Scientific Annals of Computer Science vol. 25 (1), 2015, pp. 155–170
doi: 10.7561/SACS.2015.1.155

Several Remarks on Dissimilarities and Ultrametrics

Dan A. Simovici[1]

Abstract

We investigate the relationships between tolerance relations, equivalence relations, and ultrametrics. The set of spheres associated to an ultrametric space has a tree structure that reflects a hierarchy on the set of equivalences associated to that space. We show that every ultrametric defined on a finite space is a linear combination of binary ultrametric and we introduce the notion of ultrametricity for dissimilarities, which has applications in many data mining problems.

Keywords: tolerance, equivalence, ultrametric, ultrametricity, tree of spheres

1 Introduction

Dissimilarity spaces constitute the natural framework for a number of exploratory techniques in machine learning and data mining such as certain classification methods and clustering algorithms. We examine relationships that exist between various types of dissimilarities and focus on ultrametrics.

Ultrametrics are dissimilarities that satisfy a stronger version of the triangular inequality (usually associated with metrics) and they occur in many data mining applications such as agglomerative hierarchical clustering algorithms [4, 5, 2, 6], and have applications in the study of phylogenetic trees in biology [10, 7], p-adic numbers in mathematics [12, 1], and certain physical systems [11], etc.

[1] University of Massachusetts Boston, Department of Computer Science, Boston, Massachusetts, E-mail: dsim@cs.umb.edu

We evaluate the extent of the difference between dissimilarities and ultrametrics defined on the same set by introducing the notion of ultrametricity of a dissimilarity. We show that dissimilarities can be modified to increase or decrease their level of ultrametricity. An increase in ultrametricity has an equalizing effect on dissimilarities and can be useful in certain clustering algorithms; a decrease in ultrametricity is interesting for other data mining applications such as the k-nearest neighbor technique and in outlier detection, as we have shown in [14].

The set of real numbers is denoted by $\mathbb{R}$; the set of non-negative reals is denoted by $\mathbb{R}_{\geqslant 0}$. Every other set considered in below is finite.

A *quasi-dissimilarity* on a set S is a function $d : S \times S \longrightarrow \mathbb{R}$ such that $d(x, y) \geqslant 0$, $d(x, x) = 0$, and $d(x, y) = d(y, x)$ for every $x, y \in X$. We assume that all dissimilarity spaces considered here are finite.

A quasi-dissimilarity d is a *dissimilarity* if $d(x, y) = 0$ implies $x = y$.

A quasi-dissimilarity d is a *quasi-metric* if it satisfies the triangular inequality:

$$d(x, y) \leqslant d(x, z) + d(z, y). \tag{1}$$

In addition, if $d(x, y) = 0$ implies $x = y$, then d is a *metric*. Inequality (1) is known as the *triangular inequality*.

A *quasi-ultrametric* is a quasi-dissimilarity $d : S \times S \longrightarrow \mathbb{R}_{\geqslant 0}$ that satisfies the inequality

$$d(x, y) \leqslant \max\{d(x, z), d(z, y)\} \tag{2}$$

for every $x, y, z \in S$. If, in addition, $d(x, y) = 0$ implies $x = y$, then d is an *ultrametric*.

In Section 2 we discuss the link between ultrametrics and equivalence relations and we include some preliminary results. In Section 3 we examine properties of the collection of spheres of an ultrametric space. The relationship between multivalued and binary ultrametrics is the object of Section 4. In Section 5 we introduce the notion of ultrametricity of dissimilarities. Finally, we present conclusions in Section 6.

2 Ultrametrics and Equivalences

A *tolerance* on the set S is a relation $\theta \subseteq S \times S$ that is reflexive and symmetric. In other words, $(x, x) \in \theta$ for every $x \in S$ and $(x, y) \in \theta$ if and only if $(y, x) \in \theta$ for $x, y \in S$. The set of tolerances on S is denoted by

$\mathsf{TOL}(S)$. A tolerance on S that is transitive (that is, $(x, y), (y, z) \in \theta$ imply $(x, z) \in \theta$) is an equivalence. The set of equivalences on S is denoted by $\mathsf{EQ}(S)$.

Let d be a quasi-dissimilarity on the set S. It is immediate that the relation

$$\theta_{d,r} = \{(x, y) \mid d(x, y) \leqslant r\}$$

is a tolerance on S for every $r \in \mathbb{R}$.

If d is an ultrametric on S, the relation $\theta_{d,r}$ is an equivalence on S for any $r \in \mathbb{R}_{\geqslant 0}$. Indeed, if $(x, y), (y, z) \in \theta_{d,r}$, then $d(x, y) \leqslant r$ and $d(y, z) \leqslant r$. Therefore,

$$d(x, z) \leqslant \max\{d(x, y), d(y, z)\} \leqslant r$$

because of the ultrametric inequality. Thus, $(x, z) \in \theta_{d,r}$, which proves that $\theta_{d,r}$ is transitive, so it is an equivalence.

