Measuring Distance Between Unordered Sets of Different Sizes

Andrew Gardner,* Jinko Kanno*
{abg010, jkanno}@latech.edu

Christian A. Duncan†
christian.duncan@quinnipiac.edu

Rastko Selmic*
rselmic@latech.edu

*Louisiana Tech University, Ruston, LA 71270
†Quinnipiac University, Hamden, CT 06518

Abstract

We present a distance metric based upon the notion of minimum-cost injective mappings between sets. Our function satisfies metric properties as long as the cost of the minimum mappings is derived from a semimetric, for which the triangle inequality is not necessarily satisfied. We show that the Jaccard distance (alternatively biotope, Tanimoto, or Marczewski-Steinhaus distance) may be considered the special case for finite sets where costs are derived from the discrete metric. Extensions that allow premetrics (not necessarily symmetric), multisets (generalized to include probability distributions), and asymmetric mappings are given that expand the versatility of the metric without sacrificing metric properties. The function has potential applications in pattern recognition, machine learning, and information retrieval.

1. Introduction

Measuring distance between objects plays an important role in various disciplines including data mining, machine learning, and information retrieval. Often the objects considered are comprised of multiple parts or are collections of other objects. A set or a tuple are examples of such complex objects. Assuming that the set or tuple is ordered and can be represented as a vector, one may easily define a distance (such as standard Euclidean). However, often one makes an assumption that is not always true; namely, that the i-th index of a vector $\mathbf{x}$ corresponds to the i-th index of a vector $\mathbf{y}$. If no such correspondence exists (or it is unknown), then one must usually resort to some less precise measure of distance. The Hausdorff distance, named after German mathematician Felix Hausdorff, is an example of such a distance that reduces the comparison of both tuples to a comparison of just an individual element from each. In particular, the definition of the Hausdorff distance $H(A, B)$ between two sets A and B is given by [14] as

$$H(A, B) = \sup\left[h(A, B), h(B, A)\right], \tag{1}$$

where

$$h(A, B) = \sup_{a \in A,\, b \in B} \inf D(a, b), \tag{2}$$

and $D(a, b)$ is some metric on the elements. A Hausdorff-like distance has also been established for fuzzy sets [4]. Unfortunately, the simplicity of its definition renders the Hausdorff distance vulnerable to outliers and potentially an inaccurate estimate of one's intuitive notion of distance. The Jaccard distance, or the more general Steinhaus distance [5], is useful for comparing two unordered sets and is defined to be the ratio of the sets' symmetric difference over their union. The Jaccard distance is inflexible and does not account for partial similarity between elements. The Jaccard distance, though, is what served as the inspiration for the form of our metric, and it is discussed more fully in Section 3. Another metric, similar to what we propose, is called the Earth Mover's Distance (EMD) [23] (or equivalently Wasserstein or Mallows [17] for sets of equal mass), which intuitively measures the amount of one set that must be altered to transform it into the other. Gaspard Monge laid the groundwork for the EMD in 1781 [19], and the problem was reformulated in the mid-20th century by Leonid Kantorovich [16] [15]. The EMD, however, is only a true metric for sets of the same size, which provides a motivation for our metric. Figalli and Gigli [9] propose an extension to the Wasserstein metric for distributions of unequal measure by allowing transportation to and from an external boundary. Fujita [10] proposes a generalized metric similar in spirit to our own that is based upon the average distance between sets and can even be seen as an alternative generalization of the Jaccard and other distance metrics. Eiter and Mannila [8] propose multiple distance functions based upon various types of mappings between sets, but in general their functions fail to satisfy all metric properties or are difficult, even NP-Hard, to compute. In the most similar work to our own [22], an optimal (effectively minimum-cost) matching metric between sets is proposed that assigns an arbitrary penalty to unmatched elements. A normalized form of the metric with a unit penalty is also presented, assuming that the metric used to compare elements is also bounded. In ad-

dition, the netflow distance is proposed that assigns integer weights to elements and allows them to be matched multiple times. We will show that the normalized form is a subcase of our metric.

