Taxonomy of Nominal Type Histogram Distance Measures

Sung-Hyuk Cha
Department of Computer Science, Pace University
861 Bedford rd, Pleasantville, NY 10570, USA

Abstract: - Distance or similarity measures are of fundamental importance to pattern classification, clustering, and information retrieval problems. Various distance/similarity measures that are applicable to compare two nominal type histograms are reviewed and categorized in both syntactic and semantic relationships. A correlation coefficient and a hierarchical clustering technique are adopted to reveal similarities among numerous distance/similarity measures.

Key-Words: - Distance, Metrics, Similarity

1 Introduction

Albeit the concept of *Euclidean* distance has prevailed in different cultures and regions for millennia, it is not a panacea for all types of data or pattern to be compared. The 20th century witnessed tremendous efforts to exploit new distance/similarity measures, *disim*'s in short, for a variety of applications. There are a substantial number of *disim*'s in short, encountered in many different fields such as anthropology, biology, chemistry, computer science, ecology, information theory, geology, mathematics, physics, psychology, statistics, etc.

Because it is often an essential key to solve many pattern recognition problems such as classification, clustering, and retrieval problems [1], there has been considerable effort in finding the appropriate measures among such a plethora of choices throughout different fields [2-5]. Notwithstanding, further comprehensive study is necessary because even names for certain *disim*'s are fluid and promulgated differently.

From the mathematical point of view, *distance* is defined as a quantitative degree of how far apart two objects are. Those distance measures satisfying the metric properties are simply called *metric* while other non-metric distance measures are occasionally called *divergence*. Similarity measures are often called *similarity coefficients*. A distance measure and a similarity measure are denoted as d_x and s_x, respectively throughout the rest of the paper.

The choice of *disim*'s depends on the measurement type or representation of objects. Here the *histogram*, which is one of the most popular pattern representations, is considered. Let d be the number of bins in the histogram. Two approaches vector and probability. There are different types of histograms [6]. Here only the *nominal type histogram* where each level or bin is independent from other levels or bins is considered and other types of histogram are abstained. Moreover, various *disim*'s that are applicable to compare two nominal type histograms are perambulated and categorized. All measures appearing in this paper have the shuffling invariant property [6] and thus naturally imply the level independency.

There are two approaches in histogram *disim*'s: vector and probabilistic. Since each level is assumed to be independent from other levels, a histogram can be considered as a vector, i.e., a point in the Euclidean space or a Cartesian coordinate system. Hence, numerous geometrical distances can be applied to compare histograms. There is much literature regarding discrete versions of various divergences between *probability density functions*, in short *pdf*s in probability and information theory fields [7,8]. Computing the distance between two pdfs can be regarded as the same as computing the *Bayes* (or minimum misclassification) probability [1]. This is equivalent to measuring the overlap between two pdfs as the distance. The probabilistic approach is based on the fact that a histogram of a measurement provides the basis for an empirical estimate of the pdf. A pdf for a corresponding histogram is produced by dividing each level by n. Let P and Q be the pdfs to be compared.

The rest of the paper is organized as follows. In section 2, various measures are enumerated according to their syntactic similarities. Section 3 presents the hierarchical cluster tree using the correlations between different measures. Finally, section 4 concludes this work.

2 Definitions

Table 1. L_p Minkowski family		
1. Euclidean L_2	$d_{Euc} = \sqrt{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert^2}$	(1)
2. City block L_1	$d_{CB} = \sum_{i=1}^{d} \lvert P_i - Q_i \rvert$	(2)

3. Minkowski L_p	$d_{Mk} = \sqrt[p]{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert^p}$	(3)
4. Chebyshev L_∞	$d_{Cheb} = \max_i \lvert P_i - Q_i \rvert$	(4)

A couple of thousand years ago, Euclid stated that the shortest distance between two points is a line and thus the eqn (1) is predominantly known as *Euclidean* distance. It was often called *Pythagorean* metric since it is derived from the *Pythagorean* theorem. In the late 19th century, Hermann Minkowski considered the *city block* distance [9]. Other names for the eqn (2) include *rectilinear* distance, *taxicab* norm, and *Manhattan* distance. Hermann also generalized the formulae (1) and (2) to the eqn (3) which is coined after Minkowski. When p goes to infinite, the eqn (4) can be derived and it is called the chessboard distance in 2D, the *minimax approximation*, or the *Chebyshev* distance named after Pafnuty Lvovich Chebyshev [10].