The equivalence relations α_S and ω_S on a set S are defined by:

$$\alpha_S = \{(x, x) \mid x \in S\} \text{ and } \omega_S = S \times S.$$

Let $d : S \times S \longrightarrow \mathbb{R}_{\geqslant 0}$ be a dissimilarity on S whose range is $\{0, 1\}$. We designate such functions as *binary dissimilarities*. Note that $d(x, y) \leqslant \max\{d(x, z), d(z, y)\}$ for $x, y, z \in S$. Indeed, if $d(x, y) = 0$, the ultrametric inequality is clearly satisfied. If $d(x, y) = 1$, then $x \neq y$ (since, otherwise we would have $d(x, y) = 0$). Thus, any $z \in S$ must be distinct either from x or from y, so at least one of the numbers $d(x, z), d(z, y)$ is non-zero, and the ultrametric inequality is satisfied. We conclude that every binary dissimilarity is an ultrametric.

Let θ be an equivalence on S and let d_θ be defined as the characteristic function of θ as

$$d_\theta(x, y) = \begin{cases} 0 & \text{if } (x, y) \in \theta, \\ 1 & \text{otherwise,} \end{cases}$$

for $x, y \in S$. Then, d_θ is a quasi-ultrametric. Indeed, since $(x, x) \in \theta$ it follows that $d_\theta(x, x) = 0$. If $d(x, y) = 1$, then $(x, y) \notin \theta$, so for every $z \in S$ we have $(x, z) \notin \theta$ or $(z, y) \notin \theta$. Thus, we have $\max\{d(x, z), d(z, y)\} = 1$, so the ultrametric inequality is satisfied by x, y, z.

In particular, for d_{α_S} we have

$$d_{\alpha_S}(x, y) = \begin{cases} 0 & \text{if } x = y, \\ 1 & \text{if } x \neq y \end{cases}$$

for $x, y \in S$.

If d is an ultrametric, the binary ultrametric $d_{\theta_{d,r}}$ is given by

$$d_{\theta_{d,r}}(x,y) = \begin{cases} 0 & \text{if } d(x,y) \leqslant r, \\ 1 & \text{otherwise} \end{cases}$$

for $x, y \in S$.

Dissimilarities are closely related to certain families of tolerances.

Let R be a subset of $\mathbb{R}_{\geqslant 0}$ and let $\beta : R^2 \longrightarrow R$ be an associative operation on R. A *β-directed family of tolerances* on S is a collection $\mathcal{T}_\beta = \{\theta_r \in \mathsf{TOL}(S) \mid r \in R\}$, where $\theta_r, \theta_s \in \mathcal{T}$ imply $\theta_r \theta_s \subseteq \theta_{\beta(r,s)} \in \mathcal{T}$. $\mathcal{T}_\beta$ is a bounded family if there exists $r \in R$ such that $\theta_r = \omega_S$.

Example 2.1. Let $d : S \times S \longrightarrow \mathbb{R}_{\geqslant 0}$ be a metric on the finite set S. Suppose that the range of d is $\text{Ran}(d) = \{r_1, \ldots, r_m\}$ and let $\theta_{d,r}$ be the tolerance

$$\theta_{d,r} = \{(x,y) \mid d(x,y) \leqslant r\}.$$

If $r_1 < r_2 < \cdots < r_m$ the collection $\{\theta_{d,r} \mid r \in \text{Ran}(d)\}$ is a directed β-family of tolerances, where $\beta(a,b) = a + b$. Indeed, if $(x,y) \in \theta_{d,r}\theta_s$, there exists $t \in S$ such that $(x,t) \in \theta_{d,r}$, $(t,y) \in \theta_{d,s}$, that is, $d(x,t) \leqslant r$ and $d(t,y) \leqslant s$, which implies $d(x,y) \leqslant r + s$. Thus, $(x,y) \in \theta_{d,r+s}$.

Conversely, if $\{\theta_r \mid r \in R\}$ is a β-family of tolerances, where β is defined as above, then $d : S \times S \longrightarrow \mathbb{R}_{\geqslant 0}$ defined by $d(x,y) = \min\{r \in R \mid d(x,y) \in R\}$ is a metric on S.

Indeed, if $d(x,z) = a$ and $d(z,y) = b$, then $(x,z) \in \theta_a$, $(z,y) \in \theta_b$, so $(x,y) \in \theta_a\theta_b \subseteq \theta_{a+b}$, which implies $d(x,y) \leqslant a + b = d(x,z) + d(z,y)$

Similarly, if β is replaced by $\beta(a,b) = \max\{a,b\}$, then any β-directed family of tolerances defines an ultrametric, and every ultrametric can be obtained in this manner. □

3 Spheres in Ultrametric Spaces

Let (S,d) be a dissimilarity space. The *closed sphere centered in $\boldsymbol{x}_0$ and having radius r* is the set

$$B_d(x_0, r) = \{x \in X \mid d(x_0, x) \leqslant r\}.$$

A *triangle* in a dissimilarity space (S,d) is a triple $(x,y,z) \in S^3$. To simplify the notation, we denote $t = (x,y,z)$ by xyz.

The following properties of an ultrametric space (S, d) are well-known (see [13]):

(i) for every triangle $t = xyz \in S^3$, two of the numbers $d(x, y), d(x, z), d(z, y)$ are equal and the third is not larger than the largest two numbers; thus, every triangle in an ultrametric space is isosceles;

(ii) if $B(x_0, a) \cap B(y_0, a) \neq \emptyset$, then $B(x_0, a) = B(y_0, a)$;

(iii) two spheres $B(x_0, a)$ and $B(y_0, b)$ are either disjoint or one of them is included in the other;

(iv) every $s \in B(x_0, a)$ is a center of the closed sphere $B(x_0, a)$;

(v) if $x \notin B(x_0, r)$, then the distance from x to any point of the sphere $B(x_0, r)$ is the same.

