Ultrametricity of Dissimilarity Spaces and Its Significance for Data Mining

Dan Simovici*, Rosanne Vetro**, Kaixun Hua***

University of Massachusetts Boston
* dsim@cs.umb.edu
** rvetro@cs.umb.edu
*** kingsley@cs.umb.edu

Abstract. We introduce a measure of ultrametricity for dissimilarity spaces and examine transformations of dissimilarities that impact this measure. Then, we study the influence of ultrametricity on the behavior of two classes of data mining algorithms (kNN classification and PAM clustering) applied on dissimilarity spaces. We show that there is an inverse variation between ultrametricity and performance of classifiers. For clustering, increased ultrametricity generate clusterings with better separation. Lowering ultrametricity produce more compact clusters.

1 Introduction

Ultrametrics occur in the study of agglomerative hierarchical clustering algorithms, phylogenetic trees, p-adic numbers, certain physical systems, etc.

Our goal is to evaluate the degree of ultrametricity of dissimilarity spaces and to study the impact of the degree of ultrametricity on performance of classification and clustering algorithms.

Measuring ultrametricity of metric spaces has preoccupied a number of researchers (for example, in (Rammal et al., 1985)); however, the proposed measures are usable for the special case of metrics and are linked to the subdominant ultrametric attached to a metric which requires computing a single-link clustering or a minimal spanning tree. We propose an alternative measure referred to as the weak ultrametricity that can be applied to the more general case of dissimilarity spaces.

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

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

The mapping d is a *quasi-metric* if it is a dissimilarity and it satisfies the triangular inequality $d(x, y) \leqslant d(x, z) + d(z, y)$ for $x, y, z \in S$. In addition, if $d(x, y) = 0$ implies $x = y$, then d is a *metric*.

An *quasi-ultrametric* is a dissimilarity $d : S \times S \longrightarrow \mathbb{R}_{\geqslant 0}$ that satisfies the inequality $d(x, y) \leqslant \max\{d(x, z), d(z, y)\}$ for $x, y, z \in S$. If, in addition, $d(x, y) = 0$ implies $x = y$, then d is an *ultrametric*.

In Section 2 we introduce a measure of ultrametricity for dissimilarity spaces and a weaker variant of this measure that is better from a computational point of view. Then, we examine transformations of dissimilarities that affect ultrametricity. The influence of ultrametricity of dissimilarities on the performance of classifiers is examined in Section 2 using the k-nearest neighbors classifiers. Section 4 is dedicated to the study of the impact of ultrametricity on cluster compactness and separation.

2 Evaluating Ultrametricity in Dissimilarity Spaces

Let r be a non-negative number and let $\mathcal{D}_r(S)$ be the set of dissimilarities defined on S that satisfy the inequality $d(x, y)^r \leqslant d(x, z)^r + d(z, y)^r$ for $x, y, z \in S$. Note that every dissimilarity belongs to the set $\mathcal{D}_0$; a dissimilarity in $\mathcal{D}_1$ is a semimetric.

Let $\mathcal{D}_\infty = \bigcap_{r \geqslant 0} \mathcal{D}_r$. If $d \in \mathcal{D}_\infty$, then d is an ultrametric. Indeed, let $d \in \mathcal{D}_\infty$ and assume that $d(x, y) \geqslant d(x, z) \geqslant d(z, y)$. Then, $d(x, y) \leqslant d(x, z) \left(1 + \left(\frac{d(y,z)}{d(x,z)}\right)^r\right)^{\frac{1}{r}}$ for every $r \geqslant 0$. Since $\lim_{r \to \infty} d(x, z) \left(1 + \left(\frac{d(y,z)}{d(x,z)}\right)^r\right)^{\frac{1}{r}} = d(x, z)$, it follows that $d(x, y) \leqslant d(x, z) = \max\{d(x, z), d(z, y)$ for $x, y, z \in S$, which allows us to conclude that d is an ultrametric.