Table 2. L_1 family

5. Sørensen	$d_{sor} = \dfrac{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert}{\sum_{i=1}^{d} (P_i + Q_i)}$	(5)
6. Gower	$d_{gow} = \dfrac{1}{d} \sum_{i=1}^{d} \dfrac{\lvert P_i - Q_i \rvert}{R_i}$	(6)
	$= \dfrac{1}{d} \sum_{i=1}^{d} \lvert P_i - Q_i \rvert$	(7)
7. Soergel	$d_{sg} = \dfrac{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert}{\sum_{i=1}^{d} \max(P_i, Q_i)}$	(8)
8. Kulczynski *d*	$d_{kul} = \dfrac{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert}{\sum_{i=1}^{d} \min(P_i, Q_i)}$	(9)
9. Canberra	$d_{Can} = \sum_{i=1}^{d} \dfrac{\lvert P_i - Q_i \rvert}{P_i + Q_i}$	(10)
10. Lorentzian	$d_{Lor} = \sum_{i=1}^{d} \ln(1 + \lvert P_i - Q_i \rvert)$	(11)
* L_1 family $\supset$ {Intersectoin (13), Wave Hedges (15), Czekanowski (16), Ruzicka (21), Tanimoto (23), etc}.		

Several distance measures listed in Table 2 facilitate the L_1, more precisely the absolute difference. The eqn (5), which is widely used in ecology [11], is known as *Sørensen* distance [12] or *Bray-Curtis* [2,4,13]. When it is used for comparing two pdfs, it is nothing but the L_1 divided by 2. *Gower* distance [14] in the eqn (6) scales the vector space into the normalized space and then uses the L_1. Since the pdf is already normalized space, *Gower* distance is the L_1 divided by d. Other L_1 family distances that are non-proportional to the L_1 include *Soergel* and *Kulczynski* distances given in the eqns (8) [4] and (9) [2] respectively. At first glance, *Canberra* metric given in the eqn (10) [2,15] resembles *Sørensen* but normalizes the absolute difference of the individual level. It is known to be very sensitive to small changes near zero [15]. The eqn (11) [2], attributed to *Lorentzian*, also contains the absolute difference and the natural logarithm is applied. 1 is added to guarantee the non-negativity property and to eschew the log of zero.

Table 3. Intersection family

11. Intersection	$s_{IS} = \sum_{i=1}^{d} \min(P_i, Q_i)$	(12)
$d_{non\text{-}IS} = 1 - s_{IS}$	$d_{non-IS} = \dfrac{1}{2} \sum_{i=1}^{d} \lvert P_i - Q_i \rvert$	(13)
12. Wave Hedges	$d_{WH} = \sum_{i=1}^{d} \left(1 - \dfrac{\min(P_i, Q_i)}{\max(P_i, Q_i)}\right)$	(14)
	$= \sum_{i=1}^{d} \dfrac{\lvert P_i - Q_i \rvert}{\max(P_i, Q_i)}$	(15)
13. Czekanowski	$s_{Cze} = \dfrac{2\sum_{i=1}^{d} \min(P_i, Q_i)}{\sum_{i=1}^{d} (P_i + Q_i)}$	(16)
$d_{Cze} = 1 - s_{Cze}$	$d_{Cze} = \dfrac{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert}{\sum_{i=1}^{d} (P_i + Q_i)}$	(17)
14. Motyka	$s_{Mot} = \dfrac{\sum_{i=1}^{d} \min(P_i, Q_i)}{\sum_{i=1}^{d} (P_i + Q_i)}$	(18)
$D_{Mot} = 1 - s_{Mot}$	$d_{Mot} = \dfrac{\sum_{i=1}^{d} \max(P_i, Q_i)}{\sum_{i=1}^{d} (P_i + Q_i)}$	(19)
15. Kulczynski *s*	$s_{Kul} = \dfrac{\sum_{i=1}^{d} \min(P_i, Q_i)}{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert}$	(20)
16. Ruzicka	$s_{Ruz} = \dfrac{\sum_{i=1}^{d} \min(P_i, Q_i)}{\sum_{i=1}^{d} \max(P_i, Q_i)}$	(21)
17. Tanimoto	$d_{Tani} = \dfrac{\sum_{i=1}^{d} P_i + \sum_{i=1}^{d} Q_i - 2\sum_{i=1}^{d} \min(P_i, Q_i)}{\sum_{i=1}^{d} P_i + \sum_{i=1}^{d} Q_i - \sum_{i=1}^{d} \min(P_i, Q_i)}$	(22)
	$= \dfrac{\sum_{i=1}^{d} (\max(P_i, Q_i) - \min(P_i, Q_i))}{\sum_{i=1}^{d} \max(P_i, Q_i)}$	(23)

