Short Communication

Space Science Journal

New Distance for any Finite Sets, Half the Hamming Distance

J. K. Abdurakhmanov*

Senior Lecturer, Department of Information Technology candidate of physical and mathematical sciences (i.e. PhD) Andijan State University UZBEKISTAN

***Corresponding Author**
J. K. Abdurakhmanov, Senior Lecturer, Department of Information Technology candidate of physical and mathematical sciences (i.e. PhD) Andijan State University UZBEKISTAN

Submitted: 2024, Feb 02; **Accepted:** 2024, Feb 26; **Published:** 2024, Mar 15

Citation: Abdurakhmanov, J. K. (2024). New Distance for any Finite Sets, Half the Hamming Distance. *Space Sci J, 1*(1), 1-3.

Abstract

In this paper, we introduce a new, previously unknown, distance (i.e., a new metric) in a set whose elements are some other (any) finite sets. It is proved that with such a metric the set under consideration is a metric space. A direct relationship is established between this distance and the Hamming distance: it is exactly two times smaller than the Hamming distance and it is much easier to calculate it. As an application, the set of natural numbers is considered as a discrete metric space with a new metric introduced, and a new metric criterion for the primality of a natural number is established. This is the first metric criterion in the history of mathematics for a natural number to be prime.

Keywords: Set, Finite Set, Discrete Set, Distance, Metric, Hamming Distance, Metric Space, Discrete Metric Space, Number Theory.

1. Introduction

For any finite set **A**, through |**A**| we denote the number of elements of this set, that is, in the language of set theory – the cardinality of this set. For example, if **A** = {**4, a, 7, b**}, then |**A**| = **4**.

Now let **T** be an arbitrary (finite or infinite) set, each element of which is a finite set. In other words, **T** is the set of some finite sets. Here the word "some" has a very broad meaning: the set **T** can consist of an infinite number of such "some" (that is, arbitrary, any) finite sets. These finite sets contained in **T** can be very different: heterogeneous or homogeneous. Despite this, in the most ordinary, generally accepted sense, we will use the operations of union and intersection of these finite sets, and these unions and intersections do not have to be elements of the set **T**.

For any elements **α∈T** and **β∈T**, we introduce the distance **ρ (α, β)** between them by the following formula:

$$\rho(\alpha, \beta) = ((|\alpha| + |\beta|))/2 - |\alpha \cap \beta|, \qquad (1)$$

here **α∩β** is the intersection of the subsets **α** and **β**.

For example, if **T** contains elements **α = {1, 2, 3}** and **β = {3, 4, 5, 6},** then using formula (1) it is easy to calculate the distance between them. Because

| α | = 3, | β | = 4 and **| α∩β | = 1,** then **ρ (α, β) = ((3 + 4))/2 – 1 = 3.5 – 1 = 2.5.**

We now prove the following theorem.

Theorem 1. The set T with distance (1) is a metric space.

Proof. For any elements **α∈T** and **β∈T**, the first axiom of the metric space:

1) if **α ≠ β**, then **ρ (α, β)> 0**, and if **α = β,** then **ρ (α, β)** = 0 follows directly from equality (1). Really, if **α ≠ β**, then each of the sets α and β does not have fewer elements than **α∩β** and at least one of them has more elements than **α∩β**, and therefore **| α | + | β | > 2 | α∩β |.** This and (1) imply **ρ (α, β)> 0**. If **α = β,** then **| α | + | β | = 2 | α∩β |** and therefore from (1) it follows that **ρ (α, β) = 0.** Since **α∩β = β∩α** holds for any sets **α** and **β**, the second axiom of the metric space is:

2) **ρ (α, β) = ρ (β, α)** also directly follows from equality (1).

It remains to verify the last axiom – the triangle axiom of the metric space, namely the following axiom: for any elements **α**, **β**, and **γ** of the set T, the following inequality holds:

3) **ρ (α, β) ≤ ρ (α, γ) + ρ (γ, β).** **(2)**

Substituting the values of the distances by formula (1) into inequality (2), we obtain the equivalent inequality

((| α | + | β |))/2– | α∩β | ≤ ((| α | + | γ |))/2– | α∩γ | + ((| γ | + | β |))/2– | γ∩β |,

which, after simple elementary (and, most importantly, equivalent!)

transformations, is reduced to the form:

$$|\gamma| \geq |\gamma \cap \alpha| + |\gamma \cap \beta| - |\alpha \cap \beta|. \qquad (3)$$

Therefore, it suffices to prove this inequality.

