Metric Space in Information Transmission

BY

Tsok Samuel Hwere*, Kpanja Sunday Shammah† and Okolo Patrick Noah†
*Department of Mathematics/Statistics
Nasarawa State Polytechnic, Lafia

†Department of Natural Sciences
Nasarawa State Polytechnic, Lafia

ABSTRACT

In this paper, we succinctly presented that information transmission satisfies the properties of metric space. The space of information source is infact a metric space to which a geometric picture can be ascribed. Much more in error-control codes, the measure of comparing two codes words of equal length is the hamming distance – is obviously a metric. To achieve this, we viewed the forgoing under information metric and hamming metric.

Keywords: Metric space, Information transmission, Information source, Information geometry, Information metric, Hamming metric, Error-detecting and Error- correcting Codes.

1.0 INTRODUCTION

The ratio of 'metric' is a generalization of the Euclidean metric arising from the four long-known properties of Euclidean distance. The Euclidean metric defines the distance between the lines connecting them, with this idea; we

introduce a measure of distance between information sources. It demonstrates that the space of information sources has much topological structure which here for has not been utilized directly in application of information theory. The space of information source is, in fact a metric space to which a geometric picture can be ascribed, which we called information metric [1]. The information metric quantifies the degree of recording equivalence. And also, it provides some insight into the nature of information itself. In error control codes, the measure of comparing two code wards of equal length is the hamming distance which simply is the number of places, where they differ. This is obviously a metric and we therefore called it the hamming metric.

The main objects of coding theory are metric vector or matrix spaces. Subsets of spaces are known as codes. The main problem is constructing codes of given pairwise distance and having maximal cardinality. Most known and most investigated spaces Hamming spaces. Quite a good number of papers and Books have discussed the Hamming metrics, we mention here ([2],[3],[5],[6], [7]). However this work did not consider metric in details.

A motivation of this work was the conviction that an understanding of the topological structure of the metric lattice is necessary for developing a quantitative measure and that of the Hamming metric is necessary for the developing of error-control codes for better transmission of information.

The structure of this paper is as follows: we first defined some terms used in the paper and we went further to discuss the information metric and hamming metric.

Definition 1.1:

A metric space is an ordered pair (M, d) where M is a set (which some authors require being non empty) and d is a metric on M, which is a function.

$d: M\ X\ M \rightarrow \mathcal{R}$

such that for any x, y and z in M

1. d(x,y) ≥ 0 (non – negativity)
2. d(x,y) = 0 if and only if x = y (identity of indiscernible)
3. d(x,y) = d(y,x) (symmetry)
4. d(x,y) $\leq$ d(x,y) + d(y,z) (triangular inequality)

Definition 1.2:

The hamming weight W(c) of a code word c is the number of non zero components in the code word.

Definition 1.3

The hamming distance between two code words d(x,y), is the number of places in which the code words x and y differ.

Definition 1.4

The minimum (Hamming) distance of a code c is the minimum distance between any two code words in the code.

d(c) = min {d(x,y) / x $\neq$ y, x, y ϵ c}

2.0 INFORMATION METRIC

We introduce here the idea of a measure of distance between information sources; this will demonstrate that the space of information sources has much topological structure which therefore has not been utilized directly in application of information theory or in the study of complex systems.

The space of information is in fact, a metric space to which a geometric picture can be ascribed [1].

2.1 INFORMATION GEOMETRY

Conventional development of information theory measure the complexity of a source in terms of the entropy. Entropy is a function on the space of sources I that yields real numbers, $H(s): T \to \mathcal{R}$,

With the source characterized by its probability measure. Indeed, the quantity of information, the entropy, is a measure on the space of information sources, in the same sense that probability is a measure on event space [4].

The entropy of source $X \epsilon I$ quantifies the size or volume of the equivalence class. The following establishes entropy as a measure and sets up a partially "geometric" picture.

The starting point of the development is the following four definitions:

1. The origin $\emptyset$ is the measurement set that is predictable: $H(\emptyset) = 0$
2. The norm $\|X\|$ of a source X is its entropy H(X).
3. The addition of two sources is the union of measurements:
   X + Y = {all events in either X or Y}
4. The product of two sources is the intersection of their measurement
   $X \cdot Y$ = {all those events common to X and Y}.