Our primary motivation for deriving the proposed distance metric, or the *minimum mapping metric*, arises from the desire to compare unlabeled, unordered point sets (or *frames*) generated by a motion capture camera system. We wish to define a distance between sets of points that takes into account their structure, relative cardinalities, and locations. For clarification, two identical structures in different locations or orientations should be considered different, as well as one point set that is simply a subset of another. In particular, the main idea is to address the failure of metric properties of EMD on sets of unequal size. According to EMD, two sets may be at a large distance from each other but a distance of zero from a common superset. We use the Jaccard distance as a template for deriving an expression that addresses this shortcoming. In a sense, our metric may be considered a normalized version of EMD.

2. Preliminaries

A metric M on a space X is a function $M : X \times X \mapsto \mathbb{R}$ that satisfies certain properties. The range of a metric is non-negative. The function is also symmetric; the order of the given inputs has no effect on the output. The identity property must also hold where the distance between two inputs is zero if and only if the inputs are the same. Finally, a metric M must satisfy the triangle inequality,

$$M(x, y) \leq M(x, z) + M(y, z), \tag{3}$$

where x, y and z are any three possible inputs. The Euclidean distance is a metric. We define the term semimetric to indicate satisfaction of all of the preceding properties except for the triangle inequality. An example of a semimetric is the squared Euclidean distance, and a simple example of it failing the triangle inequality may be noted with the points $x = (0, 0)$, $y = (0, 1)$, and $z = (0, 2)$ as elements of $\mathbb{R}^2$, for which the pair-wise distances are 1, 1, and 4.

A measure μ is a function on a set that generalizes the notion of area, volume, or length. The measure of a subset is less than or equal to that of its superset, *i.e.* $\mu(A) \leq \mu(B)$ if $A \subseteq B$. Measures also possess countable additivity, *i.e.* the measure of N disjoint sets is the sum of their measures. We assume that we are dealing with finite, non-negative measurable sets for the remainder of the paper. A measure space (X, μ) is a space X paired with a measure μ. Cardinality is sometimes referred to as the counting measure.

Concerning notation, we will be using $|A|$ to denote the cardinality of set A. We will use the term *discrete metric* to refer to the 0-1 distance, where $d(x, y) = 0$ if $x = y$ and $d(x, y) = 1$ otherwise. *Discrete measure* will refer to cardinality.

3. Proposed Distance Metric

In this section we define the minimum mapping metric and show that the Jaccard distance and normalized metric of Ramon and Bruynooghe are special cases of it. In essence, given two sets and a function that describes the cost of mapping two elements, we find a minimum-cost injective mapping from the smaller to the larger set. Similar to the EMD, this mapping may be thought of as the cost of transforming one set into the other. Unlike the EMD, however, each set may possess its own "point-of-view" (with restrictions) for how costly the mapping is.

The minimum mapping metric is based upon a specific interpretation of frames, namely that they are sets in the sense of set theory as opposed to geometric constructs. Consequently, it is natural to base our distance function upon the Jaccard index between two sets A and B:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}, \tag{4}$$

which is equal to the cardinality of the intersection of A and B over the cardinality of their union. Of particular importance is the fact that a distance metric J_D can be defined on the complement of $J(A, B)$ [18],

$$J_D(A, B) = 1 - J(A, B) = \frac{|A \cup B| - |A \cap B|}{|A \cup B|}. \tag{5}$$

The generalized version of the Jaccard distance, defined for arbitrary measure μ, is the Steinhaus distance S_D and can be expressed analagously to Equation 5, *e.g.*

$$S_D(A, B) = \frac{\mu(A \cup B) - \mu(A \cap B)}{\mu(A \cup B)}. \tag{6}$$

The primary obstacle that inhibits us from directly applying the Jaccard distance to frames is the fact that it is defined in terms of binary set membership and element identity; an element is either in a set or not, and two elements are either completely equal or not at all. Ideally, especially considering potential noise and error in measurements, we wish to allow a degree of uncertainty in element identity. More formally, the Jaccard distance uses a discrete metric for element comparison, and we wish to allow other metrics. Brouwer [2] discusses an extension of the Jaccard index to fuzzy sets, thereby addressing the binary set membership issue. However, there does not appear to be much discussion concerning whether the fuzzy Jaccard distance defined as the complement of the fuzzy Jaccard index satisfies metric properties.

