RAIRO-Theor. Inf. Appl. 48 (2014) 495–503
DOI: 10.1051/ita/2014022

Available online at:
www.rairo-ita.org

ON MINIMAL HAMMING COMPATIBLE DISTANCES

Parsa Bakhtary[1] and Othman Echi[2]

Abstract. A Hamming compatible metric is an integer-valued metric on the words of a finite alphabet which agrees with the usual Hamming distance for words of equal length. We define a new Hamming compatible metric and show this metric is minimal in the class of all "well-behaved" Hamming compatible metrics. This gives a negative answer to a question stated by Echi in his paper [O. Echi, *Appl. Math. Sci. (Ruse)* **3** (2009) 813–824.].

Mathematics Subject Classification. 54E35, 68R01.

1. Introduction

Similarity concepts occur in many scientific areas such as computer science, biology and medicine.

It has been pointed out that, in biology, no theory exists on how to choose the best similarity measure [1]. The most common used are distance-based measures. The choice of a similarity measure is relevant on the final result of the analysis.

In 1950, Richard Hamming has introduced a metric on words with the same length in fundamental paper [3]. This metric was named after him (the Hamming distance).

Hamming metric is useful in error detecting and error correcting codes and in telecommunication to count the number of flipped bits in a fixed-length binary word.

Keywords and phrases. Edit distance, Hamming distance, metric space.

[1] (Data Scientist) Facebook, Inc. Menlo Park, CA 94025, USA. bakhtary@gmail.com

[2] King Fahd University of Petroleum and Minerals, Department of Mathematics and Statistics PO Box 5046, Dhahran 31261, Saudi Arabia. othechi@yahoo.com; echi@kfupm.edu.sa

Article published by EDP Sciences

© EDP Sciences 2015

https://doi.org/10.1051/ita/2014022 Published online by Cambridge University Press

In biology and other sciences, Hamming distance is often used to measure similarity by interpreting, mathematically, biological structures in term of binary strings (see for instance [1, 4, 6, 7]).

In genetic, the Hamming distance between two genetic sequences is the number of nucleotide differences between the given sequences.

Ever since Richard Hamming's seminal paper, the notion of the Hamming distance has played a fundamental role in the development of coding theory, error-correcting codes, cryptography, telecommunication, and information theory. Simply put, the Hamming distance between two words (or strings) of equal length counts the number of places where the corresponding letters differ. For instance 'coma' and 'comb' have Hamming distance 1, while 'sunny' and 'burnt' have Hamming distance 3, and it is easy to check that this is in fact a metric on the set of words of a given length.

However, the classical Hamming distance is restrictive in the sense that it only measures the distance between two words of equal length. For strings with distinct lengths, a well known method to measure the similarity between two words is "edit distance". If the permitted operations are substitution, deletion, and insertion, then the edit distance [5] measures the minimum number of operations that are required to transform one word into the other one. The goal is to find such a sequence of operations of minimum length; this length is the edit distance between the two words.

One would also like to have a similar metric for words of different lengths which agrees with the Hamming distance when those words are of the same length. To this end, Echi introduced the notion of an integer-valued *Hamming compatible* metric and gave a natural example denoted by T (see [2]). The author of [2] asked whether the metric T is minimal among the Hamming compatible metrics. We provide here a negative answer.

A natural question that arises is "How small can such a metric be?"

Using only the axioms of a metric and the fact that it is integer-valued and Hamming compatible, one may try to find the smallest such metric on a language. Unfortunately, these properties alone are not enough to say something substantive about the minimality of such metrics, and we will see there are some examples which do not have the desired behavior. On the other hand, if we make some natural uniformity assumptions on the metric, we can show that there is indeed a smallest such metric.

In what follows, we fix a finite set Σ, called an alphabet, and denote by Σ_n the set of n-letter words. The collection of all words of finite length will be denoted by Σ^*, called the improper language on Σ. All metrics are assumed to be integer-valued.

