the real line: C_1 with one point at 0, C_2 with one point at 0, C_3 with one point at 1 and C_4 with one point at $1 + \delta$, where $\eta < \delta < \infty$. The first step in performing clustering is to combine two clusters C_1 and C_2. The distance between C_3 and $C_1 \cup C_2$ is $[2\alpha(\frac{1}{2})]^{1/r}$. Then the two-cluster clustering of the data set is $\{\{C_1, C_2, C_3\}, \{C_4\}\}$. Because the function $g(x) = x^{1/m}$ is strictly increasing in $x > 0$ for any $m > 0$, and the clustering algorithm is monotone admissible,

$$[2\alpha(\tfrac{1}{2})]^{1/r} < \delta^{1/m}, \quad m = 1, 2, \ldots$$

and $[2\alpha(\frac{1}{2})]^{1/r} \leq \lim_{m \to \infty} \delta^{1/m} = 1$. To prove $2\alpha(\frac{1}{2}) \geq 1$, note that there exists $0 < \eta < 1$ such that $[2\alpha(\frac{1}{2})]^{1/r} < \eta$ because $\alpha(\frac{1}{2}) > 0$. Consider a data set consisting of four clusters on the real line: C_1 with one point at 0, C_2 with one point at 0, C_3 with one point at 1, and C_4 with one point at $1 + \eta$. The distance between C_3 and $C_1 \cup C_2$ is $[2\alpha(\frac{1}{2})]^{1/r}$. Then the two-cluster clustering of the data set is $\{\{C_1, C_2\}, \{C_3, C_4\}\}$. Again apply the monotone transformation $g(x) = x^{1/m}$. Since the clustering algorithm is monotone admissible,

$$[2(\tfrac{1}{2})]^{1/r} > \eta^{1/m}, \quad m = 1, 2, \ldots$$

and $[2\alpha(\frac{1}{2})]^{1/r} \geq \lim_{m \to \infty} \eta^{1/m} = 1$. Thus, $2\alpha(\frac{1}{2}) = 1$.

For the case of $p = 2$, $q = 1$, note that

$$d(C_i, C_j) = d(\{x, x\}, \{x\}) = 0.$$

We will prove

$$d(C_i \cup C_j, \{y\}) = d(\{x, x, x\}, \{y\}) = d_0.$$

This is the same as proving $\alpha(\frac{2}{3}) + \alpha(\frac{1}{3}) = 1$.

For contradiction, first assume $\alpha(\frac{2}{3}) + \alpha(\frac{1}{3}) = \eta > 1$. Consider a data set, which is similar to the one used for the case $p = 1$ and $q = 1$, consisting of four clusters on the real line: C_1 with two points at 0, C_2 with one point at 0, C_3 with one point at 1 and C_4 with one point at $1 + \delta$, where $\eta < \delta < \infty$. The same argument can be used to produce a contradiction. To prove $\alpha(\frac{2}{3}) + \alpha(\frac{1}{3}) \geq 1$, note that there exists $0 < \eta < 1$ such that $\alpha(\frac{2}{3}) + \alpha(\frac{1}{3}) > \eta$ because $\alpha(\frac{2}{3}), \alpha(\frac{1}{3}) > 0$. Consider a data set consisting of four clusters on the real line: C_1 with two points at 0, C_2 with one point at 0, C_3 with one point at 1, and C_4 with one point at $1 + \eta$. The same technique can be used to verify the validity of the statement $\alpha(\frac{2}{3}) + \alpha(\frac{1}{3}) \geq 1$. Thus, $\alpha(\frac{2}{3}) + \alpha(\frac{1}{3}) = 1$. The general result follows by finite induction. □

Lemma 4. If an LW$(\alpha, \beta, \gamma; r)$ algorithm is monotone admissible, then

$$\alpha(u) + \alpha(1-u) = 1 \tag{2}$$

for any rational $0 < u < 1$.