Obviously, the set **(γ∩α) (γ∩β)** is a subset of the set **γ**; therefore, the inequality

$$|\gamma| \geq |(\gamma \cap \alpha) \quad (\gamma \cap \beta)|. \qquad (4)$$

For any finite sets **A** and **B**, the equality

| A B | = | A | + | B | – | A∩B |.

If we put **A** = **(γ∩α)** and **B = (γ∩β)** in this equality, then we get:

$|(\gamma\cap\alpha)\ \ (\gamma\cap\beta)| = |\gamma\cap\alpha| + |\gamma\cap\beta| - |(\gamma\cap\alpha)\cap(\gamma\cap\beta)|$ or, which is the same

$|(\gamma\cap\alpha)\ \ (\gamma\cap\beta)| = |\gamma\cap\alpha| + |\gamma\cap\beta| - |\gamma\cap(\alpha\cap\beta)|$.

From this equality, by virtue of $|\gamma\cap(\alpha\cap\beta)| \le |\alpha\cap\beta|$ the inequality follows:

Further, inequality **(3)** follows from **(4)** and **(5),** which, as we showed above, is equivalent to the triangle inequality **(2)**. Thus, for a set **T** with distance **(1)**, all the axioms of the metric space hold.

The theorem is proved.

Consider an arbitrary finite set **X** consisting of **n** elements. Let these elements be numbered, i.e.

$$X = \{x_1, x_2, x_3, \ldots, x_n\}.$$

Now let **T** be the set of some subsets of the set **X**. Then, according to **Theorem 1** just proved, the set **T** is a metric space with metric **(1).** For any $\alpha \in T$ and $\beta \in T$, we define binary vectors of length **n:**

$$\overline{\alpha} = (a_1, a_2, \ldots, a_n),\ \overline{\beta} = (b_1, b_2, \ldots, b_n),$$

where

$$a_i = \begin{cases} 1, \text{if } x_i \in \alpha, \\ 0, \text{if } x_i \notin \alpha; \end{cases}$$

$$b_i = \begin{cases} 1, \text{if } x_i \in \beta, \\ 0, \text{if } x_i \notin \beta; \end{cases}$$

$i = 1, 2, 3, \ldots, n$.

As is known, the Hamming distance $h(\overline{\alpha},\overline{\beta})$ (see [1, p.39]) between these vectors is the number of their coordinates that differ in value.

It turns out that the new distance **(1)** we introduced between the sets is exactly two times less than the Hamming distance between the corresponding binary vectors.

Theorem 2. The equality

$$h(\overline{\alpha},\overline{\beta}) = 2\rho(\alpha,\beta). \qquad (6)$$

Proof. From the above notation it follows that:
1) the number of ones in the binary vector $\overline{\alpha}$ is equal to $|\alpha|$,
2) the number of ones in the binary vector $\overline{\beta}$ is equal to $|\beta|$,
3) the number of common coordinates, where both vectors $\overline{\alpha}$ and $\overline{\beta}$ have ones, is equal to $|\alpha\cap\beta|$.

Therefore, the number of different coordinates of these vectors (i.e. the Hamming distance between these vectors) is equal to

$$h(\alpha,\beta)=(|\alpha|-|\alpha\cap\beta|)+(|\beta|-|\alpha\cap\beta|)\ .$$

This and equality **(1)** imply the assertion (equality) of the theorem.

The theorem is proved.

Thus, the new distance **(1)** introduced by us is exactly two times less than the corresponding Hamming distance in the case when the set **T** is the set of subsets of some finite set. Therefore, to calculate the Hamming distance between two binary vectors, it is enough to first calculate the distance **(1)** between the corresponding subsets, and then multiply it by two. To calculate the distance **(1)** is somewhat easier, because, as follows from the proof of **Theorem 2**, only those coordinates of the two binary vectors being compared are considered, where at least one vector has a coordinate value of **1**. And those coordinates where both binary vectors have zero values are not considered (and therefore not compared). This means that the number of comparisons when calculating the Hamming distance by formula **(6)** decreases. This means that calculating the new distance **(1)** is easier than calculating the corresponding Hamming distance.