It is easy to verify that $r \leqslant s$ implies $(d(x, z)^r + d(z, y)^r)^{\frac{1}{r}} \geqslant (d(x, z)^s + d(z, y)^s)^{\frac{1}{s}}$ (see (Simovici and Djeraba, 2014), Lemma 6.15). Thus, if $r \leqslant s$ we have the inequality $\mathcal{D}_s \subset \mathcal{D}_r$.

Let (S, d) be a dissimilarity space and let $t = xyz$ be a triangle. Following Lerman's notation (Lerman, 1981), we write $S_d(t) = d(x, y), M_d(t) = d(x, z),$ and $L_d(t) = d(y, z)$, if $d(x, y) \geqslant d(x, z) \geqslant d(y, z)$.

Definition 2.1. Let (S, d) be a dissimilarity space and let $t = xyz \in S^3$ be a triangle.

The *ultrametricity* of t is the number $u_d(t)$ defined by

$$u_d(t) = \max\{r > 0 \mid S_d(t)^r \leqslant M_d(t)^r + L_d(t)^r\}.$$

If $d \in \mathcal{D}_p$, we have $p \leqslant u_d(t)$ for every $t \in S^3$.

The notion of weak ultrametricity that we are about to introduce has some computational advantages over the notion of ultrametricity, especially from the point of view of handling transformations of metrics.

The *weak ultrametricity* of the triangle t, $w_d(t)$, is 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

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

□

The definition of $w(S,d)$ eliminates the influence of triangles whose ultrametricity is an outlier, and gives a better picture of the global ultrametric property of (S,d).

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 is equivalent to $u_d(t) \leqslant w_d(t)$.

Next we discuss dissimilarity transformations that impact the ultrametricity of dissimilarities.

Theorem 2.2. *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 2.3. 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 2.4. 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 2.5. The Schoenberg transform of a dissimilarity d described in (Deza and Laurent, 1997) 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. □

3 Classification and Ultrametricity

The k-nearest neighbors algorithm (kNN) is a classification method that is memory-based and does not require a model to fit. The classification is decided according to a simple majority decision among the most similar training set samples.

We show that the performance of kNN applied to a dissimilarity space (S, d) degrades with the increase of the ultrametricity of d. This happens because the increase of ultrametricity among the elements of S promotes the equalization of distances.

We begin with a dissimilarity space (S, d) and we obtain a new dissimilarity $d' = f(d)$, where f is one of the transformations examined in Section 2. Algorithm 1 encapsulates the above process. It runs kNNwith t-fold cross-validation and computes the confusion matrix generated for each fold as well as the cumulative classification error of the transformed space.

We limit the precision of the transformed dissimilarity d' taking into account, as observed in (Murtagh et al., 2008) that ultrametricity can decrease with the increase in precision. Limiting the precision of d' to a few decimal digits promotes the equalization of those distances. We used in our experiments the data sets *Fisheriris* and *ionosphere* available from

Algorithm 1: Runs kNN with transformed distance function

Input: A metric or dissimilarity space $S = (M, d)$, the number of nearest neighbors k, the number of folds t and a function f, such that $f(d) = d'$ and $u <= u'$ where u and u' are the ultrametricities of S and $S' = (M, d')$, respectively.
Output: The cumulative classification error of the transformed space S'
$d^{'} \leftarrow f(d)$, limited to some decimal precision
$partition\ M\ in\ t\ subsamples$
for *i=1* **to**t **do**
    $training = partition(i).training$
    $test = partition(i).test$
    $testSize(i) = size(test)$
    kNN$(training, test, k, d')$
    $err(i) = \#\ misclassified\ objects$
return $cerr = sum(err)/sum(testsSize)$

Diss.	Iris			Ionosphere			Ovarian cancer		
	$k=3$	$k=5$	$k=7$	$k=3$	$k=5$	$k=7$	$k=3$	$k=5$	$k=7$
d	0.1033	0.0467	0.0427	0.3860	0.3701	0.3852	0.1403	0.1394	0.1431
$d^{0.1}$	0.1187	0.0753	0.0567	0.3875	0.4097	0.3897	0.1454	0.1431	0.1477
$d^{0.01}$	0.2700	0.2900	0.3000	0.5211	0.5239	0.5365	0.3574	0.3181	0.3000

TAB. 1: Average of 10 computations of the classification error produced by kNN using stratified t-fold cross-validation, for different values of k and t = 10.