The *intersection* between two pdfs in the eqn (12) is a widely used form of similarity [1] where the non-overlaps between two pdfs defined in the eqn (13) is nothing but the L_1 divided by 2 [6]. Hence, most similarity measures pertinent to the intersection enumerated in Table 3 can be transformed into the L_1

based distance measures using the technique, i.e., $d_x(P,Q)$=1-$s_x(P,Q)$ with a few of exceptions. The eqn (14) is called *Wave Hedges* [16] and its L_1 based distance form is given in the eqn (15). *Czekanowski* Coefficient in the eqn (16) [15] has its distance form identical to *Sørensen* (5). Half of the *Czekanowski* Coefficient is called *Motyka* similarity in the eqn (18) [2]. The eqn (20) is known as *Kulczynski* similarity [2]. The eqn (22) is referred to as *Tanimoto* distance [1] a.k.a., *Jaccard* distance. *Soergel* distance in the eqn (8) is identical to *Tanimoto*. 1 - d_{Tani} is *Ruzicka* similarity given in the eqn (21) [2]. The eqn (23) is given to help understand their equivalencies.

Table 4. Inner Product family		
18. Inner Product	$s_{IP} = P \bullet Q = \sum_{j=1}^{d} P_i Q_i$	(24)
19. Harmonic mean	$s_{HM} = 2\sum_{i=1}^{d} \frac{P_i Q_i}{P_i + Q_i}$	(25)
20. Cosine	$s_{Cos} = \frac{\sum_{i=1}^{d} P_i Q_i}{\sqrt{\sum_{i=1}^{d} P_i^2}\sqrt{\sum_{i=1}^{d} Q_i^2}}$	(26)
21. Kumar-Hassebrook (PCE)	$s_{Jac} = \frac{\sum_{i=1}^{d} P_i Q_i}{\sum_{i=1}^{d} P_i^2 + \sum_{i=1}^{d} Q_i^2 - \sum_{i=1}^{d} P_i Q_i}$	(27)
22. Jaccard	$s_{Jac} = \frac{\sum_{i=1}^{d} P_i Q_i}{\sum_{i=1}^{d} P_i^2 + \sum_{i=1}^{d} Q_i^2 - \sum_{i=1}^{d} P_i Q_i}$	(28)
d_{Jac}= 1-s_{Jac}	$d_{Jac} = \frac{\sum_{i=1}^{d} (P_i - Q_i)^2}{\sum_{i=1}^{d} P_i^2 + \sum_{i=1}^{d} Q_i^2 - \sum_{i=1}^{d} P_i Q_i}$	(39)
23. Dice	$s_{Dice} = \frac{2\sum_{i=1}^{d} P_i Q_i}{\sum_{i=1}^{d} P_i^2 + \sum_{i=1}^{d} Q_i^2}$	(40)
d_{Dice}= 1-s_{Dice}	$d_{Dice} = \frac{\sum_{i=1}^{d} (P_i - Q_i)^2}{\sum_{i=1}^{d} P_i^2 + \sum_{i=1}^{d} Q_i^2}$	(31)