It should be emphasized that the distance **(1)** introduced by us is in a certain sense ***universal*** than the Hamming distance, since only one condition is imposed on the set T: it is only required that the set T be the set of some (any) finite sets; and these finite sets can be finite subsets of ***any***, including infinite, sets.

Now, as ***an application***, consider the set **N** of natural numbers and transform it into a completely new (unusual, previously unexplored) metric space using distance **(1)** as follows. To each natural number **n** we associate the set n_d of all its

divisors. It is clear that n_d is a finite set for any **n**. Now we introduce the distance **η(a,b)** between two natural numbers ***a*** and ***b*** by the formula

$$\eta(a,b) = \rho(a_d, b_d), \qquad (7)$$

where $\rho(a_d, b_d)$ is the distance **(1)**. Then, according to **Theorem 1,** the set **N** of natural numbers will be a metric space N_η with distance **(7)**. This new infinite discrete metric space N_η can be the subject of close study from the point of view of classical number theory. But this may already be a topic for further research. It should be noted that in the space N_η the distance between any two adjacent powers of a prime is always $\frac{1}{2}$. Moreover, the following theorem holds.

Theorem 3. A natural number $p \in N$ is prime if and only if for any natural number $m \in N$, the equality

$$\eta(p^m, p^{m+1}) = \frac{1}{2}$$

Proof. Let the number of all divisors of the number p^m be equal to d^1, and the number of all divisors of the number p^{m+1} equal to d^2. Since the set of divisors of the number p^m is a subset of the set of divisors of the number p^{m+1}, then according to **(7)** and **(1)** the distance between these numbers in the metric space N_η can be calculated by the following formula

$$\eta(p^m, p^{m+1}) = \frac{d_1+d_2}{2} - d_1\ . \qquad (8)$$

To prove the theorem, it suffices to show that for a prime p, formula **(8)** gives the value $\frac{1}{2}$, and for a non-prime number p, another value.

First case. Let the number p be prime. Then it is known from number theory that $d_1=m+1$ and $d_2=m+1+1$. Substituting these values in (8), we get the value $\frac{1}{2}$

Second case. Let the number p be non-prime, $p \neq 1$. We will split this case into two cases:

1) $p = p_1^n$,where p_1 is prime and $n \geq 2$. In this case $p^m = p_1^{nm}$ and $p^{m+1} = p_1^{n(m+1)}$.That's why $d_1=nm+1$ and $d_2=n(m+1)+1$. Substituting these formulas in **(8)**, we get: $\eta(p^m, p^{m+1}) = \frac{n}{2} \geq 1$, since $n \geq 2$.

2) $p = p_1^{n_1} p_2^{n_2} \ldots p_k^{n_k}$,where all p_i–prime, all n_i–natural,$i=1,2,\ldots k$ and $k \geq 2$. Then, for the numbers of divisors of the numbers p^m and p^{m+1}, respectively, the equalities are valid: $d_1=(mn_1+1)(mn_2+1)\ldots(mn_k+1)$,
$d_2=((m+1)n_1+1)((m+1)n_2+1)\ldots((m+1)n_2+1)$.
Let us introduce the notation: $x_i=mn_i+1, i=1,2,\ldots,k$.
Then:
$d_1=x_1x_2\ldots x_k$,
$d_2=(x_1+n_1)(x_2+n_2)\ldots(x_k+n_k)$.
After expanding the parentheses in the last expression for d_2, we get a sum consisting of 2^k terms, each of which is at least 1; one of the terms is $x_1x_2\ldots x_k=d_1$. The sum of all other terms, except for $x_1x_2\ldots x_k$, will be denoted by Y; in this sum, the number of terms is 2^k-1 and, since $k \geq 2$ and each term is not less than 1, then $Y \geq 3$. Further, by virtue of the notation, we have $d_2=d_1+Y$. Now we find the distance by formula **(8)** and estimate it:

$$\eta(p^m, p^{m+1}) = \frac{d_1+d_2}{2} - d_1 = \frac{d_1+d_1+Y}{2} - d_1 = \frac{Y}{2} \geq \frac{3}{2}.$$

Third case. Now let $p=1$. Then $\eta(p^m, p^{m+1})= \eta(1,1)=0$.

The theorem is proved.

References

1. Hamming, R. W. (1983). Coding theory and information theory: Translated from English. *Moscow,"Radio and communication, 176.*