`https://archive.ics.uci.edu/ml/data sets/` and data set *ovarian cancer* obtained from the FDA-NCI Clinical Proteomics Program Databank (`http://home.ccr.cancer.gov/ncifdaproteomics/ppatterns.asp`).

Our experiments considered a initial Euclidean space (S, d) where S corresponds to one of the data sets described above and d to the Euclidean distance. We first tested our method on the original space and compared the results to the results generated by the increase of ultrametricity of dissimilarity $d' = f(d)$, where $f(a) = a^r$ for $a \geqslant 0$. We used kNN with both t-fold cross-validation and with stratified t-fold cross-validation (where each fold has roughly equal size and roughly the same class proportions as in the entire data set). The transformed distances were limited to 2 decimal digit precision.

The classification error obtained is consistently higher for the case of the transformed space (S, d'), in both validation scenarios. In Table 2 we show the results for three values of k (the number of neighbors) in stratified 10-fold validation. Similar results are obtained for 5 folds in both validation scenarios.

4 The Impact of Ultrametricity on Cluster Compactness and Separation

Clustering validation evaluates and assesses the goodness of the results of a clustering algorithm (Maulik and Bandyopadhyay, 2002). We used internal validation measures that rely on information in the data (Tang et al., 2005), namely and compactness and separation (Tang et al., 2005; Zhao and Karypis, 2002).

Compactness measures quantify how well-related the objects in a cluster are. It provides information about the cohesion of objects in an individual cluster with respect to the other objects outside the cluster. A group of measures evaluate cluster compactness based on variance where lower values indicate better compactness. Other measures are based on distance, such as maximum or average pairwise distance, and maximum or average center-based distance.

Separation is a measure of distinctiveness between a cluster and the rest of the world. The pairwise distances between cluster centers or the pairwise minimum distances between objects in different clusters are often used as measures of separation.

The compactness of each cluster was evaluated using the average dissimilarity between the observations in the cluster and the medoid of the cluster. Separation was computed using the minimal dissimilarity between an observation of the cluster and an observation of another cluster.

We investigate the impact of ultrametricity on compactness and separation of clusters by using the Partition Around Medoids (PAM) algorithm (Kaufman and Rousseeuw, 1990) to cluster objects originally in the Euclidean Space and later in a transformed dissimilarity space with lower or higher ultrametricity.

Experiments show that a transformation on the distance matrix that decreases the ultrametricity of the original Euclidean space can actually improve compactness but also decrease separation of the clusters generated by PAM. However, the compactness improves at a faster ratio than the decrease in separation. We also observed that the increase of ultrametricity produces the reverse effect, degrading compactness and increasing separation, at different ratios. In this case, compactness decreases in a faster ratio than the increase in separation.

Let (S, d) be a dissimilarity space, (S, d') be the transformed dissimilarity space, where $d' = f(d)$ is obtained by applying one of the transformations described in Section 2 and let u and u' be the weak ultrametricities of these two dissimilarity spaces, respectively.

The increase of ultrametricity from (S, d) to (S, d') promotes the equalization of dissimilarity values. In the extreme case, we have an ultrametric space where the pairwise distances involved in all triplets of points form an equilateral or isosceles triangle. To explore how the equalization (or the reverse process) may affect clustering quality, a better study of the effects of increased (or decreased) ultrametricity on the results generated by a widely known and robust clustering algorithm was performed.

In order to study the impact of ultrametricity on cluster compactness and separation, we have implemented an algorithm that runs PAM on the original and transformed spaces, and computes those measure for each cluster from S and S'.