Table 4 deals exclusively with similarity measures which incorporate the *inner product* , P•Q explicitly in their definitions. The *inner product* of two vectors in the eqn (24) yields a scalar and is sometimes called the *scalar product* or *dot product* [1]. The inner product is also called the *number of matches* or the *overlap* if it is used for binary vectors. The eqn (25) is the *harmonic mean* [2]. The eqn (26) is the normalized inner product and called the *cosine* coefficient because it measures the angle between two vectors and thus often called the *angular metric* [2]. Other names for the *cosine* coefficient include *Ochiai* [2,4] and *Carbo* [4]. Kumar and Hassebrook utilized P•Q to measure the *Peak-to-correlation energy*, *PCE* in short [17] in the eqn (27). *Jaccard* coefficient [18], a.k.a. *Tanimoto* [19], defined in the eqn (28) is another variation of the normalized inner product. *Dice* coefficient in the eqn (30) [20] is occasionally called Sorensen, Czekannowski, Hodgkin-Richards [4] or Morisita [21]. The eqns (24,26,28,30) are frequently encountered similarity measures in the fields of information retrieval and biological taxonomy for the binary feature vector comparison (see [2,22] for the exhaustive list of distance and similarity measures for the binary feature vectors).

Table 5. Fidelity family or Squared-chord family		
24. Fidelity	$s_{Fid} = \sum_{i=1}^{d} \sqrt{P_i Q_i}$	(32)
25. Bhattacharyya	$d_B = -\ln \sum_{i=1}^{d} \sqrt{P_i Q_i}$	(33)
26. Hellinger	$d_H = \sqrt{2\sum_{i=1}^{d} (\sqrt{P_i} - \sqrt{Q_i})^2}$	(34)
	$= 2\sqrt{1 - \sum_{i=1}^{d} \sqrt{P_i Q_i}}$	(35)
27. Matusita	$d_M = \sqrt{\sum_{i=1}^{d} (\sqrt{P_i} - \sqrt{Q_i})^2}$	(36)
	$= \sqrt{2 - 2\sum_{i=1}^{d} \sqrt{P_i Q_i}}$	(37)
28. Squared-chord	$d_{sqc} = \sum_{i=1}^{d} (\sqrt{P_i} - \sqrt{Q_i})^2$	(38)
s_{sqc}= 1-d_{sqc}	$s_{sqc} = 2\sum_{i=1}^{d} \sqrt{P_i Q_i} - 1$	(39)

The sum of *geometric means* in the eqn (32) is referred to as *Fidelity* similarity, a.k.a. *Bhattacharyya* coefficient or *Hellinger affinity* [2]. *Bhattacharyya* distance given in the eqn (33), which is a value between 0 and 1, provides bounds on the *Bayes* misclassification probability [23]. Other approaches closely related to *Bhattacharyya* include *Hellinger* [2] and *Matusita* [24] in eqns (34) and (36) respectively. The basic form in the eqn (38), i.e., *Matusita* without the square root is called *Squared-chord* distance [5] and thus all *Fidelity* based measures have their alternative representation using the *squared-chord* distance.

Table 6. Squared L_2 family or χ^2 family		
29. Squared Euclidean	$d_{sqe} = \sum_{i=1}^{d} (P_i - Q_i)^2$	(40)
30. Pearson χ^2	$d_P(P,Q) = \sum_{i=1}^{d} \frac{(P_i - Q_i)^2}{Q_i}$	(41)
31. Neyman χ^2	$d_N(P,Q) = \sum_{i=1}^{d} \frac{(P_i - Q_i)^2}{P_i}$	(42)

32. Squared χ^2	$d_{SqChi} = \sum_{i=1}^{d} \frac{(P_i - Q_i)^2}{P_i + Q_i}$	(43)
33. Probabilistic Symmetric χ^2	$d_{PChii} = 2\sum_{i=1}^{d} \frac{(P_i - Q_i)^2}{P_i + Q_i}$	(44)
34. Divergence	$d_{Div} = 2\sum_{i=1}^{d} \frac{(P_i - Q_i)^2}{(P_i + Q_i)^2}$	(45)
35. Clark	$d_{Clk} = \sqrt{\sum_{i=1}^{d} \left(\frac{\lvert P_i - Q_i \rvert}{P_i + Q_i} \right)^2}$	(46)
36. Additive Symmetric χ^2	$d_{AdChi} = \sum_{i=1}^{b} \frac{(P_i - Q_i)^2 (P_i + Q_i)}{P_i Q_i}$	(47)
* Squared L_2 family $\supset$ {Jaccard (29), Dice (31)}		