It is interesting to note that for an ultrametric the equivalence classes of $\theta_{d,r}$ coincide with the spheres of the form $B(x, r)$ because $(x, y) \in \theta_{d,r}$ is equivalent to $y \in B(x, r)$. Thus, the ultrametric space (S, d) is partitioned by the set of spheres of radius r. This yields the quotient space $S/\theta_{d,r}$ whose elements are the spheres of radius r of (S, d).

Theorem 3.1. *Let d be an ultrametric (a quasi-ultrametric) on a finite set S. If $|S| = n$, then d takes at most $n - 1$ positive values.*

Proof. The proof is by induction on $n \geqslant 2$. The base case, $n = 2$ is immediate. Let $n \geqslant 3$ and suppose that the statement holds for $m \leqslant n$.

Suppose that $S = \{x_1, x_2, \ldots, x_n\}$. Without loss of generality we may assume that $d(x_1, x_2) = \min\{d(x, y) \mid x, y \in S \text{and } d(x, y) > 0\}$. Then, $d(x_k, x_1) = d(x_k, x_2)$ for all k such that $3 \leqslant k \leqslant n$.

Let $\{a_1, \ldots, a_r\} = \{d(x_k, x_1) \mid 3 \leqslant k \leqslant n\}$, where $0 < a_0 < a_1 < \cdots < a_r$. Define

$$B_j = \{x_k \mid k \geqslant 3 \text{ and } d(x_k, x_1) = a_j\}$$

for $1 \leqslant j \leqslant r$. The collection $B_1, \ldots, B_r$ is a partition of the set $\{x_3, x_4, \ldots, x_n\}$. Let $m_j = |B_j|$; then $m_1 + \cdots + m_r = n - 2$.

If $u \in B_i$ and $v \in B_j$ with $i < j$, then $d(u, x_1) = a_i < a_j = d(v, x_1)$, hence $d(u, v) = a_j$. Therefore, the values of $d(u, v)$ for u, v in distinct blocks of the partition belong to the set $\{a_1, \ldots, a_r\}$. By the induction hypothesis, for each k, $1 \leqslant k \leqslant r$, the restriction of d to B_k can take at most $m_k - 1$ distinct positive values; therefore, d can take at most $1 + r + (m_1 - 1) + \cdots + (m_r - 1) = n - 1$ distinct positive values on S. □

Table 1: Ultrametric defined by dendrogram

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	0	2	2	3	3	3	3	3
x_2	2	0	1	3	3	3	3	3
x_3	2	1	0	3	3	3	3	3
x_4	3	3	3	0	1	1	2	2
x_5	3	3	3	1	0	1	2	2
x_6	3	3	3	1	1	0	2	2
x_7	3	3	3	2	2	2	0	1
x_8	3	3	3	2	2	2	1	0

Let (S, d) be a finite ultrametric space. Since the range of values of d, $\text{Ran}(d)$ is a finite set that contains at most $|S| - 1$ values, the set $\mathsf{SPH}(S, d)$ of spheres of this space $\mathsf{SPH}(S, d) = \{B(x, r) \mid x \in S \text{ and } r \in \text{Ran}(d)\}$ is finite. Consider the graph $\mathcal{T}_{S,d} = (\mathsf{SPH}(S), V)$ having $\mathsf{SPH}(S, d)$ as its set of vertices, where an edge $(B(x, r), B(y, s))$ exists if $B(x, r) \subset B(y, s)$.

$\mathcal{T}_{S,d}$ is a rooted tree. Indeed, if $d_m = \max\{d(x, y) \mid x, y \in S\}$, then the root of the tree is the sphere $B(x, d_m)$, where x is an arbitrary element of S. Since each sphere $B(z, r)$ is included in $B(x, d_m)$, it follows that the graph $\mathcal{T}_{S,d}$ is connected. Furthermore, $\mathcal{T}_{S,d}$ is acyclic. Indeed, if we would have a cycle

$$\begin{aligned} & B(x_1, r_1) \subset B(x_2, r_2) \subset \cdots \\ & \subset B(x_{m-1}, r_{m-1}) \subset B(x_m, r_m) = B(x_1, r_1), \end{aligned}$$

an immediate contradiction would follow because of the strict inclusions that exist between the spheres of this chain. The height of the tree of sphere $\mathcal{T}_{S,d}$ of an ultrametric space (S, d) cannot exceed the number of distinct values of d.

Example 3.2. For the ultrametric space defined by Table 1 the tree of spheres is shown in Figure 1. □

4 Binary and Multivalued Ultrametrics

Let d, e be two dissimilarities on S. The dissimilarity e *dominates* d if $d(x, y) \leqslant e(x, y)$ for every $x, y \in S$. We denote this by $d \sqsubseteq e$.

