Reports of the Academy of Sciences of the USSR

1964

SovietRxiv

Abstract

Full Text

Reports of the Academy of Sciences of the USSR
1964. Vol. 157, No. 2

CYBERNETICS AND CONTROL THEORY

M. A. KARMAZIN

ON A CLASS OF ERROR-CORRECTING CODES

(Presented by Academician A. N. Kolmogorov on 4 XII 1964)

Let E^n be the set of all sequences of zeros and ones of length n. An arbitrary subset A of the set E^n will be called a **code of length** n; the sequences belonging to A, **code words**, and the number of elements in the set A, the **power of the code**. The H**-weight** $w(r)$ of a code word r will mean the number of ones in this word; the H**-distance** $d(r_1, r_2)$ between two words r_1 and r_2, the number

$$d(r_1, r_2) = w(r_1 \oplus r_2),$$

where the sign $\oplus$ denotes symbol-by-symbol addition modulo 2; the H**-distance of the code** A, the number

$$d(A) = \min_{r_i, r_j} d(r_i, r_j) \quad (r_i \neq r_j)$$

(the minimum is taken over all noncoincident pairs r_i and r_j).

The notation introduced turns out to be useful in studying an information-transmission scheme. There is (see [1]) the following

Theorem. *In order that a code correct t (detect $2t$) errors, it is necessary and sufficient that it have code distance not less than $2t + 1$.*

There is another situation in which error-correcting codes prove useful. Suppose we have some set Ω of numbers whose sums we need to obtain at the output of an adder operating with errors. In order nevertheless to avoid errors in addition, we shall add in the adder, instead of the numbers of the set Ω, numbers of some other set A, called a code, which is isomorphic under addition to the image of the set Ω. For convenience we shall henceforth assume that the set Ω consists of $2^k - 1$ numbers $\{1, \ldots, 2^k - 1\}$, and consider only the special type of codes introduced in [1], Chapter 13. We shall call a kN**-code with** k **information and** $[\log(N+1)]$ **check symbols** the collection of numbers $N, 2N, \ldots, (2^k-1)N$.

Sometimes, instead of the index kN, which specifies the parameters of the code, we shall write the index N_n, where n is the length of the word 2^{kN}, i.e., a kN-code and an $N_k + \log(N+1)$-code denote one and the same code of 2^k elements.

Let us call the P**-weight** $\bar{w}(2)$ of a number r the minimum possible number of terms on the right-hand side of the equality

$$r = \sum a_i 2^i,$$

where $a_i = \pm 1$. Let us call the P**-distance** $\bar{d}(r_1, r_2)$ between two numbers r_1 and r_2 the number $\bar{d}(r_1, r_2) = \bar{w}(r_1 - r_2)$, and the P**-distance of the code** A, the number $\bar{d}(A) = \min_{r_1, r_2} \bar{d}(r_1, r_2)$, where the minimum is taken over all noncoincident pairs r_1 and r_2. Note that the easily proved inequality holds

$$d(A) \geqslant \bar{d}(A). \qquad (*)$$

In ([1]) it is proved that a code A corrects all t-fold (detects all $2t$-fold) errors in an adder if and only if it satisfies—

the inequality

$$\bar{d}(A) \geq 2t + 1.$$

We shall also need the following theorem, contained in (1):

Theorem A. *In order that an N_n-code have P-distance not less than t, it is necessary and sufficient that no number of P-weight $t-1$ and length n be divisible by N.*

Theorem 1. Let $K(n, \bar{d})$ be the number of points of an N_n-code with code distance $\bar{d}$ ($\bar{d} = 2s + 1$). Then the inequality holds

$$K(n, \bar{d}) \leq \frac{2^{n - \frac{\bar{d}-3}{2}}}{\sum\limits_{i=0}^{\frac{\bar{d}-1}{2}-2} C_{n-i-2}^i + \sum\limits_{i=0}^{\frac{\bar{d}-1}{2}-1} C_{n-i-1}^i}. \qquad (1)$$

Theorem 2. Let

$$N_0 = \left[\sum_{i=0}^{\bar{d}-2} C_{n-i-2}^i - \sum_{i=0}^{\bar{d}-1} C_{n-i-1}^i \right] 2^{\bar{d}-1} \bar{r}(2^n). \qquad (2)$$

Then there exists an N_n-code with code distance $\bar{d}$, where $N \leq N_0$; $\bar{r}(s) = \max_{j<s} r(j)$, and $r(j)$ is the number of divisors of j (Euler' s function of j).

For the proof of the theorems we use first of all

Lemma 1. *The number of distinct odd numbers, less than 2^n, of P-weight $\leq \bar{d}$, is equal to*

$$\left(\sum_{i=0}^{\bar{d}-2} C_{n-i-2}^{i} - \sum_{i=0}^{\bar{d}-1} C_{n-i-1}^{i}\right) 2^{\bar{d}-1}.$$

The lemma follows from the following two considerations:

1. Any number r of length n can be represented uniquely in the form

   $$r = \sum a_{ij} 2^{ij} \quad (a_{ij} = \pm 1)$$

   so that the inequality

   $$\min_{l,k} |j_l - l_k| \geq 2, \quad l \neq k$$

   is satisfied.

2. There exists a one-to-one mapping of the set of sequences of H-weight $\bar{d}$, consisting of zeros and ones such that between any two ones there is at least one zero, and beginning and ending with ones, onto the entire set of sequences of zeros and ones of H-weight $\bar{d}$ and length $n - d + 2$, also beginning and ending with ones.