Distance measures containing the *Squared Euclidean* distance in the eqn (40) as the dividend are corralled in Table 6. *Jaccard* and *Dice* distance forms in the eqns (29) and (31) also belong to this family. The cornerstone to the χ^2 family (eqns (41)~(47)) is *Pearson* χ^2 divergence in the eqn (41) [25] which embodies the *Squared Euclidean* distance. *Pearson* χ^2 divergence is asymmetric. *Neyman* χ^2 in the eqn (42) [26] is $d_N(P,Q)=d_P(Q,P)$. Various symmetric versions of the χ^2 have been exploited. The eqn (43) is called the *squared* χ^2 distance [5] or *triangular discrimination* [27,28]. Twice of the eqn (44) is called the *probabilistic symmetric* χ^2 [2] which is equivalent to *Sangvi* χ^2 distance between populations [2]. The term 'divergence' is pronominal to refer non-metric distance. Notwithstanding the eqn (45) has been commonly called *divergence* [29]. The squared root of half of the divergence is called *Clark* in the eqn (46) [2]. The eqn (47) is $d_{AdChi}(P,Q) = d_P(P,Q) + d_P(Q,P)$ [2,3]. Albeit the eqn (47) is occasionally called '*symmetric* χ^2 divergence', let's call it the *additive symmetric* χ^2 here in order to distinguish other symmetric versions of χ^2.

Table 7. Shannon's entropy family		
37. Kullback–Leibler	$d_{KL} = \sum_{i=1}^{d} P_i \ln \frac{P_i}{Q_i}$	(48)
38. Jeffreys	$d_J = \sum_{i=1}^{d} (P_i - Q_i) \ln \frac{P_i}{Q_i}$	(49)
39. K divergence	$d_{Kdiv} = \sum_{i=1}^{d} P_i \ln \frac{2P_i}{P_i + Q_i}$	(50)
40. Topsøe	$d_{Top} = \sum_{i=1}^{d} \left(P_i \ln\left(\frac{2P_i}{P_i + Q_i} \right) + Q_i \ln\left(\frac{2Q_i}{P_i + Q_i} \right) \right)$	(51)
41. Jensen-Shannon	$d_{JS} = \frac{1}{2}\left[\sum_{i=1}^{d} P_i \ln\left(\frac{2P_i}{P_i + Q_i} \right) + \sum_{i=1}^{d} Q_i \ln\left(\frac{2Q_i}{P_i + Q_i} \right) \right]$	(52)
42. Jensen difference	$d_{JD} = \sum_{i=1}^{b} \left[\frac{P_i \ln P_i + Q_i \ln Q_i}{2} - \left(\frac{P_i + Q_i}{2} \right) \ln\left(\frac{P_i + Q_i}{2} \right) \right]$	(53)

Eqns in Table 7 are primary due to Shannon's concept of probabilistic uncertainty or "entropy" $H(P)=\Sigma^{d}_{i=1} P_i \ln P_i$ [30]. Kullback and Leibler [31] introduced the eqn (48) called KL divergence, relative entropy, or information deviation [2]. The symmetric form of the KL divergence using the addition method is in the eqn (49) [31-33] and it is called Jeffreys or J divergence. The eqn (50) is called the K divergence and its symmetric form is given in the eqn (51) and called Topsøe distance [2] or information statistics [5]. The half of the Topsøe distance is called Jensen-Shannon divergence [2,34]. Sibson [35] studied the idea of information radius for a measure arising due to concavity property of Shannon's entropy and introduced the Jensen difference in the eqn (53) [33]. All eqns (48~53) can be expressed in terms of entropy.