Theorem 4.1. *The dissimilarity e dominates the dissimilarity d if and only if $\theta_{e,r} \subseteq \theta_{d,r}$ for every $r \in \mathbb{R}_{\geqslant 0}$.*

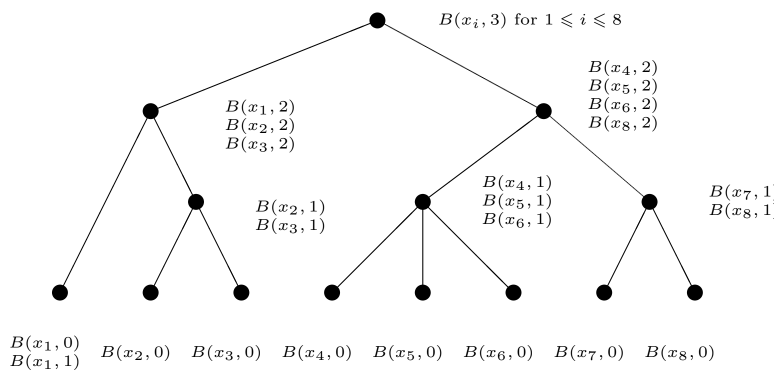

Figure 1: Tree of spheres

Proof. Indeed, suppose that e dominates d. Let $(x, y) \in \theta_{e,r}$, so $e(x, y) \leqslant r$. Since $d(x, y) \leqslant e(x, y) \leqslant r$, it follows that $(x, y) \in \theta_{d,r}$, so $\theta_{e,r} \subseteq \theta_{d,r}$.

Conversely, suppose that $\theta_{e,r} \subseteq \theta_{d,r}$ for every $r \in \mathbb{R}_{\geqslant 0}$. Since $(x, y) \in \theta_{e,e(x,y)} \subseteq \theta_{d,e(x,y)}$ it follows that $d(x, y) \leqslant e(x, y)$ for every $x, y \in S$, that is, that e dominates d. □

Lemma 4.2. *If $B(x_0, a), B(y_0, a)$ are two disjoint spheres in an ultrametric space (S, d), then for $x \in B(x_0, a)$ and $y \in B(y_0, a)$ we have $d(x, y) = d(x_0, y_0)$.*

Proof. Note that $d(x_0, y_0) > a$ because the spheres $B(x_0, a)$ and $B(y_0, a)$ are disjoint. Since $d(x, x_0) \leqslant a$ and the triangle xx_0y_0 is isosceles, we must have $d(x, y_0) = d(x_0, y_0)$. On the other hand, the triangle xy_0y is also isosceles and $d(x_0, y) < a$ it follows that $d(x, y) = d(x, y_0)$, so $d(x, y) = d(x_0, y_0)$. □

Theorem 4.3. *Let (S, d) be an ultrametric space, where* $\text{Ran}(d) = \{0, r_1, \ldots, r_m\}$, *$0 < r_1 < \cdots < r_m$, and $m \leqslant |S| - 1$. The quotient space $S/\theta_{d,r_k}$ contains no more than $n - k$ spheres.*

Proof. The positive distances between elements of S located inside a sphere $B(x, r_k)$ range in the set $\{r_1, \ldots, r_k\}$. If there are q_k spheres of the form $B(x, r_k)$ there exist at most $q_k - 1$ distinct values of d between the centers of these spheres. By Lemma 4.2 there are no more than $k + q_k - 1$ values of the distance d between the points of S and, therefore, $k + q_k - 1 \leqslant n - 1$, which implies $q_k \leqslant n - k$. □

We saw that starting from the characteristic function of an equivalence relation we can build an ultrametric whose range is the set $\{0, 1\}$. The next statement gives a method of constructing ultrametrics starting from chains of equivalence relations.

Theorem 4.4. *Let S be a finite set and let $d : S \times S \longrightarrow \mathbb{R}_{\geqslant 0}$ be a function whose range is* $\mathrm{Ran}(d) = \{r_1, \ldots, r_m\}$*, where $r_1 = 0$ such that $d(x, y) = 0$ if and only if $x = y$.*

The function d is an ultrametric on S if and only if the sequence of relations $\theta_{d,r_1}, \ldots, \theta_{d,r_m}$ is an increasing chain of equivalences on S such that $\theta_{d,r_1} = \alpha_S$ and $\theta_{d,r_m} = \omega_S$.

Proof. Suppose that d is an ultrametric on S. We have $(x, x) \in \theta_{d,r_i}$ because $d(x, x) = 0$, so all relations θ_{d,r_i} are reflexive. Also, it is clear that the symmetry of d implies $(x, y) \in \theta_{d,r_i}$ if and only if $(y, x) \in \theta_{d,r_i}$, so these relations are symmetric.

The ultrametric inequality is essential for proving the transitivity of the relations θ_{d,r_i}. If $(x, y), (y, z) \in \theta_{d,r_i}$, then $d(x, y) \leqslant r_i$ and $d(y, z) \leqslant r_i$, which implies $d(x, z) \leqslant \max\{d(x, y), d(y, z)\} \leqslant r_i$. Thus, $(x, z) \in \theta_{d,r_i}$, which shows that every relation θ_{d,r_i} is transitive and therefore an equivalence.

It is straightforward to see that $\theta_{d,r_1} \leqslant \theta_{d,r_2} \leqslant \cdots \leqslant \theta_{d,r_m}$; that is, this sequence of relations is indeed a chain of equivalences.

Conversely, suppose that $\theta_{d,r_1}, \ldots, \theta_{d,r_m}$ is an increasing sequence of equivalences on S such that $\theta_{d,r_1} = \alpha_S$ and $\theta_{d,r_m} = \omega_S$, where $\theta_{d,r_i} = \{(x, y) \in S \times S \mid d(x, y) \leqslant r_i\}$ for $1 \leqslant i \leqslant m$ and $r_1 = 0$.