2. The ρ metric

If $u = u_1 \dots u_n$ is a word with $u_1, \dots, u_n$ are letters in an alphabet Σ, then we say that u is of length n, and we write $l(u) = n$.

https://doi.org/10.1051/ita/2014022 Published online by Cambridge University Press

We begin with the following characterization of Hamming distance, H.

Proposition 2.1. *Let $d : \Sigma^* \times \Sigma^* \longrightarrow \mathbb{N}$ be a mapping such that the induced map d_n on $\Sigma_n \times \Sigma_n$ is a metric for every $n \in \mathbb{N}$. Then $d = H$ is the Hamming distance on Σ_n for every $n \in \mathbb{N}$ if and only if*

(1) *for all words u, v with $n = l(u) = l(v)$ we have $d(u, v) \leq n$, and*
(2) *$d(u_1v_1, u_2v_2) = d(u_1, u_2) + d(v_1, v_2)$ whenever $l(u_1) = l(u_2)$ and $l(v_1) = l(v_2)$.*

Proof. It is obvious by the definition that the Hamming distance satisfies properties (1) and (2). To show the converse, we induct on $n = l(u) = l(v)$. For $n = 1$, $u = a$ and $v = b$ are letters. Property (1) implies that $d(a, b) \leq 1$ and can only be 0 if $a = b$ since d is a metric. Hence, if $a \neq b$, we must have $d(a, b) = 1$. So d is the Hamming distance for letters, *i.e.* for words of length 1.

For the inductive step let u', v' be words of length $n + 1$ and write $u' = au, v' = bv$ where a, b are letters and u, v are words of length n. Using property (2), the $n = 1$ case, and the inductive hypothesis we have

$$d(u', v') = d(a, b) + d(u, v) = H(a, b) + H(u, v) = H(u', v').$$

Hence, d is the Hamming distance. □

Definition 2.2. A metric d on the language Σ^* is called *Hamming compatible* if for any $n \in \mathbb{N}$, $d(u, v) = H(u, v)$ for all $u, v \in \Sigma_n$.

Let $H(\underline{u}, \underline{v})$ be the truncated Hamming function, defined as follows:
– If $l(u) \geq l(v)$, then $\underline{u}$ drops the last $l(u) - l(v)$ letters of u and $\underline{v} = v$, so that $H(\underline{u}, \underline{v}) = H(\underline{u}, v)$ is the usual Hamming distance between two words of length $l(v)$.
– If $l(u) < l(v)$, then $\underline{v}$ drops the last $l(v) - l(u)$ letters of v and $\underline{u} = u$, so that $H(\underline{u}, \underline{v}) = H(u, \underline{v})$ is the usual Hamming distance between two words of length $l(u)$.
Observe that the truncated Hamming function is not a metric!

Example 2.3. In [2], the second author defined $T(u, v) := H(\underline{u}, \underline{v}) + |l(u) - l(v)|$, and showed that it is a Hamming compatible metric. It is easily seen to be Hamming compatible, and one checks the triangle inequality by exhausting the cases.

We define the following:

$$\rho(u, v) = H(\underline{u}, \underline{v}) + \left\lceil \frac{|l(u) - l(v)|}{2} \right\rceil,$$

for words u, v in the language Σ^* over an alphabet Σ.

Proposition 2.4. *The function $\rho(u, v) = H(\underline{u}, \underline{v}) + \lceil \frac{|l(u)-l(v)|}{2} \rceil$ is a metric on Σ^*.*

https://doi.org/10.1051/ita/2014022 Published online by Cambridge University Press

Proof. We check the triangle inequality for $\delta = \rho$. Let $u, v, w \in \Sigma^*$, and set $D = \delta(u,w) + \delta(w,v) - \delta(u,v)$. We must show $D \geq 0$ in all cases.