Proof. The proof is similar to that of Lemma 3. □

Lemma 5. If an LW$(\alpha, \beta, \gamma; r)$ algorithm is monotone admissible, then

$$\beta(u, 1-u) = 0 \tag{3}$$

for any rational $0 < u < 1$.

Proof. Suppose that $0 < u < 1$ is an arbitrary rational. Then there exist positive integers p and q such that $u = p/q$, and that p and q are relatively prime. Since $\alpha(u) + \alpha(1-u) = 1$, $|\beta(u, 1-u)| < \infty$, there exists an $\eta > 1$ such that

$$\begin{aligned}&[\alpha(u) + \alpha(1-u) + \beta(u, 1-u)]^{1/r}\\ &= [1 + \beta(u, 1-u)]^{1/r} < \eta.\end{aligned}$$

To prove $\beta(u, 1-u) \leq 0$, consider a data set consisting of four clusters: C_1 with p points at $(0, \frac{1}{2})$, C_2 with $q - p$ points at $(0, -\frac{1}{2})$, C_3 with one point at $(\sqrt{3}/2, 0)$, and C_4 with one point at $(\sqrt{3}/2 + \eta, 0)$ where $\eta > 1$. Then

$$\begin{aligned}d_{q-1}(C_1, C_2) &= d_{q-1}(C_1, C_3) = d_{q-1}(C_2, C_3)\\ &= 1 < \eta = d_{q-1}(C_2, C_4).\end{aligned}$$

The qth step in performing clustering could be to combine C_1 and C_2. The distance between C_3 and $C_1 \cup C_2$ is $[1 + \beta(u, 1-u)]^{1/r}$. Then the two-cluster clustering of the data set is $\{\{C_1, C_2, C_3\}, \{C_4\}\}$. Again use $g(x) = x^{1/m}$, because the clustering algorithm is monotone admissible,

$$[1 + \beta(u, 1-u)]^{1/r} < \eta^{1/m}, \quad m = 1, 2, \ldots .$$

Thus

$$[1 + \beta(u, 1-u)]^{1/r} \leq \lim_{m \to \infty} \eta^{1/m} = 1.$$

This implies $\beta(u, 1-u) \leq 0$.

To prove $\beta(u, 1-u) \geq 0$, consider a data set consisting of four clusters: C_1 with p points at $(0, \varepsilon)$, C_2 with $q - p$ points at $(0, -\varepsilon)$, C_3 with one point at $((1-\varepsilon^2)^{0.5}, 0)$ and C_4 with one point at $((1-\varepsilon^2)^{0.5} + \eta, 0)$ where ε is a small positive number and $\eta > 0$. Previously, it was shown that $\beta(u, 1-u) \leq 0$. Since $\beta(u, 1-u) > -\infty$, for any $0 < \eta < 1$, ε can be selected such that $0 < 2\varepsilon < \eta < [1 + (2\varepsilon)^r \beta(u, 1-u)]^{1/r} \leq 1$. The distance between C_3 and $C_1 \cup C_2$ is $[1 + (2\varepsilon)^r \beta(u, 1-u)]^{1/r}$. Then the two-cluster clustering of the data set is $\{\{C_1, C_2\}, \{C_3, C_4\}\}$. Again consider the transformation $g(x) = x^{1/m}$, because the clustering algorithm is monotone admissible,

$$[1 + (2\varepsilon)^{r/m} \beta(u, 1-u)]^{1/r} > \eta^{1/m}, \quad m = 1, 2, \ldots$$

and

$$\lim_{m \to \infty} [1 + \beta(u, 1-u)(2\varepsilon)^{r/m}]^{1/r} \geq \lim_{m \to \infty} \eta^{1/m}.$$

Thus, $\beta(u, 1-u) \geq 0$. □

Lemma 6. If an LW$(\alpha, \beta, \gamma; r)$ algorithm is monotone admissible, then

$$|\gamma| \leq \min\{\alpha(u), \alpha(1-u)\} \tag{4}$$

for any rational $0 < u < 1$.