Table 8. Combinations		
43. Taneja	$d_{TJ} = \sum_{i=1}^{d} \left(\frac{P_i + Q_i}{2} \right) \ln\left(\frac{P_i + Q_i}{2\sqrt{P_i Q_i}} \right)$	(54)
44. Kumar-Johnson	$d_{KJ} = \sum_{i=1}^{d} \left(\frac{(P_i^2 - Q_i^2)^2}{2(P_i Q_i)^{3/2}} \right)$	(55)
45. Avg(L_1, L_∞)	$d_{ACC} = \frac{\sum_{i=1}^{d} \lvert P_i - Q_i \rvert + \max_i \lvert P_i - Q_i \rvert}{2}$	(56)

Table 8 exhibits distance measures utilizing multiple ideas or measures. Taneja utilized both arithmetic and geometric means came up with the arithmetic and geometric mean divergence in the eqn (54) [36]. Symmetric χ^2, arithmetic and geometric mean divergence is given in the eqn (55) [37]. The average of city block and Chebyshev distances in the eqn (56) appears in [9].

3 Hierarchical Clustering

Hitherward, the focus is moved from the syntactic similarity to the semantic similarity between *disim*'s. So as to assess how similar distance measures are, the following experiments were conducted using the cluster analysis. n samples whose values are between 1 and d are randomly selected to build a histogram. Next, each bin is divided by n to produce the pdf. Let R be the set of r number of reference pdfs and q be a query pdf. Then r number of distance values are produced using a certain distance measure $d_x(r_i,q)$ for $\forall i$. r_i and q are randomly generated pdfs.

Fig 1 presents the upper triangle matrix of correlation between $d_x(r_i,q)$ and $d_y(r_i,q)$ plots for selected distance or similarity measures where n = 20, b = 8, and r = 30. Each plot in Figure 2 represents the relation between two distance measures. In order to quantify the correlation between *disim*'s, a correlation coefficient measure in the eqn (57) is used.

$$\mathrm{Corr}(d_x, d_y) = \frac{\sum_{i=1}^{r}(d_x(r_i,q) - \overline{d_x})(d_y(r_i,q) - \overline{d_y})}{\sqrt{\sum_{i=1}^{r}(d_x(r_i,q) - \overline{d_x})^2 \sum_{i=1}^{r}(d_y(r_i,q) - \overline{d_y})^2}} \qquad (57)$$

$$\text{where } \overline{d_x} = \frac{\sum_{i=1}^{r} d_x(r_i,q)}{r}$$

It indicates the strength and direction of a linear relationship between two distance measures. If the value gets close to 1, it represents a good fit, i.e., two distance measures are semantically similar. As the fit gets worse, the correlation coefficient approaches zero. When either two distance or two similarity measures are compared, the correlation coefficient is a positive value. When a distance measure and a similarity measure are compared, the correlation coefficient is a negative value e.g., the squared χ^2 and probabilistic symmetric χ^2 divergences have d_{SsqChi}= .5 d_{PrChi} and Corr (d_{SsqChi}, d_{PrChi}) = 1 whereas Motyka similarity (20) and Sørensen (5) have s_{Mot} = 1 - d_{Sor} and Corr (s_{Mot}, d_{Sor}) = -1.

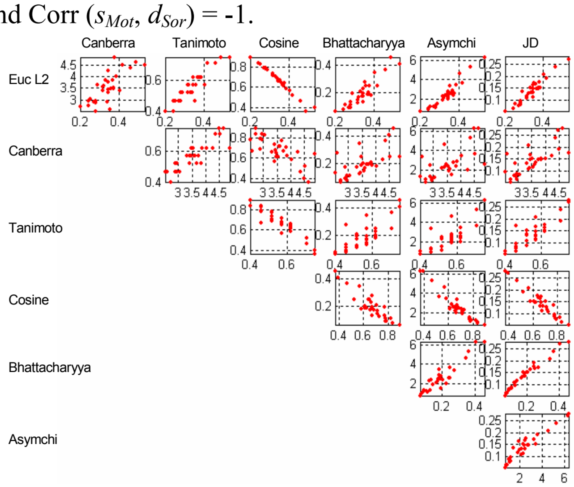

Fig 1. Upper triangle matrix of correlation plots between two selected disim's.

To adequately understand the similarities among *disim*'s, cluster analysis is adopted. The correlation coefficient is converted into the distance in the eqn (59) to find clusters of *disim*'s shown in Fig 2.

$$d_{DM}(d_x, d_y) = 1 - |\mathrm{Corr}(d_x, d_y)| \qquad (58)$$

The dendrogram representing the hierarchical clusters of disim's is produced by averaging 30 independent trials of the above experiment. It is built using the agglomerative single linkage with the average clustering method [1]. The vertical scale represents various *disim*'s and the horizontal scale represents the closeness between two clusters.