Our experiments considered a initial Euclidean space (S, d) where S corresponds to a set of objects and d to the Minkowski distance with exponent 2. To obtain a valid comparison of compactness and separation, the clusters obtained from a specific data set S must contain the same elements in the original and transformed spaces.

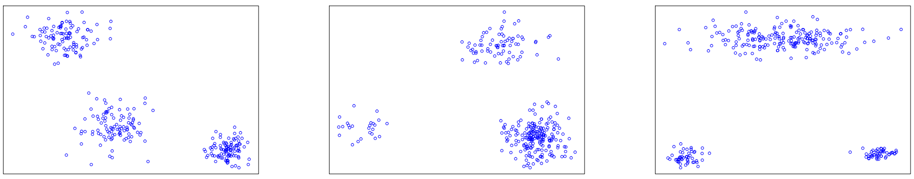
Dissimilarities d^x where $x > 1$ tend to decrease the ultrametricity of the original space, whereas dissimilarities where $0 < x < 1$ tend to increase ultrametricity.

Current existing clustering validation measures and criteria can be affected by various data characteristics (Liu et al., 2010). For instance, data with variable density is challenging for several clustering algorithms. It is known that k-means suffers from an uniformizing effect which tends to divide objects into relatively equal sizes (Xiong et al., 2009). Likewise, k-means and PAM do not have a good performance when dealing with skewed distribution data sets where clusters have unequal sizes. To determine the impact of ultrametricity in the presence of any of those characteristics, experiments were carried considering 3 different data aspects: good separation, density, and skewed distributions in three synthetic data sets named *WellSeparated*, *DifferentDensity* and *SkewDistribution*, respectively.

Figure 1 shows the synthetic data that was generated for each aspect. Each data set contains 300 objects.

Tables 2 shows the results for data sets *WellSeparated*, *DifferentDensity* and *SkewDistribution*, respectively. The measure (compactness or separation) ratio is computed dividing the transformed space measure by the original space measure. The average measure ratio computed for the 3 clusters is presented in each table.

Note that the average measure ratio is less than one for spaces with lower ultrametricity (obtained with dissimilarities d^5 and d^{10}). In this case, the average compactness ratio is also lower than the average separation ratio, showing that the transformations generated intra-cluster dissimilarities that shrunk more than the inter-cluster ones, relatively to the original dissimilarities. In spaces with higher ultrametricity (obtained with dissimilarities $d^{0.1}$ and $d^{0.01}$), the average measure ratio is higher than one. The average compactness ratio is also higher than the average separation ratio, showing that the transformations generated intra-cluster dissimilarities that expanded more than the inter-cluster ones. This explain the equalization effect obtained with the increase in ultrametricity.



(a) *Well Separated* (b) *Different Density* (c) *Skewed Distribution*

FIG. 1: Synthetic data containing 3 different data aspects: 1a: good separation, 1b: different density and 1c: skewed distributions

Figures 2a, 2b and 2c show the relation between compactness a separation ratios for each data set.

In Figure 2 we show the relationship between compactness and separation ratios for the three synthetic data sets and for the *Fisheriris* data set which exhibit similar variation patterns.

As previously mentioned, data with characteristics such as different density and different cluster sizes might impose a challenge for several clustering algorithms.

Diss.	Compactness Avg.	Compactness Ratio Avg.	Separation Avg.	Separation Ratio Avg.
d	0.159852		0.462208	
d^{10}	6.782346E-008	0.00000036	2.113074E-006	0.000004433
d^5	0.000150339	0.00082451	0.002830298	0.006041492
$d^{0.1}$	0.835946	5.493823933	0.955770	2.073671831
$d^{0.01}$	0.973616	6.433067888	0.995943	2.161429967

Results for a data set with well-separated clusters

Diss.	Compactness Avg.	Compactness Ratio Avg.	Separation Avg.	Separation Ratio Avg.
d	0.226611		0.883006	
d^{10}	0.000000299	1.225126E-006	0.0085821266	0.0067414224
d^5	0.000414	0.001758	0.120677	0.101145
$d^{0.1}$	0.862157	3.829475	1.019217	1.247117
$d^{0.01}$	0.968235	4.302930	1.002328	1.234965