We will first introduce subsidiary terms of the minimum mapping metric. We assume that we have been given a measure space (X, μ) along with a semimetric $m_D : X \times X \mapsto \mathbb{R}$. We will call m_D the ground distance. In addition we assume that we have been given a function,

$\nu : \mathcal{F}(X) \mapsto \mathbb{R}$, where $\mathcal{F}(X)$ is the family of all subsets of X with non-negative, finite measure, such that for any A and B, $\nu(A) \leq \nu(B)$ if $A \subseteq B$. For example, ν could be set diameter, constant, or even μ. Given two measurable sets A and B and assuming $\mu(A) \leq \mu(B)$, we seek a minimum-cost injective mapping Ψ^* from A to B such that $\mu(\Psi^*(A)) = \mu(A)$, *i.e.* the mapping preserves measure. In particular, the cost function we wish to minimize is

$$\Psi^* = \underset{\Psi: A \mapsto B}{\arg\inf} \left[\delta_A(\Psi) + \delta_B(\Psi) + \delta_{AB}(\Psi)\right], \tag{7}$$

where

$$\delta_A(\Psi) = \int_A \min\left[1, \frac{m_D(a, \Psi(a))}{\nu(A)}\right] d\mu(a), \tag{8}$$

$$\delta_B(\Psi) = \int_A \min\left[1, \frac{m_D(a, \Psi(a))}{\nu(B)}\right] d\mu(a), \tag{9}$$

$$\delta_{AB}(\Psi) = \int_A \min\left[1, \frac{m_D(a, \Psi(a))}{\nu(A \cup B)}\right] d\mu(a). \tag{10}$$

Let

$$\delta_Z = \delta_Z(\Psi^*), \tag{11}$$

where Z stands for A, B, or AB. Let us also define

$$\mu(A, B) = \max\left[\mu(A), \mu(B)\right], \tag{12}$$

$$\mu'(A, B) = \min\left[\mu(A), \mu(B)\right]. \tag{13}$$

We are now ready to define the minimum mapping metric.

DEFINITION 1. Given a measure space (X, μ) and a semi-metric $m_D : X \times X \mapsto \mathbb{R}$, define $\mathcal{F}(X)$ to be the family of all subsets of X with non-negative, finite measure and the minimum mapping metric $M : \mathcal{F}(X) \times \mathcal{F}(X) \mapsto \mathbb{R}$ to be

$$M(A, B) = \frac{\mu(A, B) + \delta_A + \delta_B - \mu'(A, B)}{\mu(A, B) + \delta_A + \delta_B - \delta_{AB}} \tag{14}$$

where $A, B \in \mathcal{F}(X)$.

EXAMPLE 1. The Jaccard distance is the special case where μ is the discrete measure, the ground distance is the discrete metric and $\nu(X) \leq 1$. The Steinhaus distance is the general Jaccard distance for any measure with the discrete metric and $\nu(X) \leq 1$.

EXAMPLE 2. The normalized form of the metric proposed in [22] is the special case of the minimum mapping metric where μ is the discrete measure, the ground distance is a normalized metric with range $[0, 1]$, and ν is constant.

An upper bound and lower bound for $M(A, B)$ may be easily computed given two sets A and B, and are presented here without proof.

THEOREM 1. Given two sets A and B with $\mu(A) \leq \mu(B)$, then

$$\frac{\mu(B) - \mu(A)}{\mu(B)} \leq M(A, B) \leq \frac{\mu(A \cup B) - \mu(A \cap B)}{\mu(A \cup B)}. \tag{15}$$

One may note that the upper bound is the Steinhaus distance. Of particular importance is the fact that $M(A, B)$, in all of its possible forms, is a metric:

THEOREM 2. The function $M(A, B)$ as defined by Equation 14 is a metric.

The entire proof (mostly concerned with the triangle inequality) is too long to present here, so a short sketch is provided instead. Since the sets and measures are not necessarily discrete or countable, we use a variant of continuous mathematical induction [3]. The induction is indexed by a real variable k over the interval $[0, 1]$, progressing in infinitesimal steps of size dk from 0 to 1. In essence, we start with a combination of sets for which we know the triangle inequality is satisfied. We then make small modifications to one of the sets such that we produce the cases for which we wish to prove the triangle inequality. An example of a case considered is that of three sets A, B, and C where A is a subset of C and $\mu(B) \leq \mu(C)$. The initial step for this example is that of three sets A, B, and C with $A \cup B \subseteq C$, where at each step of the induction, we move part of C out of $B \setminus A$.