These operators yield an algebra of measurements. The first step is to establish that the entropy is a measure.

1. $0 \leq H(X) \leq \infty$ for every source X and $\exists$ X_o such that $H(X_o) < \infty$.
2. $H(X + Y) = H(X) + H(Y)$, whenever $X \cdot Y = \emptyset$; that is, the sources are independent.

It also follows from these that:

1. If $X \rightarrow Y$, then $H(Y) \leq H(X)$;
2. $H(\emptyset) = 0$; and
3. $H(X + Y) \leq H(X) + H(Y)$ for any sources X and Y.

To determine the distance d(X,Y) between two sources X and Y requires a measure of their difference $X \Delta Y$, where Δ is the symmetric difference in the set theoretic sense. $X \Delta Y$ is itself an information source and so formally we define $d(X,Y) = \|X \Delta Y\|$ (1)

This yields a generalized picture of the norm of a source as being the "distance" from the origin of predictable, zero – entropy sources, since

$\|X\| = \|X \Delta \emptyset\| = d(X, \emptyset)$ (2)

The information source $X \Delta Y$ can be explain as below

In set theoretic terms there are two constituent, independent sources:

(i) X – Y are those events in X and not in Y and

(ii) Y – X corresponds to those in Y and not in X. The entropy H(X - Y) of the source measurements in X - Y defines a conditional entropy H(X|Y) for measurements x_i of X given y_j of Y, such that x_i is not determined from y_j with probability one.

We define a source Z that is the union of these two sources,

$Z \equiv X \Delta Y = Z_1 + Z_2 = (X - Y) + (Y - X)$. (3)

From the algebra of measurements it follows that

$Z = (X + Y) \cdot \overline{X \cdot Y} = X \cdot \overline{X \cdot Y} + Y \cdot \overline{X.Y}$ (4)

In informational terms, we have

$H(X - Y) = H(X \cdot \overline{X \cdot Y})$ *and* $H(Y - X) = H(Y \cdot \overline{X \cdot Y})$ (5)

A measure of the size or entropy of Z will be a measure of the non-common or distance between x and y.

$H(Z) = H(Z_1 + Z_2) = H(Z_1) + H(Z_2/Z_1) = H(Z_1)$ (6)

The finaly step follows from the independence of Z_1 and Z_2.

$$Z_1 \cdot Z_2 = (X \cdot \overline{X \cdot Y}) \cdot (Y \cdot \overline{X \cdot Y}) = (X \cdot Y) \cdot \overline{X \cdot Y} = \Phi \quad (7)$$

We can further obtain

$$H(Z) = H(X \cdot \overline{X \cdot Y}) + H(Y \cdot \overline{X \cdot Y}) = H(X/Y) + H(Y/X) \quad (8)$$

With this we have established, starting with the entropy as a norm of information source that algebra of measurement allows us to define the conditional source x-y and y-x. From this it readily follows that

$$\| X - Y \| = d(X-Y, \Phi)$$

The associate pseudo-geometric picture is shown in fig 1.