Note that $d(x, y) = 0$ is equivalent to $(x, y) \in \theta_{d,r_1} = \alpha_S$, that is, to $x = y$.

We claim that

$$d(x, y) = \min\{r \mid (x, y) \in \theta_{d,r}\}. \tag{3}$$

Indeed, since $\theta_{d,r_m} = \omega_S$, it is clear that there is an equivalence θ_{d,r_i} such that $(x, y) \in \theta_{d,r_i}$. If $(x, y) \in \theta_{d,r_i}$, the definition of θ_{d,r_i} implies $d(x, y) \leqslant r_i$, so $d(x, y) \leqslant \min\{r \mid (x, y) \in \theta_{d,r}\}$. This inequality can be easily seen to become an equality since $(x, y) \in \theta_{d,d(x,y)}$. This implies immediately that d is symmetric.

To prove that d satisfies the ultrametric inequality, let x, y, z be three members of the set S. Let $p = \max\{d(x, z), d(z, y)\}$. Since $(x, z) \in \theta_{d,d(x,z)} \subseteq \theta_{d,p}$ and $(z, y) \in \theta_{d,d(z,y)} \subseteq \theta_{d,p}$, it follows that $(x, y) \in \theta_{d,p}$, due to the transitivity of $\theta_{d,p}$. Thus, $d(x, y) \leqslant p = \max\{d(x, z), d(z, y)\}$, which proves the triangular inequality for d. □

Theorem 4.5. *Let S be a finite set and let $d : S \times S \longrightarrow \mathbb{R}_{\geqslant 0}$ be a a quasi-ultrametric whose range is $\mathrm{Ran}(d) = \{r_1, \ldots, r_m\}$, where $r_1 = 0$. Then d is a linear combination with positive coefficients of binary quasi-ultrametrics,*

$$d(x, y) = a_1 d_1(x, y) + a_2 d_2(x, y) + \cdots + a_{m-1} d_{m-1}(x, y)$$

for $x, y \in S$.

Proof. Note that the range of d consists of $m - 1$ positive values, so $m \leqslant |S|$ by Theorem 3.1. Assume that $0 = r_1 < r_2 < \cdots < r_m$.

Consider the equivalences θ_{d,r_i} for $1 \leqslant i \leqslant m$ and the corresponding binary quasi-ultrametrics d_i given by

$$d_i(x, y) = \begin{cases} 0 & \text{if } d(x, y) \leqslant r_i, \\ 1 & \text{otherwise} \end{cases}$$

for $(x, y) \in S \times S$. Note that $d_m(x, y) = 0$ for every $x, y \in S$.

We claim that there exist $m - 1$ numbers $a_1, \ldots, a_{m-1}$ such that

$$\begin{aligned} d(x, y) &= a_1 d_1(x, y) + a_2 d_2(x, y) + \cdots \\ &\quad + a_{m-1} d_{m-1}(x, y) \end{aligned}$$

for $x, y \in S$.

Indeed, note that $d(x, y) \in \{r_1, \ldots, r_m\}$ by the definition of the range of d. Suppose that we have

$$r_1 < \cdots < r_{k-1} < d(x, y) = r_k < r_{k+1} < \cdots < r_m.$$

This implies

$$d_i(x, y) = \begin{cases} 0 & \text{if } k \leqslant i \leqslant m \\ 1 & \text{otherwise,} \end{cases}$$

so $r_k = a_1 + \cdots + a_{k-1}$ for $1 \leqslant k \leqslant m$. This implies $a_1 = r_2, a_2 = r_3 - r_2, \ldots, a_{m-1} = r_m - r_{m-1}$ and we obtain the necessary equality. □

5 Ultrametricity of Dissimilarities

Let $F_p : \mathbb{R}^2_{\geqslant 0} \longrightarrow \mathbb{R}_{\geqslant 0}$ be the function defined by $F_p(a, b) = (a^p + b^p)^{\frac{1}{p}}$, where $p > 0$.

An *p-dissimilarity* is a dissimilarity $d : S^2 \longrightarrow S$ such that

$$d(x, y) \leqslant F_p(d(x, z), d(z, y)) \tag{4}$$

for $x, y, z \in S$. We denote by $\mathcal{D}_p(S)$ the collection of F_p-dissimilarities on S. The *ultrametricity* of a dissimilarity space (S, d) is the number $\mathsf{u}(S, d) = \sup\{p \in \mathbb{R}_{\geqslant 0} \mid d \in \mathcal{D}_p\}$.

We note that every dissimilarity on a set S belongs to $\mathcal{D}_0(S)$. Also, every dissimilarity in $\mathcal{D}_1(S)$ satisfies the triangular axiom $d(x, y) \leqslant d(x, z) + d(z, y)$ for x, y, z, so it is a metric on S.

Lemma 5.1. *Let p, q be two positive numbers. If $p \geqslant q$ then we have $F_p(a, b) \leqslant F_q(a, b)$.*

Proof. Consider the function $\phi : \mathbb{R}_{\geqslant 0} \longrightarrow \mathbb{R}_{\geqslant 0}$ given by

$$\phi(p) = (a^p + b^p)^{\frac{1}{p}}.$$

We have

$$\frac{\phi'(p)}{\phi(p)} = \frac{a^p \ln a + b^p \ln b}{p(a^p + b^p)} - \frac{\ln(a^p + b^p)}{p^2}.$$

Since $\frac{a^p}{a^p+b^p} \ln \frac{a^p}{a^p+b^p} + \frac{b^p}{a^p+b^p} \ln \frac{b^p}{a^p+b^p} < 0$ it follows that $\phi'(p) < 0$, which shows that ϕ is a decreasing function and this implies the statement of the lemma. □

Note also that for any $p > 0$ the function F_p is monotonic in each of its arguments.