The costs for each mapping should have the same physical interpretation as that of EMD with the exception that we allow it to be more costly to move the earth one way than the other. The relative difference of areas represents wholesale creation of new earth. Physically, ν may represent a variety of things, *e.g.* the elements of one set are "heavier" than the other's. The restrictions on nu() are those that allow triangle inequality satisfaction. Thresholding represents the point where it would be cheaper to just create earth at the destination rather than move from the source.

We tried to be as general as possible in the definition of the minimum mapping metric, including non-geometric spaces for which rotation and translation are not defined. We assume that rotation, translation, or any other such preprocessing or normalization is applied prior to the distance.

3.1. Extensions

The following extensions may be used to expand the versatility of the metric without any loss of metric properties. In short, these extensions expand the definition to allow premetrics (defined below), multisets, and asymmetric mappings. These extensions are not mutually exclusive and may be combined if desired.

3.1.1 Premetrics

Define a premetric to be a function $p : X \times X \mapsto \mathbb{R}$ that satisfies only the non-negativity and identity properties of a metric. We interpret $p(a, b)$ and $p(b, a)$ to be, respectively, the distance from a to b and b to a. An example of a premetric is the directed Hausdorff distance from Equation 2. The minimum mapping metric may be modified to accept premetric ground distances by altering the definitions of $\delta_B(\Psi)$ and $\delta_{AB}(\Psi)$:

$$\delta_B(\Psi) = \int_A \min\left[1, \frac{m_D(\Psi(a), a)}{\nu(B)}\right] d\mu(a), \tag{16}$$

$$\delta_{AB}(\Psi) = \int_A \min\left[1, \frac{m_D(a, \Psi(a))}{\nu(A \cup B)}, \frac{m_D(\Psi(a), a)}{\nu(A \cup B)}\right] d\mu(a), \tag{17}$$

or possibly

$$\delta_{AB}(\Psi) = \min\left[\int_A \min\left[1, \frac{m_D(a, \Psi(a))}{\nu(A \cup B)}\right] d\mu(a)\ , \int_A \min\left[1, \frac{m_D(\Psi(a), a)}{\nu(A \cup B)}\right] d\mu(a)\right]. \tag{18}$$

Note that we have simply reversed the arguments of $m_D(a, \Psi(a))$ in a few locations. One may note that any expression for $\delta_{AB}(\Psi)$ is possible so long as it remains less than or equal to δ_A and δ_B.

Note that we cannot relax the identity constraint on the ground distance without also relaxing it for $M(A, B)$. However, if one were able to collapse all identical elements into a single entity via an equivalence relation on $m_D(x, y) = 0$, then one could use the following extension to multisets as a form of relaxed identity. Physically, a premetric may represent the difference between moving earth up versus downhill.

3.1.2 Multisets

A multiset is traditionally a set containing multiple copies of the same element. The multiplicity of an element x is a positive integer indicating how many copies of x are contained in a given multiset X. In the following discussion, we will generalize the definition of a multiset to include any set containing at least one measurable subset with non-unitary membership, where the membership is a positive real number indicating an element's or subset's multiplicity. With this definition, we may also include probability distributions and other continuous functions. In fact, we open the possibility for weighted elements. Let $g(a)$ be the mass density (or multiplicity or distribution) function of the multiset A and $h(b)$ be the density function of the multiset B for $a \in A$ and $b \in B$. For a multiset X with distribution function $D(x)$,

$$\mu(X) = \int_X D(x)\, d\mu(x)\,. \tag{19}$$

Note that for a standard set X (*i.e.* not multiset), $D(x) = 1$ for all $x \in X$. For any element x not contained in X, $D(x) = 0$. The membership of an element $x \in X$ is contingent upon $D(x) > 0$. The density function completely defines a multiset. We will generalize the definition of a subset $A \subseteq B$ in the space X to be such that $g(x) \leq h(x)$ for all $x \in X$. The density function $I(x)$ for the intersection of two multisets $A \cap B$ is defined as

$$I(x) = \min\left[g(x), h(x)\right], \tag{20}$$

and the union $U(x)$ is defined in a similar manner:

$$U(x) = \max\left[g(x), h(x)\right]. \tag{21}$$

Assuming $\mu(A) \leq \mu(B)$, we seek the optimal transport plan or flow that injectively maps the mass of A into B. In particular, let $f(a, b)$ denote the flow of mass from $a \in A$ to $b \in B$, subject to the following constraints:

$$f(a, b) \geq 0, \tag{22}$$

$$\int_B f(a, b)\, d\mu(b) = g(a), \tag{23}$$

$$\int_A f(a, b)\, d\mu(a) \leq h(b), \tag{24}$$

$$\int_A \int_B f(a, b)\, d\mu(b)\, d\mu(a) = \mu(A). \tag{25}$$

Let Φ be an arbitrary flow from A to B. We may now express the δ terms for multisets.