Case 1: $l(w) \leq l(u) \leq l(v)$

Then we may write $u = u_1u_2, v = v_1v_2v_3$, where $l(w) = l(u_1) = l(v_1)$ and $l(u_2) = l(v_2)$. We also have
$\delta(u,v) = H(u_1,v_1) + H(u_2,v_2) + \frac{l(v_3)+\alpha_1}{2}$
$\delta(u,w) = H(u_1,w) + \frac{l(u_2)+\alpha_2}{2}$
$\delta(w,v) = H(v_1,w) + \frac{l(v_2)+l(v_3)+\alpha_3}{2}$, where each $\alpha_i \in \{0,1\}$. Then we have

$$D = [H(u_1,w) + H(v_1,w) - H(u_1,v_1)] + [l(u_2) - H(u_2,v_2)] + \frac{\alpha_2+\alpha_3-\alpha_1}{2}.$$

Since H is a metric on Σ_n for $n = l(w)$, the first term is nonnegative. The second term is nonnegative, as it is just property (1) from the previous proposition. It suffices to show the third term $L = \frac{\alpha_2+\alpha_3-\alpha_1}{2}$ is nonnegative.

Subcase 1A: $l(u) - l(v)$ is even. Then $l(u) - l(w)$ and $l(v) - l(w)$ are either both even or both odd. If both are even, then $\alpha_1 = \alpha_2 = \alpha_3 = 0$, so $L = 0 \geq 0$. If both are odd, then $\alpha_1 = 0, \alpha_2 = \alpha_3 = 1$ so $L = 1 \geq 0$.

Subcase 1B: $l(u) - l(v)$ is odd. Then exactly one of $l(u) - l(w)$ and $l(v) - l(w)$ is odd, and the other is even. Hence $\alpha_1 = 1$ and $\{\alpha_2, \alpha_3\} = \{0,1\}$, so $L = 0 \geq 0$.

Case 2: $l(u) \leq l(w) \leq l(v)$

Then we may write $w = w_1w_2, v = v_1v_2v_3$, where $l(u) = l(w_1) = l(v_1)$ and $l(w_2) = l(v_2)$. We also have
$\delta(u,v) = H(u,v_1) + \frac{l(v_2)+l(v_3)+\alpha_1}{2}$
$\delta(u,w) = H(u,w_1) + \frac{l(w_2)+\alpha_2}{2}$
$\delta(w,v) = H(w_1,v_1) + H(w_2,v_2) + \frac{l(v_3)+\alpha_3}{2}$, where each $\alpha_i \in \{0,1\}$. Then we have

$$D = [H(u,w_1) + H(w_1,v_1) - H(u,v_1)] + H(w_2,v_2) + \frac{\alpha_2+\alpha_3-\alpha_1}{2}.$$

Again, it suffices to show the third term $L = \frac{\alpha_2+\alpha_3-\alpha_1}{2}$ is nonnegative.

Subcase 2A: $l(u) - l(v)$ is even. Then $l(u) - l(w)$ and $l(v) - l(w)$ are either both even or both odd. If both are even, then $\alpha_1 = \alpha_2 = \alpha_3 = 0$, so $L = 0 \geq 0$. If both are odd, then $\alpha_1 = 0, \alpha_2 = \alpha_3 = 1$ so $L = 1 \geq 0$.

Subcase 2B: $l(u) - l(v)$ is odd. Then exactly one of $l(u) - l(w)$ and $l(v) - l(w)$ is odd, and the other is even. Hence $\alpha_1 = 1$ and $\{\alpha_2, \alpha_3\} = \{0,1\}$, so $L = 0 \geq 0$.

Case 3: $l(u) \leq l(v) \leq l(w)$

Then we may write $v = v_1v_2, w = w_1w_2w_3$ where $l(u) = l(v_1) = l(w_1)$ and $l(v_2) = l(w_2)$. We also have
$\delta(u,v) = H(u,v_1) + \frac{l(v_2)+\alpha_1}{2}$
$\delta(u,w) = H(u,w_1) + \frac{l(w_2)+l(w_3)+\alpha_2}{2}$
$\delta(w,v) = H(w_1,v_1) + H(w_2,v_2) + \frac{l(w_3)+\alpha_3}{2}$, where each $\alpha_i \in \{0,1\}$. Then we have

$$D = [H(u,w_1) + H(w_1,v_2) - H(u,v_1)] + H(w_2,v_2) + l(w_3) + \frac{\alpha_2+\alpha_3-\alpha_1}{2}.$$

https://doi.org/10.1051/ita/2014022 Published online by Cambridge University Press

Again, it suffices to show the third term $L = \frac{\alpha_2+\alpha_3-\alpha_1}{2}$ is nonnegative. An argument analogous to the previous cases shows that $L \geq 0$.