Lemma 5.1 implies that if $p \geqslant q$, then $\mathcal{D}_p(S) \subseteq \mathcal{D}_q(S)$.

Theorem 5.2. *Let d be a dissimilarity on a set S. We have $d \in \bigcap_{r \geqslant 0} \mathcal{D}_r$ if and only if d is an ultrametric on S.*

Proof. If $d \in \bigcap_{r \geqslant 0} \mathcal{D}_r$ we have $d(x, y) \leqslant (d(x, z)^r + d(z, y)^r)^{\frac{1}{r}}$ for every $r \geqslant 0$ and $x, y, z \in S$. Therefore, $d(x, y) \leqslant \lim_{r \to \infty}(d(x, z)^r + d(z, y)^r)^{\frac{1}{r}} = \max\{d(x, z), d(z, y\}$, which implies that d is an ultrametric.

Conversely, if d is an ultrametric, we have $d(x, y) \leqslant \max\{d(x, z), d(z, y)\} \leqslant (d(x, z)^r + d(z, y)^r)^{\frac{1}{r}}$ for every $r > 0$, so $d \in \mathcal{D}_r$ for every $r > 0$. Since $d \in \mathcal{D}_0$, it follows that $d \in \bigcap_{r \geqslant 0} \mathcal{D}_r$. □

Theorem 5.3. *Let r, s be two positive numbers such that $s < r$, let $d \in \mathcal{D}_s$. The family of r-dissimilarities dominated by d has a largest element.*

Proof. Let $\mathcal{D} = \{d_i \mid i \in I\}$ be a collection of dissimilarities on a set S such that $\mathcal{D} \subseteq \mathcal{D}_r$. Then, the dissimilarity d defined by $d(x,y) = \max_{i\in I} d_i(x,y)$ for $x, y \in S$ belongs to $\mathcal{D}_r$.

It is immediate that d itself is a dissimilarity on S and we have

$$\begin{aligned} d(x,y) &\leqslant d_i(x,y) \leqslant F_r(d_i(x,z), d_i(z,y)) \\ &\quad \text{(for every } i \in I \text{ because } d_i \in \mathcal{D}_r) \\ &\leqslant F_r(d(x,z), d(z,y)) \\ &\quad \text{(because } F_r \text{ is monotonic in both} \\ &\quad \text{its arguments),} \end{aligned}$$

so $d \in \mathcal{D}_r$.

Let e be a dissimilarity in $\mathcal{D}_s$ and let $\mathcal{D}$ be the set of dissimilarities in $\mathcal{D}_r$ dominated by e. If $d'(x,y) = \max\{d \in \mathcal{D}, d \leqslant e\}$, then $d' \leqslant e$ and $d \leqslant d'$ for every $d \in \mathcal{D}$, so d' is the largest dissimilarity in $\mathcal{D}$. □

Theorem 5.3 implies that given a dissimilarity d on a set S there exists a largest metric (in $\mathcal{D}_1$) that is dominated by d. Another, well-known result that follows from this theorem is the fact that given a dissimilarity, there exists the largest ultrametric that is dominated by this dissimilarity [8, 13]. This ultrametric is obtained by clustering the metric space using the single-link hierarchical clustering.

Let r, s be two positive numbers. An (r,s)-*transformation* is a function $g : \mathbb{R}_{\geqslant 0} \longrightarrow \mathbb{R}_{\geqslant 0}$ such that

(i) $g(x) = 0$ if and only if $x = 0$;

(ii) g is a strictly monotonic function on $\mathbb{R}_{\geqslant 0}$, and

(iii) $g(F_r(a,b)) \leqslant F_s(g(a), g(b))$ for $a, b \in \mathbb{R}_{\geqslant 0}$.

Theorem 5.4. *If d is a dissimilarity on S such that $d \in \mathcal{D}_r$ and g is an (r,s)-transformation, then gd is a dissimilarity in $\mathcal{D}_s$.*

Proof. Let d be an r-dissimilarity. It is immediate that gd is a dissimilarity. Since g is an (r,s)-transformation, we have

$$\begin{aligned} F_s(g(d(x,z)), g(d(z,y))) &\geqslant g(F_r(d(x,z), d(z,y)) \\ &\geqslant g(d(x,y)), \end{aligned}$$

so gd is an s-dissimilarity. □

Example 5.5. The function g given by $g(x) = \ln(x+1)$ for $x \geqslant 0$ is a $(2,1)$-transformation. Indeed, for $a, b \geqslant 0$ we have the immediate inequality:

$$\begin{aligned} g(F_2(a,b)) &= \ln(\sqrt{a^2+b^2}+1) \\ &\leqslant \ln(a+1) + \ln(b+1). \end{aligned}$$

Note that if g is an (r,s)-transformation, then g^{-1} is an (s,r)-transformation. Therefore, the function $h(x) = e^x - 1$ is a $(1,2)$-transformation. ∎