$$\delta_A(\Phi) = \int_A \int_B \Phi(a, b) \min\left[1, \frac{m_D(a, b)}{\nu(A)}\right] d\mu(b)\, d\mu(a), \tag{26}$$

$$\delta_B(\Phi) = \int_A \int_B \Phi(a, b) \min\left[1, \frac{m_D(a, b)}{\nu(B)}\right] d\mu(b)\, d\mu(a), \tag{27}$$

$$\delta_{AB}(\Phi) = \int_A \int_B \Phi(a, b) \min\left[1, \frac{m_D(a, b)}{\nu(A \cup B)}\right] d\mu(b)\, d\mu(a)\,. \tag{28}$$

The optimal flow Φ^* is defined by

$$\Phi^* = \underset{\Phi : A \times B \mapsto \mathbb{R}}{\arg\inf} \left[\delta_A(\Phi) + \delta_B(\Phi) + \delta_{AB}(\Phi)\right]. \tag{29}$$

Note that our mapping Ψ^* may be expressed as a discrete flow where $\Psi^*(a, b) = 1$ if $\Psi^*(a) = b$ and $\Psi^*(a, b) = 0$ otherwise.

3.1.3 Asymmetric Mappings

We may also allow multiple minimum-cost maps to be defined, one for each δ term. In other words, we allow a minimum-cost mapping from A to B and vice versa; each δ term is minimized independently. If we let $M'(A,B)$ denote the metric with an independent mapping for each of δ_A, δ_B, and δ_{AB}, we may note that $M'(A,B) \leq M(A,B)$ (each independent δ term is less than or equal to its dependent counterpart). That is,

$$\delta'_A = \delta_A(\Psi_1^*),\ \delta'_B = \delta_B(\Psi_2^*),\ \delta'_{AB} = \delta_{AB}(\Psi_3^*), \quad (30)$$

where

$$\Psi_1^* = \underset{\Psi:A\mapsto B}{\arg\inf}\,\delta_A(\Psi), \quad (31)$$

$$\Psi_2^* = \underset{\Psi:A\mapsto B}{\arg\inf}\,\delta_B(\Psi), \quad (32)$$

$$\Psi_3^* = \underset{\Psi:A\mapsto B}{\arg\inf}\,\delta_{AB}(\Psi). \quad (33)$$

We may now formally define the multiple minimum mapping metric.

DEFINITION 2. Given a measure space (X,μ) and a semimetric $m_D : X \times X \mapsto \mathbb{R}$, define $\mathcal{F}(X)$ to be the family of all subsets with non-negative, finite measure of X and the multiple minimum mapping metric $M' : \mathcal{F}(X) \times \mathcal{F}(X) \mapsto \mathbb{R}$ to be

$$M'(A,B) = \frac{\mu(A,B) + \delta'_A + \delta'_B - \mu'(A,B)}{\mu(A,B) + \delta'_A + \delta'_B - \delta'_{AB}} \quad (34)$$

where $A, B \in \mathcal{F}(X)$.

The multiple minimum mapping metric would be especially appropriate for use with a premetric.

3.2. Computation

Computation of the minimum mapping metric is primarily limited by computation of the optimal mapping, assuming one exists. Generally, though, one should use the exact same algorithms that one would use for EMD. In the case of a discrete measure, the globally optimal mapping always exists and may be found in $O(n^3)$ time, where n is the cardinality of the larger set. The algorithm given by Edmonds and Karp [7] for the assignment problem can be used to find the optimal mapping between two sets A and B with arbitrary non-negative symmetric costs. Let c be the cost function used to define the separation between elements of A and B, where $c(x,y)$ denotes the cost to map two items x and y. In the context of $M(A,B)$,

$$c(x,y) = \min\left[1, \frac{m_D(x,y)}{\nu(A)}\right] + \min\left[1, \frac{m_D(x,y)}{\nu(B)}\right] + \min\left[1, \frac{m_D(x,y)}{\nu(A\cup B)}\right]. \quad (35)$$