Results for a data set with clusters with varied densities

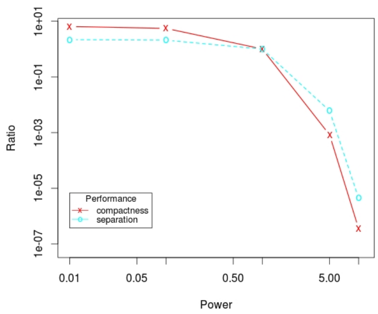
Diss.	Compactness Avg.	Compactness Ratio Avg.	Separation Avg.	Separation Ratio Avg.
d	0.152911		1.088650	
d^{10}	5.001356E-005	0.0001674944	0.0202263733	0.0185757406
d^5	0.001707	0.005744	0.240466	0.220866
$d^{0.1}$	0.815746	7.502117	1.042825	0.957924
$d^{0.01}$	0.966675	9.123531	1.004683	0.922859

Results for a data set with skewed distributions.

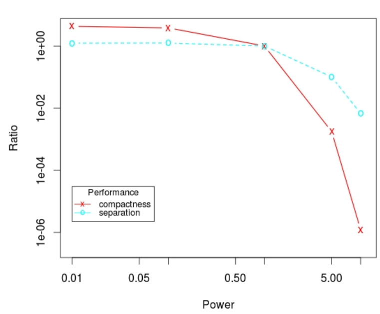
Diss.	Compactness Avg.	Compactness Ratio Avg.	Separation Avg.	Separation Ratio Avg.
d	2.564313e-01		2.841621e-01	
d^{10}	4.495584e-07	1.753134e-06	1.171608e-05	4.123026e-05
d^5	7.628527e-04	2.974881e-03	4.583216e-03	1.612888e-02
$d^{0.1}$	8.664974e-01	3.379062e+00	8.715969e-01	3.067252e+00
$d^{0.01}$	9.630195e-01	3.755467e+00	9.858841e-01	3.469442e+00

Results for the *Fisheriris* data set.

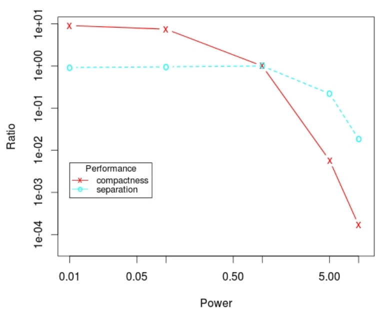
TAB. 2: Cluster compactness and separation using PAM on three synthetic data sets and *Fisheriris*. dissimilarities. Both ratio averages are computed relative to the data set cluster compactness and separation values given by the original dissimilarity d.