Proof. Suppose that $0 < u < 1$ is an arbitrary rational. Let p and q be positive integers such that $u = p/q$, and that p and q are relatively prime. To prove $\gamma \leq \min\{\alpha(u), \alpha(1-u)\}$, select four points 0, $0.5 - \varepsilon$, $1 - \varepsilon$, and $2 - \varepsilon$ on the real line where $0 < \varepsilon < 0.5$. Consider a data set consisting of four clusters: C_1 with p points at 0, C_2 with $q - p$ points at $0.5 - \varepsilon$, C_3 with one point at $1 - \varepsilon$, and C_4 with one point at $2 - \varepsilon$. After combining clusters C_1 and C_2, the distance between C_3 and $C_1 \cup C_2$ is

$$[(1-\varepsilon)^r \alpha(u) + 0.5^r \alpha(1-u) + ((1-\varepsilon)^r - 0.5^r)\gamma]^{1/r}.$$

By Lemma 4, the distance between C_3 and $C_1 \cup C_2$ is

$$[(1-\varepsilon)^r + ((1-\varepsilon)^r - 0.5^r)(\gamma - \alpha(1-u))]^{1/r}.$$

Assume for contradiction that $\gamma > \alpha(1-u)$. Then ε can be selected so small that

$$[(1-\varepsilon)^r + ((1-\varepsilon)^r - 0.5^r)(\gamma - \alpha(1-u))]^{1/r} > 1.$$

Then the two-cluster clustering of the data set is $\{\{C_1, C_2\}, \{C_3, C_4\}\}$. Apply $g(x) = x^m$, $m > 0$, to the distance matrix, the distance between C_3 and $C_1 \cup C_2$ is

$$[(1-\varepsilon)^{mr} \alpha(u) + 0.5^{mr} \alpha(1-u) + ((1-\varepsilon)^{mr} - 0.5^{mr})\gamma]^{1/r}.$$

When m is sufficiently large, the distance between C_3 and $C_1 \cup C_2$ is very close to 0, and the distance between C_3 and C_4 is still 1. Then the two-cluster clustering of the data set is $\{\{C_1, C_2, C_3\}, \{C_4\}\}$. This contradicts the assumption that the clustering algorithm is monotone admissible. Then $\gamma \leq \alpha(1-u)$. Similarly, $\gamma \leq \alpha(u)$. Thus $\gamma \leq \min\{\alpha(u), \alpha(1-u)\}$.

Combining this with the restriction (ii) in Section 1, the result follows. □

Lemma 7. If an $LW(\alpha, \beta, \gamma; r)$ algorithm is monotone admissible, then, for any rational $0 < u < 1$, γ cannot be in the interval $(0, \min\{\alpha(u), \alpha(1-u)\})$.

Proof. Suppose that $0 < u < 1$ is an arbitrary rational. Then there exist positive integers p and q such that $u = p/q$, and that p and q are relatively prime. Assume, for contradiction, that

$$0 < \gamma < \min\{\alpha(u), \alpha(1-u)\}.$$

Select four points 0, ε, 1, and $1+\eta$ on the real line where $0 < \varepsilon < 1-\varepsilon < 1$, $\varepsilon < \eta < 1$. Consider a data set consisting of four clusters: C_1 with p points at 0, C_2 with $q-p$ points at ε, C_3 with one point at 1, and C_4 with one point at $1+\eta$. After combining clusters C_1 and C_2, the distance between C_3 and $C_1 \cup C_2$, by Lemma 5, is

$$[\alpha(u) + (1-\varepsilon)^r \alpha(1-u) + (1-(1-\varepsilon)^r)\gamma]^{1/r}, \quad (5)$$

which, by Lemma 4, is

$$[1 + (1-(1-\varepsilon)^r)(\gamma - \alpha(1-u))]^{1/r}.$$

Since $\gamma < \alpha(1-u)$, then ε and η can be selected such that

$$0 < [1 + (1-(1-\varepsilon)^r)(\gamma - \alpha(1-u))]^{1/r} < \eta.$$

Then the two-cluster clustering of the data set is $\{\{C_1, C_2, C_3\}, \{C_4\}\}$. Apply $g(x) = x^m$, $m > 0$, to the distance matrix, then the distance between C_3 and $C_1 \cup C_2$ is

$$[\alpha(u) + (1-\varepsilon)^{mr}\alpha(1-u) + (1-(1-\varepsilon)^{mr})\gamma]^{1/r}.$$