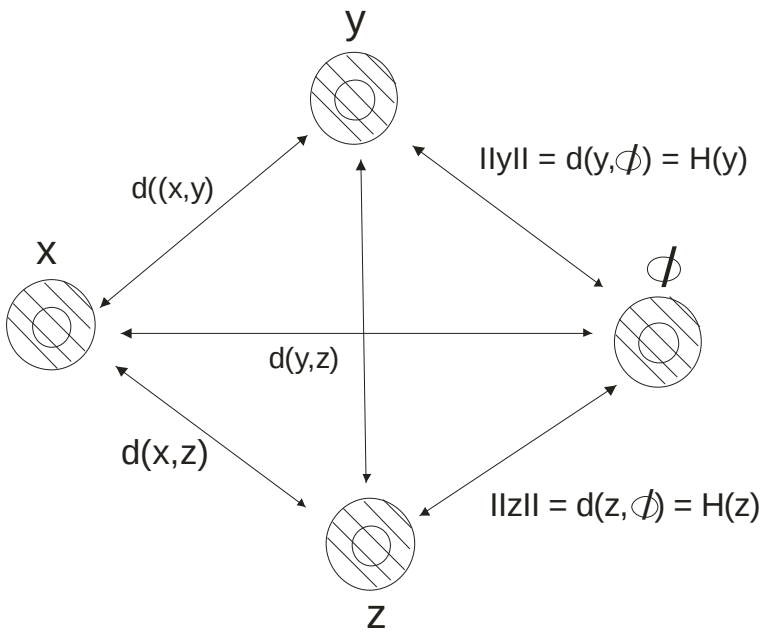

Fig 1. The geometric interpretation of space information sources. Shown are the distances and information vectors for three sources.

2.2 INFORMATION METRIC

The information quantity $d(x,y) = h(x/y) + H(y/x)$

Can be interpreted as the total independent information. We now establish its metric properties. Although the proofs are straight forward.

THEOREM 1

D is a metric and (I,d) is a metric space.

Proof

i. Symmetry: $d(x,y) = d(y,x)$ this follows directly from the symmetry of definition

ii. Equivalence: $d(x,y) = o$ if and only if $x \sim y$

iii. "Only if": assume $d(x,y) = 0$ As the conditional entropies themselves are positive or zero and their sum is zero, they individually vanish. Consider one of the zero conditional entropies $H(x/y) = 0$

The measurements of x knowing those of y provide no new information and so may be inferred with probability one from y. thus, there is a recording, that may be many-to-one, of measurements of y into x measurements. Similarly, since H(y/x) vanishes, there is a recording of x measurements into those from source y. Taking these together, there is a one-to-one coding between measurements from x and from y and so they are equivalent sources: x~y, [1].

For the "if" portion: assume x~y, then there is a one-to-one recording between measurements from x and from y. measurements of source x can be deduced with probability one from those of y and vise-versa. It follows from this that the conditional entropies vanish, as does the distance between them.

2.3 TRIANGULAR INEQUALITY

$d(x,z) \leq d(x,y) + d(y,z)$ we consider expressions of the three variables joint entropy.

$H(x,z) \leq H(x,y,z)$ or

$$H(x,z) \leq H(x/yz) + H(y,z) \quad (9)$$

Noting that additional measurements can not increase the entropy, i.e.

$H(x/y) \geq H(x/yz)$, we have

$H(x,z) \leq H(x/y) + (y,z)$ or

$$H(x,z) \leq H(x,y) - H(y) + H(y,z) \quad (10)$$

Subtracting the average independent entropy

$\frac{1}{2}(H(x) + H(y))$ yields

$H(x,z) - \frac{1}{2} H(x) - \frac{1}{2} H(z) \leq H(x,y) - \frac{1}{2} H(y) - \frac{1}{2} H(x) + H(y,z) - 1/2\, H(z) - \frac{1}{2} H(y)$

Or

$$H(x/z) + H(z/x) \leq H(x/y) + H(y/x) + H(y/z) + H(z/y)$$

Or

$$d(x,z) \leq d(x,y) + d(y,z) \qquad (11)$$

thus, d(-,-) is a metric. With this it follows that the pair (I,d), where I is the space of recording – equivalent information sources, is a metric space. This completes the proof.

The theorem indicates that the space of information has quite a bit of topological structure. For example, the notion of ϵ- balls of "close" information sources. The continuity of function on information sources can be developed.

These are numerical computations of information distances will follow in a sequel.

We can define a normalized metric as follows:

$$d(x,y) = \frac{H(x/y) + H(y/x)}{H(x,y)} \qquad (12)$$

Note that in the case of independent sources $d = 1$

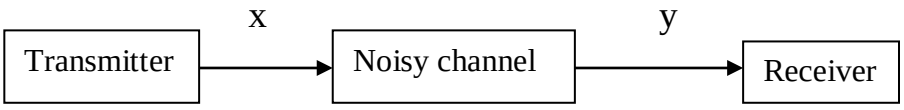

Fig 2. Discrete channel

A simple model of the process is shown in fig 2. Where x represents the space of messages transmitted and y the space of messages received during a unit time over the channel.