Hence, in all cases $\delta = \rho$ satisfies the triangle inequality and obviously symmetry and separation properties; so it is an integer-valued metric, and it is obvious that it is Hamming compatible. □

Remark 2.5. It was asked by the second author in [2], Example 2.3 if the metric T in is minimal in its class. That is to say, if δ is a Hamming compatible metric, must it always be the case that $\delta \geq T$? As we have just seen, ρ is such a metric that satisfies $\rho \leq T$ by definition, with strict inequality holding for appropriate pairs of words. However, as we will see in Section 3, there exist Hamming compatible metrics that take values even smaller than ρ.

3. Minimality of the ρ metric

Here we explore lower bounds of a Hamming compatible metric δ on a language Σ^*. Recall that ε is the empty word.

Definition 3.1. We say that two words $u, v \in \Sigma^*$ of the same length $n = l(u) = l(v)$ are *Hamming opposites* if the Hamming distance between them is maximal, *i.e.* $H(u, v) = n$.

Definition 3.2. We say that δ is *weakly uniform* if for any Hamming opposites u, v we have that $\delta(u, \varepsilon) = \delta(v, \varepsilon)$.

Proposition 3.3. *If δ is weakly uniform, then $\delta(u, \varepsilon) \geq \frac{l(u)}{2}$ for all $u \in \Sigma^*$. In particular, $\delta(u, \varepsilon) \to \infty$ as $l(u) \to \infty$.*

Proof. Let u be a word of length $l(u) = n$, and choose a Hamming opposite v. Then the triangle inequality gives

$$n = H(u, v) = \delta(u, v) \leq \delta(u, \varepsilon) + \delta(v, \varepsilon) = 2\delta(u, \varepsilon),$$

hence $\delta(u, \varepsilon) \geq \frac{n}{2}$. □

Denote by $H = H(\underline{u}, \underline{v})$ the truncated Hamming function of $u, v \in \Sigma^*$. Given a Hamming compatible metric δ, set $\gamma = \delta - H$. Thus, we may write $\delta = H + \gamma$ where H is the truncated Hamming function. Observe that δ is Hamming compatible if and only if $\gamma(u, v) = 0$ whenever $l(u) = l(v)$.

Since $H(\underline{u}, \underline{\varepsilon}) = 0$ implies $\delta(u, \varepsilon) = \gamma(u, \varepsilon)$, requiring that δ be weakly uniform is equivalent to requiring that $\gamma(u, \varepsilon) = \gamma(v, \varepsilon)$ for all Hamming opposites u, v.

Definition 3.4. Given $\delta = H + \gamma$ as above, we say that δ is *uniform* if given any pair of Hamming opposites $u, v \in \Sigma_n$, we have $\gamma(u, w) = \gamma(v, w)$ for all $w \in \Sigma^*$.

https://doi.org/10.1051/ita/2014022 Published online by Cambridge University Press

Example 3.5. For $T(u,v) := H(\underline{u},\underline{v}) + |l(u) - l(v)|$, so here

$$\gamma(u,v) = |l(u) - l(v)|.$$

Thus the metric T is uniform.

Example 3.6. For $\rho(u,v) = H(\underline{u},\underline{v}) + \lceil \frac{|l(u)-l(v)|}{2} \rceil$, so here $\gamma(u,v) = \lceil \frac{|l(u)-l(v)|}{2} \rceil$. Consequently, ρ is uniform.