The dendrogram representing the hierarchical clusters of *disim*'s is produced by averaging 30 independent trials of the above experiment. It is built using the agglomerative single linkage with the average clustering method [1]. The vertical scale on the left represents various *disim*'s and the horizontal scale represents the closeness between two clusters of *disim*'s. The dendrogram provides intuitive groupings of *disim*'s. Some distance measures in syntactic groups are interspersed in the semantic groups. Here are a few simple observations.

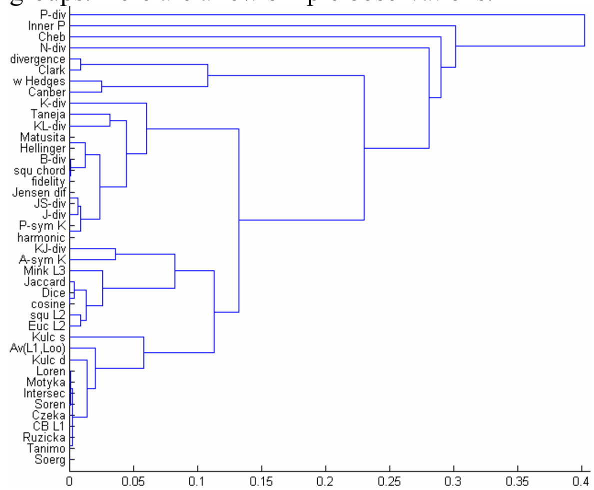

Fig 2. Hierarchical Clusters of disim's.

Observation 1: if two measures are proportional to each other, i.e., $d_x = cd_y$, $d_{DM}(d_x, d_y) = 0$.

Observation 2: if two measures are in distance/ similarity relation s.t. $d_x = 1 - s_y$, $d_{DM}(d_x, d_y) = 0$.

Observation 3: if two measures are in distance/ similarity relation such that $s_y = 1/d_x$, $d_{DM}(d_x, d_y) \geq 0$. e.g, *Kulczynski* has $s_{kul} = 1/d_{kul}$ and $d_{DM}(s_{kul}, d_{kul}) > 0$.

4 Conclusion

This article built the edifice of *disim*'s by enumerating and categorizing a large variety of *disim*'s for comparing nominal type histograms. Grouping aforementioned measures has concentrated upon two general aspects: syntactic similarity and semantics. The importance of finding suitable *disim*'s cannot be overemphasized. There is a continual demand for better ones.

References:

1. R.O. Duda, P.E. Hart, and D.G. Stork, Pattern Classification, 2nd ed. Wiley, 2001
2. E. Deza and M.M. Deza, Dictionary of Distances, Elsevier, 2006
3. P. Zezula, G. Amato, V. Dohnal, and M. Batko, Similarity Search The Metric Space Approach, Springer, 2006
4. V. Monev, Introduction to Similarity Searching in Chemistry, MATCH Commun. Math. Comput. Chem. 51 pp. 7-38 , 2004
5. D.G. Gavin, W.W. Oswald, E.R. Wahl, and J.W. Williams, A statistical approach to evaluating