3.0 HAMMING METRIC

We shall here work with an alphabet, which is simply a finite set A of symbols. If A has size q, then we call A a q-ary alphabet (although we say binary and denary rather than 2-ary and 3-ary.

For most purposes, it is sufficient to take A to be F_q, the field of order q. A word of length n is simply a string consisting of n (not necessarily distinct elements of A, i.e. an element of A, and a block code of length is simply a set of words of length n, i.e. a sub set of A^n.

If A is a q-ary alphabet, we say that any code over A is a q-ary code. Although there are words which have different lengths, but our entity is with the block code.

3.1 ERROR DETECTION AND ERROR CORRECTION

DIFINITION 1.41

Let $A = \{a_1, a_2, \ldots, a_r\}$ be a finite set, which we call a code alphabet.

An r-ary code over A is a sub set C of the set A of all words over A. the elements of C are called code words. The number r is called the radix of the code.

When we talk about the number of errors in a received code word, we are talking about the distance between a received word and a transmitted word. Suppose a codeword $x = x_0x_1,\ldots\ldots x_{n-1}$ is sent through a channel and the received vector is sent through a channel and the received vector is $y=y_0y_1\ldots..y_{n-1}$. The error vector is defined as $e = y\text{-}x = e_0e_1\ldots\ldots e_{n-1}$,[6].

The job of the decoder is to decide which code word was most likely transmitted, or equivalently decide which error vector most likely occurred. Many codes use a nearest neighbor decoding scheme which chooses the code word that minimizes the distance between the received vector and possible transmitted vector. A nearest neighbor decoding scheme for a q-ary

code maximizes the decoder's likelihood of correcting errors provided the following assumptions are made about the channel.

1. Each symbol transmitted has the probability $p(<1/2)$ of being received in error.
2. If a symbol is received in error, that each of the q-1 possible is equally likely.

Such a channel is called a q-ary symmetric channel, throughout we shall assume the channels involves as symmetric.

The steps of the encoding and decoding that concern us are as follows:

$M \rightarrow \text{Encode} \rightarrow \overline{C} \rightarrow \text{Noise} \rightarrow C + e = r \rightarrow \text{Decode} \rightarrow M$

Where m is the message, c is the codeword, e is the error vector due to noise, r is the received word or vector.

The hope is that $\widetilde{m} = m$

Informally a code is t- error detecting if, whenever we take a codeword and change at most t of the symbols in it, we don't reach a different code word. So if we send the new word to someone without telling him which symbols we changed, he will be able to tell us whether we changed any symbols.

A code is t-error connecting if whenever we take a code word and change at most t of the symbols in it ,we don't reach a different code word, and we don't even reach word which can be obtained from different starting code word by changing at most t of the symbols. So if we sent the new word to someone without telling him which symbols we changed, he will be able to tell us which code word we started with.

Formally we define a metric (distance function on A^n as follows: given two words x and y differ, i.e.

if $x = x_1 \ldots\ldots\ldots x_n$ and $y = y_1 \ldots\ldots\ldots y_n$, then

d(x,y) is the number of values I for which is called the hamming distance d.

LEMMA 3.41

D is a metric on A^n

1. $d(x,y) = 0$ for all $x \in A^n$
2. $d(x,y) > 0$ for all $x \neq A^n$
3. $d(x,y) = d(y,x)$ for all $x, y \in A^n$
4. $d(x,z) \leq d(x,y) + d(y,z)$ for all $x, y, z \in A^n$

Proof

(1), (2) and (3) are very easy, so let us do (4)

Now d(x,z) is the number of values I for which $xi \neq zi$. Note that if $xi \neq zi$, then either $xi \neq yi$ or $yi \neq zi$. Hence

$\{i \mid xi \neq zi\} \subseteq \{i \mid xi \neq yi\} \cup \{i \mid yi \neq zi\}$;

So $|\{i \mid xi \neq z\}| \leq |\{i \mid xi \neq yi\} \cup \{i \mid yi \neq zi\}$

$\leq |\{i \mid xi \neq yi\}| + |\{i \mid yi \neq zi\}|$

$d(x,z) \leq d(x,y) + d(y,z)$

now we can talk about error detection and correction. We say that a code c is t-error detecting if $d(x,y) > t$ for any two distinct words in C. we say that c is t-error correcting if there do not exist words $x,y \in c$ and $z \in A^n$ such that $d(x,y) \leq t$ and $d(y,z) \leq t$. Therefore the function $d: AxA \rightarrow \Re$ is the hamming distance which satisfies all the property of metric space, hence the pair (A_2^n, d) is known as the Hamming metric.