For multisets, one must instead solve the discrete transportation problem, which has an $O(n^3\log(n))$ solution [20]. Since our costs are thresholded, however, it may be faster in practice to use the algorithm of Pele and Werman [21]. In the case of a continuous measure, the solution is less clear and depends upon the choice of m_D. In general, techniques associated with solving the Monge-Kantorovich transportation problem would need to be used [11], including potentially finding numerical solutions to ordinary and partial differential equations. Computation with general continuous measures is beyond the scope of this paper.

The Hausdorff and average distance may be naively computed in $O(n^2)$ time, which offers one advantage in their consideration. The Hausdorff distance may be computed in linear time for various computer vision applications [1] [24].

4. Application

We compared the minimum mapping metric to the Hausdorff distance, EMD, and normalized matching metric for a nearest neighbor search and classification scheme among frames. The k-Nearest Neighbor Graph (kNNG) generation scheme of Dong *et al*. [6] was used due to the relative efficiency of its generation and unbiased nature towards generic metrics. The kNNG search of Hajebi *et al*. [12] was then used to retrieve the nearest neighbor. We would like to point out that these are approximate nearest neighbor algorithms used in the interest of reduced time and space complexity, as one might desire in a practical application. Note that for this application, $X = \mathbb{R}^3$ and μ is the discrete measure. Note also that this test is not designed to show that any one metric is necessarily better than another, but rather that our metric and its variations are competitive alternatives to traditional functions.

4.1. Dataset

The dataset is comprised of 5 static gestures (hand poses) captured for 12 users using a Vicon motion capture camera system and a glove with attached infrared markers on certain joints. A rigid pattern on the back of the glove was used to establish a translation and rotation invariant local coordinate system. The five gestures captured were fist, pointing with one finger, pointing with two fingers, stop (hand flat), and grab (fingers curled) (Figure 1). Several hundred instances of each gesture were captured as part of streams where the user held the gesture for a short time. Instances were preprocessed by removing all markers more than 200 mm from the origin and transforming to local coordinates.

4.2. Methodology

Given that each user's gestures were generally captured as parts of streams of data, a given instance is likely to have a near-duplicate in its user's dataset separated by a small

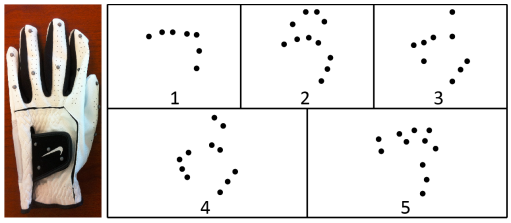

Figure 1. The glove used to capture data along with a sample from each class of static gesture projected onto the local XY plane. The classes are fist (1), stop (2), point with one finger (3), point with two fingers (4), and grab (5).

time-interval. Therefore, we adopted a leave-one out approach where one user was taken as the test set and the kNNG was built upon the remaining 11 users. This procedure thus measures the generalization of the system to new users. A random 5% of the instances from each user (≈ 4000 total instances) were selected during preprocessing and used to build each kNNG and test the classifications. Since random samples are used to initiate graph construction and neighbor search, an arbitrarily set random seed was used to ensure identical choices when possible across the various metrics. The kNNG was built with $k = 12$, and classification of a query frame was determined by a majority vote by its 6 nearest neighbors (with a tie broken by closest average distances). Eight metrics were chosen, including the Hausdorff, EMD, a variant of the average set distance [10], and two variants (scaled Manhattan and Euclidean) of the normalized matching metric of Ramon and Bruynooghe [22]. The average distance chosen was the average minimum distance from each point in one set to the other, based on the given generalization to power means. In a sense, this average distance reflects a minimum cost surjective mapping from each set onto any subset of the other. Variants of the minimum mapping metric (including the normalized matching metric) are denoted by the form $M^3(m_D, \nu)$ with E and M for Euclidean and Manhattan ground distance and C and D for constant and set diameter ν. A constant of 200 mm was chosen so that thresholding of the ground distance could be mostly avoided. One semi-metric ground distance E^2, the square Euclidean distance, was explored.