When m is sufficiently large, the distance between C_3 and $C_1 \cup C_2$ is very close to $[\alpha(u) + \gamma]^{1/r}$, and the distance between C_3 and C_4 is very close to 0. Because $\alpha(u) + \gamma > 0$, the two-cluster clustering of the data set is $\{\{C_1, C_2\}, \{C_3, C_4\}\}$. This contradicts the assumption that the clustering algorithm is monotone admissible. □

Lemma 8. If an $LW(\alpha, \beta, \gamma; r)$ algorithm is monotone admissible, then, for any rational $0 < u < 1$, γ cannot be in the interval $(-\min\{\alpha(u), \alpha(1-u)\}, 0)$.

Proof. Assume, for contradiction, that

$$-\min\{\alpha(u), \alpha(1-u)\} < \gamma < 0.$$

Consider the same data set as in the proof of Lemma 7. By equation (5) and Lemma 4, the distance between C_3 and $C_1 \cup C_2$ is

$$[(1-\varepsilon)^r + (1-(1-\varepsilon)^r)(\gamma + \alpha(u))]^{1/r} \equiv (h_{r,\gamma,\alpha(u)}(\varepsilon))^{1/r}.$$

Because $(d/d\varepsilon)\, h_{r,\gamma,\alpha(u)}(\varepsilon) \le 0$, $h_{r,\gamma,\alpha(u)}(\varepsilon)$ is decreasing in $\varepsilon > 0$. Note that $\lim_{\varepsilon \to 0} h_{r,\gamma,\alpha(u)}(\varepsilon) = 1$ and that

$$\begin{aligned} \lim_{\varepsilon \to 0.5} h_{r,\gamma,\alpha(u)}(\varepsilon) &= 0.5^r + (1-0.5^r)(\gamma + \alpha(u)) \\ &< 0.5^r + (1-0.5^r)\alpha(u) \\ &\le 0.5^r + (1-0.5^r) = 1. \end{aligned}$$

Then ε and η can be selected such that

$$0 < [(1-\varepsilon)^r + (1-(1-\varepsilon)^r)(\gamma + \alpha(u))]^{1/r} < \eta < 1.$$

Then the two-cluster clustering of the data set is $\{\{C_1, C_2, C_3\}, \{C_4\}\}$. Apply $g(x) = x^m$, $m > 0$, to the distance matrix, then the distance between C_3 and $C_1 \cup C_2$ is

$$[\alpha(u) + (1-\varepsilon)^{mr}\alpha(1-u) + (1-(1-\varepsilon)^{mr})\gamma]^{1/r}.$$

When m is sufficiently large, the distance between C_3 and $C_1 \cup C_2$ is very close to $(\alpha(u) + \gamma)^{1/r}$, and the distance between C_3 and C_4 is very close to 0. Because $\alpha(u) + \gamma > 0$, the two-cluster clustering of the data set is $\{\{C_1, C_2\}, \{C_3, C_4\}\}$. This contradicts the assumption that the clustering algorithm is monotone admissible. □

Lemma 9. If an $LW(\alpha, \beta, \gamma; r)$ algorithm is monotone admissible, then $\gamma \ne 0$.