Sphere: For $a \in A_2^n$ and positive integer k

$S(a, k) = \{x \in A_2^n \mid d(a, k) \leq k\}$ is called sphere radius with centre a.

For a = 011 hence S(a,2) = {000,100,010,110,101,011,111}

EXAMPLE

The simplest kinds of error detecting codes are repetition code of length n over A simple consists of all words aa……a, for a ϵ A. For this code, any two distinct code words differ in every position, and so d(x,y) = n for all $x \neq y$ in c.

So the code is t-error detecting for every $t \leq n-1$, and is t-error correcting for every $t \leq \frac{n-1}{2}$.

We solve Some examples of minimum distance code words (minimum E or min E = δ) base on [3].

Example 1. If $E: A_2^2 \rightarrow A_2^6$ with min E is 5 then How many

(1) errors can be detected

(2) errors can be corrected

E(00)=000000 E(11)=111111 E(10)=101010 E(01)=010101

Solution1

δ=7

(1) errors of weight≤ δ-1=5-1=4 be detected

(2) errors of weight≤(δ-1)/2=2 can be corrected

Example 2. If E: $A_2^2 \rightarrow A_2^6$ Find (1) Min E then

(2) errors detecting capacity

(3) errors correcting capacity

Where E(00)=000000 E(11) =111111 E(10)=101010 E(01)=010101

Solution 2

d(E(00),E(11))= 6 d(E(01),E(11))=3

d(E(00),E(01))=3 d(E(00),E(10))=3

d(E(01),E(10))=6 d(E(10),E(11))=3

(1) min E= δ is 3

(2) error detecting capacity: E can detect errors with weight ≤ δ-1=3-1=2

(3) errors correcting capacity: E can correct errors of weight≤(δ-1)/2=1

Example 3. If E: $A_2^3 \to A_2^6$ Find (1) min E then

(2) error detecting capacity

(3) errors correcting capacity

Where E(000)=000111 E(001)=001001 E(010)=010010 E(100)=100100

E(110)=110001 E(101)=101010 E(011)=011100 E(111)=111000

Solution 3

First Distance are found

d(E(000),E(001))=3 d(E(100),E(010))=4

d(E(000),E(010))=3 d(E(100),E(001))=4

d(E(000),E(100))=3 d(E(100),E(110))=3

d(E(000),E(110))=4 d(E(100),E(101))=3

d(E(000),E(101))=4 d(E(100),E(011))=3

d(E(000),E(011))=4 d(E(100),E(111))=3

d(E(000),E(111))=6

d(E(010),E(001))=4 d(E(001),E(110))=3

d(E(010),E(110))=3 d(E(001),E(101))=3

d(E(010),E(101))=3 d(E(001),E(011))=2

d(E(010),E(011))=3 d(E(001),E(111))=2

d(E(010),E(111))=3

d(E(110),E(101))=4 d(E(101),E(011))=3

d(E(110),E(011))=4 d(E(101),E(111))=2

d(E(110),E(111))=2 d(E(011),E(111))=2

Min E=2=δ

(2) Error detecting capacity: E can detect errors with weight ≤ δ-1=2-1=1

(3) Error correcting capacity: E can correct errors of weight$\leq (\delta-1)/2=1/2$ can be corrected. This means E cannot correct any error.

LEMMA 3.2

A code C is t-error detecting if and only $d(C) \geq t + 1$, and is t-error-correcting if and only if $d(C) \geq 2t + 1$

Proof

The first fact is immediate from the definition of "t-error-detecting".