4.3. Results

Results are presented in two forms, an information coverage plot [13] (Figure 2) and an accuracy scale (Figure 3). An information coverage plot is grounded in information theory and provides an easy to visualize comparison of multi-class classifiers similar to an ROC curve according to certain entropy measures of the confusion matrix. Basically, the horizontal and vertical axes respectively measure the amount of false or true information captured, and the overall score of the classifier is equal to the Manhattan distance from the perfect score $(0, 1)$ (lower scores are better). Considering two classifiers with the same accuracy, the one with false positives spread among fewer classes will achieve a lower score. Visually similar to an ROC curve, the closer the plot is to the top-left, the better our evaluation of the classifier. The diagonal lines correspond to 0.1-interval steps in score.

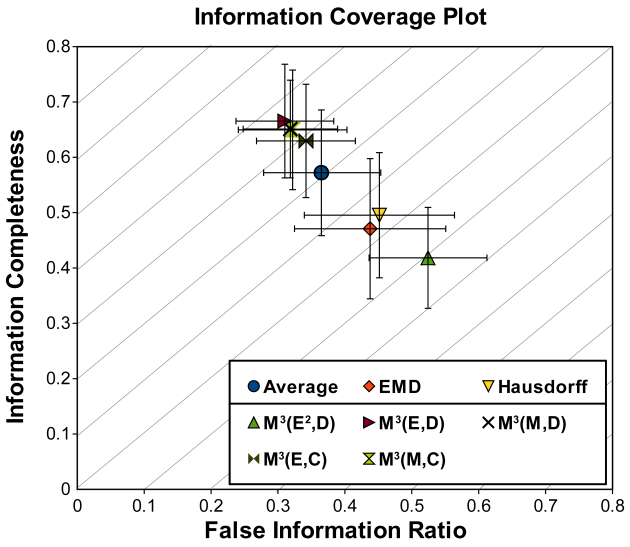

Figure 2. An information coverage plot for comparing the performance of each metric. Similar to an ROC curve, the top-left of the plot is preferred for a good evaluation. Variants of our metric, denoted by M^3, performed better than the alternatives.

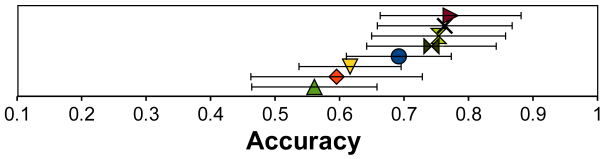

Figure 3. The average accuracy of each metric with standard deviations. Refer to the legend in Figure 2.

As can be seen, variants of the minimum mapping metric performed better than the alternatives (with one exception). The average distance performed quite well, the closest in performance to the minimum mapping metric. We may also note that the square Euclidean semi-metric, for which there was no particularly compelling reason to use with this data in the first place, performed relatively poorly compared to the remaining metrics. Whether this indicates that a metric ground distance is preferable to semi-metric in the general case is unknown. Hausdorff and EMD performed quite similarly, likely due to the removal of most outliers during preprocessing. The minimum mapping metrics with set diameter marginally outperformed their constant alternatives (Ramon-Bruynooghe metrics). The diameters of most experimental sets should not have varied by more than a factor

of two or so, possibly accounting for the minimal improvement. Together, these metrics significantly outperformed the traditional EMD and Hausdorff. In the absence of partial matches, however, we would not expect much of a difference from EMD. Increasing k for a more strongly connected kNNG would likely result in improvements for all metrics.

5. Conclusions and Future Work

We have introduced a new parameterizable metric that measures the distance between unordered sets of different sizes with non-negative, finite measure. We also presented extensions of the metric that allow the use of premetrics, multisets, and multiple minimum-cost mappings from each set's perspective. Comparison to existing methods demonstrated certain cases of our metric to be competitive. The metric space induced by the minimum mapping metric has not been completely characterized with respect to topological properties such as completeness. We also have reason to believe that the current constraints on ν and its usage do not completely characterize the possible functions that could be employed while retaining metric properties. Though our function addresses the discrete comparison of elements present in the classical Jaccard distance, it does not necessarily address binary set membership. An extension to fuzzy sets could likely be performed through some manipulation of the ground distance and/or use of multisets, but a more rigorously researched method may be worth investigating.