Proof. Assume that $\gamma = 0$. Select four points 0, ε, 1, and $1+\eta$ on the real line where $0 < \varepsilon < 1-\varepsilon < 1$, $0 < \varepsilon < \eta < 1$. Consider a data set consisting of four clusters: C_1 with one points at 0, C_2 with one points at ε, C_3 with one point at 1, and C_4 with one point at $1+\eta$. After combining clusters C_1 and C_2, the distance between C_3 and $C_1 \cup C_2$ is

$$[\alpha(0.5) + (1-\varepsilon)^r \alpha(0.5)]^{1/r} < [\alpha(0.5) + \alpha(0.5)]^{1/r} = 1.$$

Then η can be selected such that

$$[\alpha(0.5) + (1-\varepsilon)^r \alpha(0.5)]^{1/r} < \eta < 1.$$

Then the two-cluster clustering of the data set is $\{\{C_1, C_2, C_3\}, \{C_4\}\}$. Apply $g(x) = x^m$, $m > 0$, to the distance matrix, the distance between C_3 and $C_1 \cup C_2$ is

$$[\alpha(0.5) + (1-\varepsilon)^{mr} \alpha(0.5)]^{1/r}.$$

When m is sufficiently large, the distance between C_3 and $C_1 \cup C_2$ is very close to $\{\alpha(0.5)\}^{1/r}$, and the distance between C_3 and C_4 is very close to 0. Because $\{\alpha(0.5)\}^{1/r} > 0$, the two-cluster clustering of the data set is $\{\{C_1, C_2\}, \{C_3, C_4\}\}$. This contradicts the assumption that the clustering algorithm is monotone admissible. □

Combine Lemmas 3–9, the necessary and sufficient conditions for an agglomerative algorithm to be monotone admissible are listed in the following theorem.

Theorem 10. An LW$(\alpha, \beta, \gamma; r)$ algorithm is monotone admissible if and only if

(i) $\alpha(u) + \alpha(1 - u) = 1$.
(ii) $\beta(u, 1 - u) = 0$.
(iii) $|\gamma| = \min\{\alpha(u), \alpha(1 - u)\}$.

for any rational $0 < u < 1$.

Proof. The necessity part of the theorem is the combination of Lemmas 3–9. On the other hand, if an agglomerative clustering algorithm LW$(\alpha, \beta, \gamma; r)$ satisfies the above conditions (i)–(iii), then $\alpha(u) = \alpha(1 - u) = 0.5$, $\beta(u, 1 - u) = 0$, $\gamma = \pm 0.5$ for any $0 < u < 1$. Then the clustering algorithm is either the nearest-neighbor algorithm or the farthest-neighbor algorithm, and hence is monotone admissible. □

Corollary 11. An LW$(\alpha, \beta, \gamma; r)$ algorithm is monotone admissible if and only if the algorithm is either the nearest neighbor algorithm or the farthest neighbor algorithm.

3. CHARACTERIZATION OF NEAREST- AND FARTHEST-NEIGHBOR ALGORITHMS

The necessary and sufficient conditions for monotone admissible agglomerative clustering algorithms reveal that the only monotone admissible LW algorithms are the nearest-neighbor algorithm and the farthest- neighbor algorithm. This shows that the family of monotone admissible LW clustering algorithms is very limited. In this sense, monotone admissibility characterizes the nearest-neighbor and the farthest- neighbor algorithms. Chen and Van Ness[(5)] proposed another clustering admissibility condition, metric admissibility.

Definition 12. An agglomerative clustering algorithm is called metric admissible if the distances between the clusters generated at each stage, t, during clustering satisfy the axioms of a metric for any data whose original distances between the clusters are metric, or satisfy the axioms of a pseudo-metric for any data whose original distances between the clusters are pseudo-metric.

Using this, the nearest-neighbor and the farthest-neighbor algorithms can be characterized. It can be shown that the nearest-neighbor algorithm is not metric admissible, while the farthest-neighbor algorithm is metric admissible. So the following theorem holds.