Example 5.6. Let $\alpha > 0$ and let $g : \mathbb{R}_{\geqslant 0} \longrightarrow \mathbb{R}_{\geqslant 0}$ be the monotonic function defined by $g(x) = x^\alpha$ for $x \geqslant 0$. If $s \leqslant \frac{r}{\alpha}$, then g is an (r,s)-transformation. Indeed, we have

$$\begin{aligned} F_s(g(a), g(b)) &= (a^{\alpha s} + b^{\alpha s})^{\frac{1}{s}} \\ &= (F_{\alpha s}(a,b))^\alpha \\ &\geqslant (F_r(a,b))^\alpha \\ & \text{(by Lemma 5.1)} \\ &= g(F_r(a,b)), \end{aligned}$$

which shows that g is an (r,s)-transformation. For this transformation we have $\mathsf{u}(S, gd) = \frac{\mathsf{u}(S,d)}{\alpha}$. Thus, by choosing α we can modulate the ultrametricity of the transformed dissimilarity space. ∎

The notion of ultrametricity introduced above involves satisfying Inequality (4) for all triangles xyz of the dissimilarity space. Therefore, a few triangles in the dissimilarity space which have very different side lengths can unduly influence the value of the ultrametricity. This motivates considering yet another variant of ultrametricity of dissimilarities.

Let $t = xyz \in S^3$ be a triangle in the is a dissimilarity space (S,d). Following Lerman's notation in [9], if $d(x,y) \geqslant d(x,z) \geqslant d(y,z)$, we write $S_d(t) = d(x,y)$ for the longest side of t, $M_d(t) = d(x,z)$ for the middle side of t, and $L_d(t) = d(y,z)$ for the shortest side of the triangle. We consider two local variants of ultrametricity of a triangle.

The *strong ultrametricity* of t is the number $u_d(t) = \max\{r > 0 \mid S_d(t) \leqslant F_r(M_d(t), L_d(t))\}$. The *weak ultrametricity* of t is the number $w_d(t)$ given by

$$w_d(t) = \begin{cases} \frac{1}{\log_2 \frac{S_d(t)}{M_d(t)}} & \text{if } S_d(t) > M_d(t) \\ \infty & \text{if } S_d(t) = M_d(t). \end{cases}$$

If $w_d(t) = \infty$, then t is an *ultrametric triple.*

The *weak ultrametricity* of the dissimilarity space (S, d) is the number $w(S, d)$ defined by

$$w(S, d) = \mathsf{median}\{w_d(t) \mid t \in S^3\}.$$

For a triangle t we have

$$0 \leqslant S_d(t) - M_d(t) = \left(2^{\frac{1}{w_d(t)}} - 1\right) M_d(t) \leqslant \left(2^{\frac{1}{w(S,d)}} - 1\right) M_d(t)$$

Thus, if $w_d(t)$ is sufficiently large, the triangle t is almost isosceles. For example, if $w_d(t) = 5$, the difference between the length of longest side $S_d(t)$ and the median side $M_d(t)$ is less than 15%.

For every triangle $t \in S^3$ in a dissimilarity space we have $u_d(t) \leqslant w_d(t)$. Indeed, since $S_d(t)^{u_d(t)} \leqslant M_d(t)^{u_d(t)} + L_d(t)^{u_d(t)}$ we have $S_d(t)^{u_d(t)} \leqslant 2M_d(t)^{u_d(t)}$, which implies immediately $u_d(t) \leqslant w_d(t)$.

Theorem 5.7. *Let (S, d) be a dissimilarity space and let $f : \mathbb{R}_{\geqslant 0} \longrightarrow \mathbb{R}_{\geqslant 0}$ be a strictly increasing function on $\mathbb{R}_{\geqslant 0}$.*

If the function $g : \mathbb{R}_{\geqslant 0} \longrightarrow \mathbb{R}_{\geqslant 0}$ given by

$$g(a) = \begin{cases} \frac{f(a)}{a} & \text{if } a > 0, \\ 0 & \text{if } a = 0 \end{cases}$$

is strictly decreasing, then the function $e : S \times S \longrightarrow \mathbb{R}_{\geqslant 0}$ defined by $e(x, y) = f(d(x, y))$ for $x, y \in S$ is a dissimilarity and $w_d(t) \leqslant w_e(t)$ for every triangle $t \in S^3$.

Proof. It is immediate that $e(x, y) = e(y, x)$ and $e(x, x) = 0$ for $x, y \in S$. Let $t = xyz \in S^3$ be a triangle. Since $S_d(t) > M_d(t)$ and g is strictly decreasing, $g(S_d(t)) \leqslant g(M_d(t))$, which implies $\frac{f(S_d(t))}{S_d(t)} \leqslant \frac{f(M_d(t))}{M_d(t)}$. Since f is a strictly increasing function we have $S_e(t) = f(S_d(t))$ and $M_e(t) = f(M_d(t))$. This allows us to write:

$$\frac{S_e(t)}{M_e(t)} = \frac{f(S_d(t))}{f(M_d(t))} \leqslant \frac{S_d(t)}{M_d(t)}.$$

Therefore,

$$w_d(t) = \frac{1}{\log_2 \frac{S_d(t)}{M_d(t)}} \leqslant \frac{1}{\log_2 \frac{S_e(t)}{M_e(t)}} = w_e(t).$$