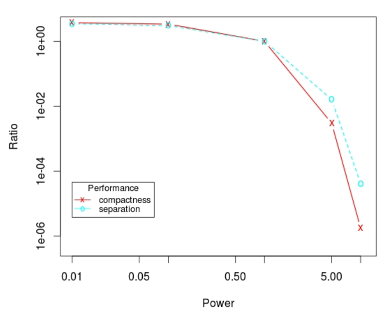
(a) WellSeparated Data Set



(b) DifferentDensity Data Set



(c) SkewDistr Data Set



(d) Fisheriris Data Set

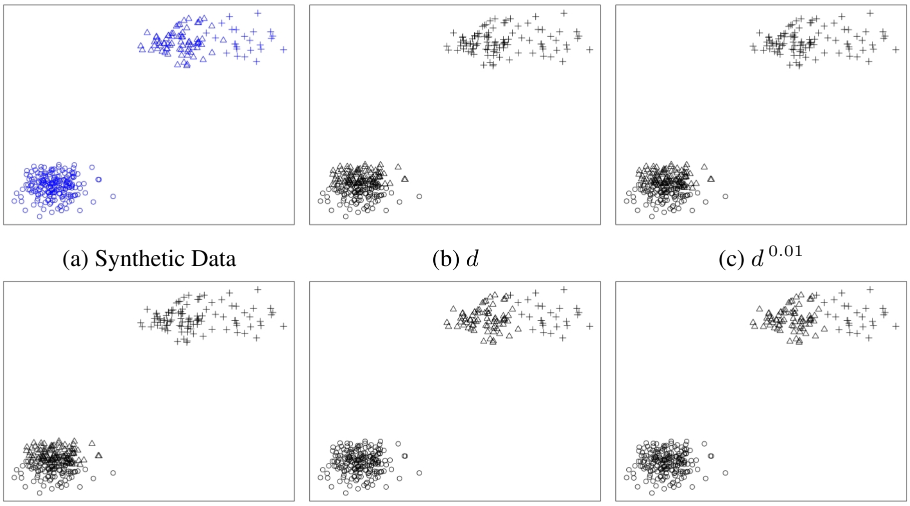
FIG. 2: Relation between Compactness and Separation Ratios for three synthetic data set and for the Fisheriris data set

We show a scenario where PAM, when applied to the original Euclidean space, does not perform well. Nevertheless, we are able to improve the PAM's results by applying a transformation that decreases the ultrametricity of the original space and running PAM on the transformed space.

Consider the data set presented in Figure 3a which was synthetically generated in an Euclidean Space with pairwise metric d by three normal distributions with similar standard deviation but different densities. It has 300 points in total, with the densest group including 200 points and the other two containing 75 and 25 points.

Note that the somewhat sparse groups are also located very close to each other. Different symbols are used to identify the three distinct distributions. PAM's objective function tries to minimize the sum of the dissimilarities of all objects to their nearest medoid. However, it may fail to partition the data into the original distributions when dealing with different density data since the split of the densest cluster may occur. In our example, PAM does exactly that and also combines the two sparse clusters that are not well separated. Notice that unlike k-means (which also does not perform well in these scenarios but eventually can find the right partition due to the randomness on the selection of the centroids), PAM will most likely fails due to the determinism of its BUILD and SWAP steps combined and the choice of the objective function.

To explore the positive effect of increased intra-cluster compactness generated by new spaces with lower ultrametricity on data containing those characteristics, we applied the same transformations with positive integer exponents to the original Euclidean distance matrix obtained from d. Results show significant improvement of the clustering. Figure 3b shows the result of applying PAM to cluster the synthetic data with dissimilarity d. Note that the clustering result does not correspond to a partition resembling the distributions that were used to generate the data. Figures 3d and 3c show that PAM also fails to provide a good partition with dissimilarities $d^{0.1}$ and $d^{0.01}$ since the increase in ultrametricity promotes equalization of dissimilarities which may degrade even more the results. Note however that the partitions obtained by PAM using the dissimilarities d^5 and d^{10} form similar clusters to the ones generated by the original distributions. Indeed, the increase in compactness helps PAM to create boundaries that are compliant with the original normal distributions.



(a) Synthetic Data (b) d (c) $d^{0.01}$

(d) $d^{0.1}$ (e) d^5 (f) d^{10}

FIG. 3: 3a shows the synthetic data generated from distributions with different density. 3b to 3f show the results of PAM using Euclidean distance d and other dissimilarities obtained by transformations on d.

Table 3 shows the measures and ratios for this data set. Figure 4 shows the relationship between compactness and separation ratios.

5 Conclusions and Further Work

We examined the influence of ultrametricity of dissimilarity spaces regarding classification and clustering.

Diss.	Compactness Avg.	Compactness Ratio Avg.	Separation Avg.	Separation Ratio Avg.
d	0.138692		0.460486	
d^{10}	1.295368e-09	9.339889e-09	0.011426	0.024814
d^{5}	2.868980e-05	2.068598e-04	0.104837	0.227665
$d^{0.1}$	0.842801	6.076787	0.816082	1.772218
$d^{0.01}$	0.974571	7.026878	0.978284	2.124458

TAB. 3: Data set comprising clusters with different density.