Theorem 13. (i) An LW$(\alpha, \beta, \gamma; r)$ algorithm is the nearest-neighbor algorithm if and only if it is monotone admissible but not metric admissible. (ii) An LW$(\alpha, \beta, \gamma; r)$ algorithm is the farthest-neighbor algorithm if and only if it is monotone admissible and metric admissible.

Proof. Suppose an LW$(\alpha, \beta, \gamma; r)$ algorithm is the nearest-neighbor algorithm. By Corollary 11, the algorithm is monotone admissible. It has been shown [see reference (5)] that the nearest-neighbor clustering algorithm is not metric admissible. Thus, the necessity part of (i) holds. On the other hand, let LW$(\alpha, \beta, \gamma; r)$ be monotone admissible but not metric admissible. By Corollary 11, the algorithm must be either the nearest-neighbor or farthest-neighbor algorithm. Because the farthest-neighbor algorithm is metric admissible, the algorithm must be the nearest-neighbor algorithm. Thus, (i) holds. Part (ii) of the theorem can be proved similarly. □

4. CONCLUSIONS

There is a confusion of possible clustering algorithms. Clustering admissibility is one of the few objective criteria for selection among these algorithms. There are many clustering situations in which the relative scale of variables is meaningless (for example, blood pressure and age). One is tempted in these situations to seek a clustering algorithm which does not depend on the scale of the distance function, that is the algorithm is scale invariant. This gives rise to the monotone admissibility condition—one of several clustering admissibility conditions [see reference (2)]. This article states that if the algorithm is a Lance and Williams algorithm, then scale invariance implies that one must use either the nearest-neighbor or the farthest-neighbor algorithm. Both nearest-neighbor and farthest-neighbor are legitimate-clustering algorithms, but both are known to have serious shortcomings. Thus, one might conclude, in many applications, that scale invariance is too much to ask for because it leads to such a limited range of clustering algorithms.

A quite reasonable clustering admissibility condition is metric admissibility [reference (5)]. If one requires both monotone admissibility and metric admissibility, then the only Lance and Williams clustering algorithm satisfying both requirements is farthest-neighbor.

REFERENCES

1. L. Fisher and J. Van Ness, Admissible clustering procedures, *Biometrika* **58,** 91–104 (1971).
2. J. Van Ness, Admissible clustering procedures II, *Biometrika*, **60,** 422–424 (1973).
3. G. Lance and W. Williams, A general theory of classificatory sorting strategies. 1. Hierarchical systems, *Comput. J.* **9,** 373–380 (1967).
4. Z. Chen and J. Van Ness, Space-contracting, space-dilating, and positive admissible clustering algorithms, *Pattern Recognition* **27,** 853–857 (1994).
5. Z. Chen and J. Van Ness, Metric admissibility and agglomerative clustering, *Commun. Statist. Simulation Comput.* **23,** 833–845 (1994).

About the Author—Zhenmin Chen is an Assistant Professor of Statistics, Department of Statistics, Florida International University. He received his BS in mathematics from Shanghai Teachers University and PhD in statistics from University of Texas at Dallas. His research interests include: applied statistics, reliability, and clustering analysis.

About the Author—John Van Ness received his BS in mechanical engineering from Northwestern University in 1959 and his PhD in applied mathematics from Brown University in 1964. He has served as Visiting Assistant Professor, Department of Statistics, Stanford University, 1964–1966; Assistant Professor, Department of Mathematics, University of Washington, Seattle, 1966–1971; Associate Professor, Department of Statistics, Carnegie-Mellon University, 1971–1973; and Associate Professor and Professor, Programs in Mathematical Sciences, UT-Dallas, from 1973 to the present. He also served as Head, Programs in Mathematical Sciences, UT-Dallas, 1973–1978 and 1980–1992; and Associate Dean, School of Natural Sciences and Mathematics, UT-Dallas, 1978–1983. He is a Fellow, American Statistical Association, and Fellow, Institute of Mathematical Statistics. His research interests include non-standard and robust stochastic models, measurement error models, classification, and pattern recognition and areas of application include physical and medical sciences and engineering.