□

Example 5.8. Let (S, d) be a dissimilarity space and let e be the dissimilarity defined by $e(x, y) = d(x, y)^r$, where $0 < r < 1$. If $f(a) = a^r$, then f is strictly increasing and the function $g : \mathbb{R}_{\geqslant 0} \longrightarrow \mathbb{R}_{\geqslant 0}$ given by

$$g(a) = \begin{cases} \frac{f(a)}{a} & \text{if } a > 0, \\ 0 & \text{if } a = 0 \end{cases} = \begin{cases} a^{r-1} & \text{if } a > 0, \\ 0 & \text{if } a = 0 \end{cases}$$

is strictly decreasing. Therefore, the weak ultrametricity $w_e(t)$ is greater than $w_d(t)$, where $e(x, y) = (d(x, y))^r$ for $x, y \in S$. □

Example 5.9. Let $f : \mathbb{R}_{\geqslant 0} \longrightarrow \mathbb{R}_{\geqslant 0}$ be defined by $f(a) = \frac{a}{a+1}$. It is easy to see that f is strictly increasing on $\mathbb{R}_{\geqslant 0}$ and

$$g(a) = \begin{cases} \frac{1}{1+a} & \text{if } a > 0, \\ 0 & \text{if } a = 0 \end{cases}$$

is strictly decreasing on the same set. Therefore, the weak ultrametricity of a triangle increases when d is replaced by e given by

$$e(x, y) = \frac{d(x, y)}{1 + d(x, y)}$$

for $x, y \in S$. □

Example 5.10. For a dissimilarity space (S, d), the Schoenberg transform of d described in [3] is the dissimilarity $e : S^2 \longrightarrow \mathbb{R}_{\geqslant 0}$ defined by

$$e(x, y) = 1 - e^{-kd(x,y)}$$

for $x, y \in S$. Let $f : \mathbb{R}_{\geqslant 0} \longrightarrow \mathbb{R}_{\geqslant}$ be the function $f(a) = 1 - e^{-ka}$ that is used in this transformation. It is immediate that f is a strictly increasing function. For $a > 0$ we have $g(a) = \frac{1-e^{-ka}}{a}$, which allows us to write

$$g'(a) = \frac{e^{-ka}(ka + 1) - 1}{a^2}$$

for $a > 0$. Taking into account the obvious inequality $ka + 1 < e^{ka}$ for $k > 0$, it follows that the function g is strictly decreasing. Thus, the weak ultrametricity of a triangle relative to the Schoenberg transform is greater than the weak ultrametricity under the original dissimilarity. □

6 Conclusions and Future Work

Theorem 5.3 highlights the common nature of the results used in two rather areas of data mining: obtaining the subdominant ultrametric for a dissimilarity through the single-link hierarchical clustering, and obtaining a configuration of points in a metric space whose distances approximate object dissimilarities. The later process is known as non-metric multidimensional scaling. We propose to explore computing the largest metric that approximates a dissimilarity between objects without the intermediate calculation of a representation of the objects in $\mathbb{R}^n$ equipped with a Minkowski metric.

We introduced the notion of ultrametricity of dissimilarities. Transformations that increase or diminish ultrametricity, decrease or accentuate discrepancies between dissimilarity values, respectively. We will examine the impact of these transformations on various clustering algorithms and classification methods (such as the k^{th} nearest neighbor). Another possible application of these transformations lies in the area of outlier detection.

References

[1] Y. Amice. *Les nombres p-adiques.* Presses Universitaires de France, Paris, 1975.

[2] P. Contreras and F. Murtagh. Fast, linear time hierarchical clustering using the Baire metric. *Journal of Classification*, 29(2):118–143, 2012. doi:10.1007/s00357-012-9106-3.

[3] M. M. Deza and M. Laurent. *Geometry of Cuts and Metrics.* Springer, Heidelberg, 1997. doi:10.1007/978-3-642-04295-9.

[4] A. D. Gordon. *Classification.* Chapman and Hall, London, 1981.

[5] A. D. Gordon. A review of hierarchical classification. *Journal of the Royal Statistical Society, Series (A)*, 150(2):119–137, 1987. doi:10.2307/2981629.

[6] N. Jardine and R. Sibson. *Mathematical Taxonomy.* Wiley, New York, 1971.

[7] M. Kimura. *The Neutral Theory of Molecular Evolution.* Cambridge University Press, Cambridge, UK, 1983.

[8] B. Leclerc. La comparaison des hiérarchies: indices et métriques. *Mathématiques et sciences humaines*, 92:5–40, 1985.

[9] I. C. Lerman. *Classification et Analyse Ordinale des Données*. Dunod, Paris, 1981.

[10] J. Ninio. *Molecular Approaches to Evolution*. Princeton University, Princeton, NJ, 1983.

[11] R. Rammal, G. Touluse, and M. A. Virasoro. Ultrametricity for physicists. *Reviews of Modern Physics*, 58:765–788, 1986. doi:10.1103/RevModPhys.58.765.

[12] W. H. Schikhof. *Ultrametric Calculus*. Cambridge University Press, Cambridge, UK, 1984.

[13] D. A. Simovici and C. Djeraba. *Mathematical Tools for Data Mining*. Springer, London, second edition, 2014. doi:10.1007/978-1-4471-6407-4.

[14] D. A. Simovici, R. Vetro, and K. Hua. Ultrametricity of dissimilarity spaces and its significance for data mining. In *Proceedings of EGC 2015, Luxembourg*. EGC, 2015. (to appear).