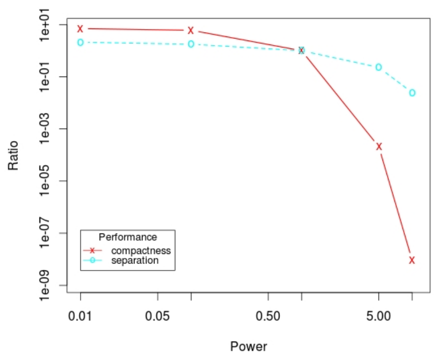


FIG. 4: Relation between Compactness and Separation Ratios for the test data set

We have shown that there is an inverse variation between ultrametricity and the performance of classifiers. The increase of such measure, obtained by transformations applied to the original space, promotes the equalization of distances. This equalization raises the level of uncertainty during the classification process and degrades the quality of the results generated by classifiers.

For clustering, increased ultrametricity generates clusterings with better separation. However, it also decreases compactness faster than the increase in separation. Lowering ultrametricity produces clusters that are more compact but not as well separated as in the original space. In this case, compactness grows at a faster ratio than the decrease in separation.

There are numerous applications that can benefit from this study. For example, changing the ultrametricity of the original space may help finding patterns in data that do not conform to the expected behavior, in a classical example of anomaly detection. The impact of ultrametricity on various hierarchical clustering algorithms also seems a promising subject of investigation.

References

Deza, M. M. and M. Laurent (1997). *Geometry of Cuts and Metrics*. Heidelberg: Springer.

Kaufman, L. and P. J. Rousseeuw (1990). *Finding Groups in Data – An Introduction to Cluster Analysis*. New York: John Wiley & Sons.

Lerman, I. C. (1981). *Classification et Analyse Ordinale des Données*. Paris: Dunod.

Liu, Y., Z. Li, H. Xiong, X. Gao, and J. Wu (2010). Understanding of internal clustering validation measures. In *2010 IEEE 10th International Conference on Data Mining*, pp. 911–916. IEEE.

Maulik, U. and S. Bandyopadhyay (2002). Performance evaluation of some clustering algorithms and validity indices. *IEEE Transactions on Pattern Analysis and Machine Intelligence 24*(12), 1650–1654.

Murtagh, F., G. Downs, and P. Contreras (2008). Hierarchical clustering of massive, high dimensional data sets by exploiting ultrametric embedding. *SIAM Journal on Scientific Computing 30*(2), 707–730.

Rammal, R., J. C. A. d'Auriac, and D. Doucot (1985). On the degree of ultrametricity. *Le Journal de Physique - Letteres 45*, 945–952.

Simovici, D. A. and C. Djeraba (2014). *Mathematical Tools for Data Mining* (second ed.). London: Springer.

Tang, P., M. Steinbach, and V. Kumar (2005). *Introduction to Data Mining*. Reading, MA: Addison-Wesley.

Xiong, H., J. Wu, and J. Chen (2009). K-means clustering versus validation measures: a data-distribution perspective. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 39*(2), 318–331.

Zhao, Y. and G. Karypis (2002). Evaluation of hierarchical clustering algorithms for document datasets. In *Proceedings of the Eleventh International Conference on Information and Knowledge Management*, pp. 515–524. ACM.

Résumé

Nous introduisons une mesure d'ultramétricité pour les dissimilaritées et examinons les transformations des dissimilaritées et leurs impact sur cette mesure. Ensuite, nous étudions l'influence de l'ultramétricité sur la comportement de deux classes d'algorithmes d'exploration de données (le kNN algorithme de classification et l'algorithme de regroupement PAM) appliqués sur les espaces de dissimilarité. On montre qu'il existe une variation inverse entre ultramétricité et la performance des classificateurs. Pour les clusters, une augmentation d'ultramétricité genere regroupements avec une meilleure séparation. Une diminution de la ultramétricité produit groupes plus compacts.