From Lemma 1, Theorem 1 is obtained analogously to the way in which the Hamming bound is derived in the theory of error-correcting codes.

For the proof of Theorem 2, consider the sets $V(n, \bar{d})$ of numbers of length n, of P-weight not greater than $\bar{d}$. According to Lemma 1, the set $V(n, \bar{d})$ contains

$$\left(\sum_{i=0}^{\bar{d}-2} C_{n-i-2}^{i} - \sum_{i=0}^{\bar{d}-1} C_{n-i-1}^{i}\right) 2^{\bar{d}-1}$$

elements. The set of numbers that are divisors of at least one of the numbers belonging to $V(n, \bar{d})$ contains no more than

$$\left(\sum_{i=0}^{\bar{d}-2} C_{n-i-2}^{i} + \sum_{i=0}^{\bar{d}-1} C_{n-i-1}^{i}\right) 2^{\bar{d}-1} \bar{r}(2^n)$$

elements, and it is obvious that among the first N numbers, where N is the same as in the formulation of Theorem 2, there is at least one which does not belong to this set. Theorem 2 follows from this.

The formulas (1) and (2) take on a simpler form in the most natural asymptotic case, when $n \to \infty$ and $\bar{d} \simeq dn$. In this situation the following theorems are valid:

Theorem 1′.

$$\frac{\log k(n, \bar{d})}{n} \gtrsim \left[1 - \left(1 - \frac{d}{2}\right) H\left(\frac{d}{2(1 - d/2)}\right) - \frac{d}{2}\right]. \tag{3}$$

Theorem 2′. There exists an N such that

$$\frac{\log N}{n} \lesssim \left[1 - (1 - d) H\left(\frac{d}{1 - d}\right) + d\right], \tag{4}$$

and such that the N_n-code has code distance $\bar{d}$ ($\bar{d} \simeq dn$).

Here we write that $f_1(n) \lesssim f_2(n)$ if the condition

$$\lim_{n \to \infty} \frac{f_1(n)}{f_2(n)} \leq 1$$

is satisfied.

In the right-hand side of relation (4) we may omit the term $\log \bar{r}(n)$, since the following holds.

Lemma 2. For any s, for all k,

$$\bar{r}(s) \lesssim (\ln n)^{\sqrt{\ln n}} \cdot 2^{k \frac{\ln n}{\ln \ln n}}.$$

Lemma 2 can also be used to replace the function by its upper estimate in Theorem 2′. The following is true.

Theorem 3. If the multiplier N of a code of prescribed length n is chosen at random among the first M numbers, then the probability that the code obtained has distance less than $\bar{d}$ does not exceed $\dfrac{N_0}{M}$, where N_0 is determined by the equality

$$N_0 = \left[\sum_{i=0}^{\bar{d}-2} C_{i-n-2}^{i} + \sum_{i=0}^{\bar{d}-1} C_{i-n-1}^{i}\right] 2^{\bar{d}-1} \bar{r}(2^n).$$

In the asymptotic case considered above, when $n \to \infty$ and $\bar{d} \simeq dn$, N_0 is given by the equality

$$N \simeq 2^{[1-(1-d)H(\frac{d}{1-d})+d]n}.$$

Theorem 3 is proved from the same considerations as Theorem 2.

In the theory of correcting codes, effectively constructible codes with linearly growing code distance are still unknown (with the exception of Elias iterative codes, for which the code distance grows very slowly), and, when necessary, random codes are used. In this situation, the random codes obtained on the basis of Theorem 3 are also applicable. Note that it follows from formula (4) that, with probability tending exponentially to 1, these codes have code distance only slightly worse than that of the best known codes lying on the Varshamov-Gilbert bound. (The last argument may seem incorrect, since the Varshamov-Gilbert theorem involves not the P-distance but the H-distance of codes; however, this is immaterial in view of inequality (*).)

In works ([1],[2]), codes with code distance 3 are constructed, i.e., codes that correct 1 error and detect 2 errors. We indicate here a method for constructing codes with code distance 4 and 5.

Theorem 4. 1) If $t > \sqrt{n}$, then the $(2^t-1)(2^{t+1}-1)_n$-code has code distance 4.

2) If $t > \sqrt{n}$ and is odd, then the $(2^t-1)(2^{t+1}-1)(2^{t+2}-1)_n$-code has code distance 5.

3) If the condition $2 \cdot 3 \cdots p_i \cdots p_k > n^2$ is satisfied, where p_i is the i-th prime number, then the

$$\prod_1^k (2^{p_i}-1)_n$$

-code has code distance 4.

To prove the theorem, one should consider the set of numbers r representable in the form of the sum $r = \sum a_i 2^i$ ($a_i = \pm 1$) with 3 and 4 summands. The theorem makes essential use of the following two simple facts:

1. If the expression $2^l \pm 1$ is divisible by $2^{i_1}-1$ and by $2^{i_2}-1$, then it is also divisible by $2^{i_1 i_2}-1$.
2. The remainders upon division of numbers of the form 2^α by numbers $2^\beta - 1$ have the form α^γ.

Let us note that simple methods for decoding kN-codes are not yet known, and therefore, in their practical application, it is apparently advisable to use not so much their error-correcting capability as their error-detecting capability. Detecting errors by means of kN-codes is quite simple: for this it is only necessary to perform division by a fixed number.

Received
4 XII 1963

References

[1] W. Peterson, *Error-Correcting Codes*, N. Y., 1961. [2] D. T. Brown, IRE Trans. Electr. Comput., 9, 333 (1960).

Note: Figure translations are in progress. See original paper for figures.