Whereas weak uniformity is the simple notion that Hamming opposites are equidistant from the empty word ε, this definition of uniformity is slightly more mysterious. If we fix $w = \varepsilon$ this condition reduces to weak uniformity. Notice both examples above are uniform, and more generally, if $\gamma(u,w)$ depends only on the lengths $l(u), l(w)$ of the inputs u, w, then δ is uniform.

Lemma 3.7. *Suppose we are given words $u, w \in \Sigma^*$ with $l(u) \geq l(w)$. Then there exists a Hamming opposite v for u such that*

$$H(\underline{u},\underline{w}) + H(\underline{v},\underline{w}) = l(w).$$

Proof. Write $u = u_1 u_2 \dots u_n$ and $w = w_1 w_2 \dots w_m$, where $m \leq n$ by assumption. Then let $k = H(\underline{u},\underline{w}) = H(u', w)$, where $u' = u_1 u_2 \dots u_m$. Now choose v_i by the following method. For $1 \leq i \leq m$, set $v_i = w_i$ if $w_i \neq u_i$. If $w_i = u_i$, then take v_i to be any letter in Σ different from u_i. For $m < i \leq n$, just take v_i to be any letter in Σ different from u_i.

Now, it is clear by the construction that $v = v_1 v_2 \dots v_n$ is a Hamming opposite for u since they differ in each letter place. Let $v' = v_1 v_2 \dots v_m$, so that $H(\underline{v},\underline{w}) = H(v', w)$. Since u' and w differ in k places, and v' agrees with w in precisely those k places and differs elsewhere, $H(v', w) = m - k$. Hence, we have

$$H(u', w) + H(v', w) = k + (m - k) = m = l(w). \qquad \square$$

The following theorem is the main result.

Theorem 3.8. *Suppose δ is a uniform Hamming compatible metric on a language Σ^*. Then $\delta(u,v) \geq \rho(u,v)$ for all $u, v \in \Sigma^*$. In particular, $\delta(u,v) \to \infty$ as $|l(u) - l(v)| \to \infty$.*

Proof. We must show that for all $u, w \in \Sigma^*$, $\delta(u,w) \geq \rho(u,w)$. Writing $\delta = H + \gamma$, this is equivalent to showing that $\gamma(u,w) \geq \lceil \frac{|l(u)-l(w)|}{2} \rceil$. Assume without loss of generality that $l(u) \geq l(w)$, and choose a Hamming opposite v for u that satisfies the equality of the above lemma. Now, we have

$$\begin{aligned} l(u) = H(u,v) = \delta(u,v) &\leq \delta(u,w) + \delta(v,w) \\ &= H(\underline{u},\underline{w}) + \gamma(u,w) + H(\underline{v},\underline{w}) + \gamma(v,w). \end{aligned}$$

By our choice of v, $H(\underline{u},\underline{w}) + H(\underline{v},\underline{w}) = l(w)$, so

$$l(u) \leq l(w) + \gamma(u,w) + \gamma(v,w) = l(w) + 2\gamma(u,w)$$

by the uniformity assumption. Hence, $\gamma(u,w) \geq \frac{l(u)-l(w)}{2}$. Because γ is integer valued, it follows that $\gamma(u,w) \geq \lceil \frac{|l(u)-l(w)|}{2} \rceil$, which completes the proof. $\square$

https://doi.org/10.1051/ita/2014022 Published online by Cambridge University Press