distance metrics and analog assignments for pollen records, Quaternary Research 60, pp 356–367, 2003
6. S. Cha and S. N. Srihari, On Measuring the Distance between Histograms, in Pattern Recognition, Vol 35/6, pp 1355-1370, June 2002
7. T. Kailath, The divergence and bhattacharyya distance measures in signal selection, IEEE Trans. Commun. Technol. COM-15 (1) (1967) 52–60.
8. G.T. Toussaint, Bibliography on estimation of misclassification, IEEE Trans. Inform. Theory 20 (4) (1974) 472–479. pp. 21–24.
9. E.F. Krause, Taxicab Geometry An Adventure in Non-Euclidean Geometry
10. F. Heijden, R. Duin, D. Ridder, D. Tax, Classification, Parameter Estimation and State Estimation: An Engineering Approach Using MATLAB. John Wiley and Sons, 2004.
11. J. Looman and J.B. Campbell, Adaptation of Sorensen's K (1948) for estimating unit affinities in prairie vegetation. Ecology 41 (3): 409-416, 1960
12. T. Sørensen, A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. Biologiske Skrifter / Kongelige Danske Videnskabernes Selskab, 5 (4): 1-34, 1948.
13. J.R. Bray and J.T. Curtis, An ordination of the upland forest of the southern Winsconsin. Ecological Monographies, 27, 325-349, 1957.
14. J.C. Gower, General Coefficient of Similarity and Some of Its Properties, Biometrics 27, pp857-874 1971
15. A.D. Gordon, Classification. 2nd ed. Chapman & Hall/CRC, 1999
16. T.S. Hedges, 1976, An empirical modification to linear wave theory, Proc. Inst. Civ. Eng., 61, 575-579.
17. B. V. K. Vijaya Kumar and L. G. Hassebrook, "Performance measures for correlation filters," Appl. Opt. 29, 2997-3006, 1990.
18. P. Jaccard, Étude comparative de la distribution florale dans une portion des Alpes et des Jura. Bulletin del la Société Vaudoise des Sciences Naturelles 37, 1901, 547-579.
19. T.T. Tanimoto, IBM Internal Report Nov. 1957.
20. L. R. Dice, Measures of the amount of ecologic association between species, Ecology, 26:297-302, 1945
21. M. Morisita, Measuring of interspecific association and similarity between communities. Mem. Fac. Sci. Kyushu Univ. Ser. E (Biol.) 3:65-80, 1959.
22. S. Cha, S. Yoon, and C.C. Tappert, Enhancing Binary Feature Vector Similarity Measures, in Journal of Pattern Recognition Research, Vol 1 No 1, pp 63-77, 2006
23. A. Bhattacharyya, On a measure of divergence between two statistical populations defined by probability distributions, Bull. Calcutta Math. Soc., vol. 35, pp. 99–109, 1943.
24. K. Matusita, Decision rules, based on the distance, for problems of fit, two samples, and estimation, Ann. Math. Statist. 26 631–640, 1955
25. K. Pearson, On the Criterion that a given system of deviations from the probable in the case of correlated system of variables is such that it can be reasonable supposed to have arisen from random sampling, Phil. Mag., 50, 157-172, 1900.
26. J. Neyman. Contributions to the theory of the χ^2 test. In Proc. of the First Berkley Symposium on Mathematical Statistics and Probability, 1949.
27. S. S. DRAGOMIR, J. SUNDE and C. BUSE, New Inequalities for Jeffreys Divergence Measure, Tamsui Oxford Journal of Mathematical Sciences, 16(2), 295-309, 2000.
28. F. TOPSØE, Some Inequalities for Information Divergence and Related Measures of Discrimination, IEEE Trans. on Inform. Theory, IT-46, 1602-1609, 2000.
29. T.F. Cox and M.A.A. Cox, Multidimensional Scaling, Chapman & Hall/CRC 2nd ed. 2001
30. C.E. Shannon, A mathematical theory of communication, Bell System Tech. J. vol 27, 379–423, 623–656, 1948.
31. S. Kullback, R.A. Leibler, On information and sufficiency, Ann. Math. Statist. 22, 79–86, 1951.
32. H. Jeffreys, An Invariant Form for the Prior Probability in Estimation Problems, Proc. Roy. Soc. Lon., Ser. A, 186, 453-461, 1946.
33. I.J. Taneja, Generalized Information Measures and Their Applications, 2001 on-line book: www.mtm.ufsc.br/~taneja/book/book.html
34. J. Lin, Divergence measures based on the Shannon entropy. IEEE Trans. on Information Theory, 37(1):145–151, 1991.
35. R. Sibson, Information Radius, Z. Wahrs. und verw Geb., 14, 149-160, 1969.
36. I.J. Taneja, New Developments in Generalized Information Measures, in Advances in Imaging and Electron Physics, 37-135, 1995.
37. P. Kumar and A. Johnson, On a symmetric divergence measure and information inequalities, Journal of Inequalities in pure and applied Mathematics, Vol 6, Issue 3, article 65, 2005.