For the second part, assume that C is not t-error-correcting. Then there exist distinct codeword x,y, ϵ C and a word $z \epsilon A^n$ such that

$$d(x,z) \leq t \text{ and } d(z,y) \leq t.$$

By triangle inequality $d(x,y) \leq d(x,z) + d(z,y) \leq 2t$ and have $d(C) \leq 2t$. contradicting the assumption that $d(C) \geq 2t +1$.

Conversely, suppose that C can correct up to t errors. If $d(C) \leq 2t$ then there are two codewords that differ in 2t bits. Changing k of the bits in one of these codewords produces a bit string that differs from each of these two codewords in exactly t positions, thus making it impossible to correct these t errors.

LEMMA 3.3

A code c is t-error-correcting if and only if for any distinct words x, y ϵ c, the spheres S(x,t) and S(y,t) are disjoint.

This lemma gives us a useful bound on the size of a t-error-correcting code. We begin by counting the words in a sphere.

Hence, the binomial coefficient

$$\binom{n}{r} = \frac{n!}{(n-r)!r!}$$

LEMMA 3.4

If A is a q-ary alphabet, x is a word over A of length n an $r \leq n$, then the sphere S(x,r) contained exactly.

$$\binom{n}{0}+(q-1)\binom{n}{1}+(q-1)^2\binom{n}{2}+\cdots+(q-1)^r\binom{n}{r}$$

Words.

Proof

We claim that for any i, the number of words y such that d(x,y) equals I is $(q-1)^i\binom{n}{I}$; the lemma then follows by summing for i = 0, I, …, r.

d(x,y) = I means that x and y differ in exactly I position. Given x, in how many ways can we choose such a y ? we begin by choosing the position in which x and y differ, this can be done in $\binom{n}{I}$ ways.

Then we choose what symbols will appear in these positions in y. for each position we can choose any symbol other than the symbol which appears in that position in x – this gives us q-1 choices. So we have $(q-1)^i$ choices for these i symbols altogether.

THEOREM 3.1

Hamming bound. If c is a t-error-correcting code of length n over a q-ary alphabet A, then

$$|C| \leq \frac{(q)^n}{\binom{n}{I}+(q-1)\binom{n}{I}+(q-1)^2\binom{n}{2}+\cdots+(q-1)^t\binom{n}{t}}$$

Proof

Each code word has a sphere of radius t around it, and by lemma 2.4. This sphere is disjoint. So the total number of words in all these spheres together is

$$M \times (\binom{n}{0}+(q-1)\binom{n}{1}+(q-1)^2\binom{n}{2}+\cdots+(q-1)^t\binom{n}{t})$$

And this can't be bigger than the total number of possible words, which q^n. The hamming bound is also known as the sphere-packing bound.

4.0 CONCLUSION

We have successfully shown here that information transmission can be conveniently be discussed in terms of information metric, thus A measure of distance between information source is derived from algebra of measurement of which the space of information sources is shown to be a metric space; and the hamming metric, thus the typical measure of how "close" two (equal-length) strings are, is the hamming distance which is simply the number of places where they differ. We can conclude that metric space is a very useful tool in the field of Information theory and as well as many other fields of studies.

References

[1] J.P. Crutchfield, *Information and its Metric*, Nonlinear structures in physical systems-pattern formation, Chaos and Waves, Springer-Verlag, New York. pp,119-130, (1990)

[2] D. Joyner, R. Kreminski, and J.Turisco, *Applied Abstract Algebra*, The John Hopkins University press pp,179-180 , (2003)

[3] J.Thomas, A. Stephen, *Abstract Algebra theory and application*, GNU free Documentation license Version 1.2 pp,121-125, (2012) retrieved from (abstract.ups.edu)

[4] R. S. Ingarden, and K. Urbanik, "Information without probability", Colloq. Math., vol. IX, p,131, (1962)

[5] S. G. Vladul, D. J. Nogin, and M. A. Tsfasman, "Algebraic- Geometric codes", Fundamentals (in Russian) M:MCCME, p,504, (2003)

[6] S. A. Sarah, *Introduction to Algebraic Coding theory*, Cornell University Department of Mathematics pp, 4-7 ,(2002)

[7] I. Kra, *Abstract Algebra with Application*, State University of New York at Stony Brook and University of California at Berkeley pp,132-134, (2003)