Example 3.9. Uniformity is necessary for the above minimality result to hold. Let $\Sigma = \{0, 1\}$ and define

$$\delta(\varepsilon, u) = \delta(u, \varepsilon) = \begin{cases} 0 & \text{if } u = \varepsilon \\ 1 & \text{if } u = 000 \\ \rho(u, 000) & \text{otherwise} \end{cases}$$

where ρ is the metric defined previously. Then take $\delta(u, v) = \rho(u, v)$ for all $u, v \in \Sigma^* - \{\varepsilon\}$. We first show δ is a metric. Since we have only changed distances to ε, and ρ has been shown to be a metric, we only need to check the triangle inequality for expressions involving ε. First consider

$$\delta(u, \varepsilon) \leq \delta(u, v) + \delta(v, \varepsilon).$$

If $u \in \{\varepsilon, 000\}$ then this is trivial. If $u \notin \{\varepsilon, 000\}$ then this becomes

$$\rho(u, 000) \leq \delta(u, v) + \delta(v, \varepsilon).$$

If $v \in \{\varepsilon, 000\}$ then again this is trivial, so assuming $v \notin \{\varepsilon, 000\}$ this becomes

$$\rho(u, 000) \leq \rho(u, v) + \rho(v, 000)$$

which is true because ρ is a metric. Now consider the expression

$$\delta(u, v) \leq \delta(u, \varepsilon) + \delta(v, \varepsilon).$$

Again, if $u \in \{\varepsilon, 000\}$ then this is trivial, and by symmetry the same is true if $v \in \{\varepsilon, 000\}$. If neither u nor v are these words, then this becomes

$$\rho(u, v) \leq \rho(u, 000) + \rho(v, 000)$$

which is again true because ρ is a metric. Hence, δ is a Hamming compatible integer valued metric.

It is easy to see that δ is not even weakly uniform, since $\delta(0, \varepsilon) = \rho(0, 000) = 1$ and $\delta(1, \varepsilon) = \rho(1, 000) = 2$. Moreover, by construction $\delta(000, \varepsilon) = 1 < 2 = \lceil \frac{3}{2} \rceil = \rho(000, \varepsilon)$, so it takes values smaller than ρ.

Example 3.10. The following variant of the above example shows that there exist weakly uniform metrics which are not uniform. As before, let $\Sigma = \{0, 1\}$ and define

$$\delta(0, u) = \delta(u, 0) = \begin{cases} 0 & \text{if } u = 0 \\ 1 & \text{if } u \in \{\varepsilon, 1, 11\} \\ T(u, 11) & \text{otherwise} \end{cases}$$

where T is the metric defined previously. Then take $\delta(u, v) = T(u, v)$ for all $u, v \in \Sigma^* - \{0\}$. We first show δ is a metric. Again, since we have only changed distances

https://doi.org/10.1051/ita/2014022 Published online by Cambridge University Press

to 0, we have not changed the distance from ε to 0 or from 1 to 0, and T is a metric, we only need to check the triangle inequality for expressions involving 0. First consider the inequality

$$\delta(u,0) \leq \delta(u,v) + \delta(v,0).$$

If $u \in \{\varepsilon, 0, 1, 11\}$, then the required inequality is trivial. If $u \notin \{\varepsilon, 0, 1, 11\}$, then the inequality becomes

$$T(u,11) \leq \delta(u,v) + \delta(v,0).$$

Now we may assume $l(u) \geq 2$. If $v = \varepsilon$ this is $T(u,11) \leq l(u)+1$, which is true since $T(u,11) = H(\underline{u},11)+l(u)-2 \leq l(u)$. If $v = 0$ this is $T(u,11) \leq \delta(u,0)+0 = T(u,11)$. If $v = 1$ this is $T(u,11) \leq \delta(u,1)+1$. This holds, because the left hand side is

$$H(\underline{u},11) + l(u) - 2$$

and the right hand side is

$$H(\underline{u},1) + l(u) - 1 + 1 = H(\underline{u},1) + l(u).$$

If $v = 11$, then the required inequality is

$$T(u,11) \leq \delta(u,11) + 1 = T(u,11) + 1.$$

Now, assuming $v \notin \{\varepsilon, 0, 1, 11\}$ the required inequality becomes

$$T(u,11) \leq T(u,v) + T(v,11)$$

which is true because T is a metric.