Acknowledgments

This research is supported in part by the AFOSR grants FA9550-09-1-0479 and FA9550-09-1-0289.

References

[1] M. J. Atallah. A linear time algorithm for the hausdorff distance between convex polygons. *Information Processing Letters*, 17(4):207 – 209, 1983.

[2] R. K. Brouwer. Extending the rand, adjusted rand and jaccard indices to fuzzy partitions. *Journal of Intelligent Information Systems*, 32(3):213–235, June 2009.

[3] Y. R. Chao. A note on "continuous mathematical induction". *Bulletin of the American Mathematical Society*, 26(1):17–18, 1919.

[4] B. Chaudhur and A. Rosenfeld. On a metric distance between fuzzy sets. *Pattern Recognition Letters*, 17(11):1157 – 1160, 1996.

[5] M. M. Deza and E. Deza. *Encyclopedia of Distances*. Springer, 1 edition, Aug. 2009.

[6] W. Dong, C. Moses, and K. Li. Efficient k-nearest neighbor graph construction for generic similarity measures. In *Proceedings of the 20th international conference on World Wide Web*, WWW '11, pages 577–586, New York, NY, USA, 2011. ACM.

[7] J. Edmonds and R. M. Karp. Theoretical improvements in algorithmic efficiency for network flow problems. *Journal of the Association for Computing Machinery*, 19(2):248–264, 1972.

[8] T. Eiter and H. Mannila. Distance measures for point sets and their computation. *Acta Informatica*, 34:103–133, 1997.

[9] A. Figalli and N. Gigli. A new transportation distance between non-negative measures, with applications to gradients flows with dirichlet boundary conditions. *Journal de Mathmatiques Pures et Appliques*, 94(2):107 – 130, 2010.

[10] O. Fujita. Metrics based on average distance between sets. *Japan Journal of Industrial and Applied Mathematics*, 30(1):1–19, 2013.

[11] W. Gangbo. An introduction to the mass transportation theory and its applications. *Gangbos notes from lectures given at the 2004 Summer Institute at Carnegie Mellon University and at IMA in March 2005*, March 2005.

[12] K. Hajebi, Y. Abbasi-Yadkori, H. Shahbazi, and H. Zhang. Fast approximate nearest-neighbor search with k-nearest neighbor graph. In *Proceedings of the 22nd international joint conference on Artificial Intelligence - Volume Two*, IJCAI'11, pages 1312–1317. AAAI Press, 2011.

[13] R. S. Holt, P. A. Mastromarino, E. K. Kao, and M. B. Hurley. Information theoretic approach for performance evaluation of multi-class assignment systems. In *Proceedings of SPIE*, volume 7697, 2010.

[14] D. Huttenlocher, G. Klanderman, and W. Rucklidge. Comparing images using the hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9):850–863, 1993.

[15] L. V. Kantorovich. On the translocation of masses. *C. R. (Doklady) Acad. Sci. USSR*, 321:199–201, 1942.

[16] L. V. Kantorovich. On a problem of monge. *Uspekhi Matematicheskikh Nauk*, 3(2):225–226, 1948.

[17] E. Levina and P. Bickel. The earth mover's distance is the mallows distance: some insights from statistics. In *Proceedings of the Eighth IEEE International Conference on Computer Vision, 2001.*, volume 2, pages 251–256 vol.2, 2001.

[18] A. H. Lipkus. A proof of the triangle inequality for the Tanimoto distance. *Journal of Mathematical Chemistry*, 26:263–265, 1999.

[19] G. Monge. Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences*, pages 666–704, 1781.

[20] J. B. Orlin. A faster strongly polynomial minimum cost flow algorithm. In *Operations Research*, pages 377–387, 1988.

[21] O. Pele and M. Werman. Fast and robust earth mover's distances. In *Proceedings of the IEEE 12th International Conference on Computer Vision*, pages 460–467, 2009.

[22] J. Ramon and M. Bruynooghe. A polynomial time computable metric between point sets. *Acta Informatica*, 37(10):765–780, 2001.

[23] Y. Rubner, C. Tomasi, and L. J. Guibas. The earth movers distance as a metric for image retrieval. *International Journal of Computer Vision*, 40:2000, 2000.

[24] R. Shonkwiler. An image algorithm for computing the hausdorff distance efficiently in linear time. *Information Processing Letters*, 30(2):87 – 89, 1989.