Now consider the expression

$$\delta(u,v) \leq \delta(u,0) + \delta(v,0).$$

If $u = 0$ this is trivial. If $u = 1$, this becomes

$$\delta(1,v) \leq 1 + \delta(v,0).$$

If $v \in \{\varepsilon, 0, 1\}$ this is trivial, and if $v = 11$ this is $1 = T(1,11) \leq 1+1$. If $v \notin \{\varepsilon, 0, 1, 11\}$ then $l(v) \geq 2$ and this becomes $T(1,v) \leq 1 + T(v,11)$. This is

$$H(\underline{v},1) + l(v) - 1 \leq 1 + H(\underline{v},11) + l(v) - 2$$

which holds because $H(\underline{v},1) \leq H(\underline{v},11)$ when $v \neq 11$. Now suppose $u = \varepsilon$. Then our inequality becomes $\delta(\varepsilon,v) \leq 1 + \delta(v,0)$. If $v \in \{\varepsilon, 0, 1, 11\}$ this is obvious, so assume $v \notin \{\varepsilon, 0, 1, 11\}$, so $l(v) \geq 2$, and this is $l(v) \leq 1+T(v,11) = 1+H(\underline{v},11)+l(v)-2$. But this is just saying that $H(\underline{v},11) \geq 1$, which is true since $v \neq 11$ and $l(v) \geq 2$.

https://doi.org/10.1051/ita/2014022 Published online by Cambridge University Press

If $u = 11$ then our inequality becomes $\delta(11, v) \leq 1 + \delta(v, 0)$. If $v \in \{\varepsilon, 0, 1, 11\}$ this is clear, and if $v \notin \{\varepsilon, 0, 1, 11\}$ this is just saying that $T(11, v) \leq 1 + T(v, 11)$ which is trivial. The final case is when $u, v \notin \{\varepsilon, 0, 1, 11\}$, in which case our inequality becomes

$$T(u, v) \leq T(u, 11) + T(v, 11)$$

which is again true because T is a metric. Hence, δ is a Hamming compatible integer valued metric.

Now, δ is weakly uniform because $\delta(0, \varepsilon) = \delta(1, \varepsilon) = 1$, and for $l(u) > 1$, we have $\delta(u, \varepsilon) = T(u, \varepsilon) = T(v, \varepsilon) = \delta(v, \varepsilon)$ for all $u, v \in \Sigma_n$. However, δ is not uniform because $\gamma(11, 0) = \delta(11, 0) - H(\underline{11}, \underline{0}) = 1 - 1 = 0$, while $\gamma(00, 0) = \delta(00, 0) - H(\underline{00}, \underline{0}) = 2 - 0 = 2$.

Acknowledgements. The authors thank the referee for his valuable comments and suggestions. The second author thanks King Fahd University of Petroleum and Minerals (Saudi Arabia) for its support.

References

[1] A. Brazma and J. Vilo, Gene expression data analysis. *FEBS Lett.* **480** (2000) 17–24.
[2] O. Echi, The cardinality of a sphere relative to an edit distance. *Appl. Math. Sci. (Ruse)* **3** (2009) 813–824.
[3] R.W. Hamming, Error detecting and error correcting codes. *Bell System Technical Journal* **29** (1950) 147–160.
[4] E. Halperin, J. Buhler, R. Karp, R. Krauthgamer and B. Westover, Detecting protein sequence conservation via metric embeddings. *Bioinformatics* **19** (2003) i122–i129.
[5] V.I. Levenshtein, Binary codes capable of correcting, deletions, insertions and reversals. *Soviet Phys. Dokl.* **10** (1966) 707–710.
[6] M.X. He, S.V. Petoukhov and P.E. Ricci, Genetic code, Hamming distance and stochastic matrices. *Bull. Math. Biol.* **66** (2004) 1405–1421.
[7] C.D. Pilcher, J.K. Wong and S.K. Pillai, Inferring HIV Transmission Dynamics from Phylogenetic Sequence Relationships. *PLoS Med.* **5** (2008) e69.

Communicated by D. Augot.
Received September 20, 2013. Accepted May 26, 2014.

https://doi.org/10.1051/ita/2014022 Published online by Cambridge University Press