

Math 3500-101
Instructor: Dr. V. Rykov

End of Class Report

A new construction of superimposed code

Prepared: Giang Nguyen

Advisor: Dr Rykov, Vladimir Ufimsev

University of Nebraska at Omaha
June 28, 2007

Abstract

Superimposed code has a wide application in group testing. There's several ways of construct the superimposed code. In this paper, we will introduce a new way of construct the code based from the trivial substitution then we construct the code and then test the code using computer aids tools.

Introduction

Group testing is the mathematical technique that commonly employed in the design of screening experiments [1], with objective of finding a specified number of defective units among a large population of units with no defects, in the least amount of tests. There are two types of algorithm in which group testing is used, adaptive and non adaptive. Non adaptive group testing is employed when there is a constraint on time (all tests must be carried out simultaneously) or when there is a cost constraint (obtaining information from other tests during the experiment will cost too much money). Non adaptive group testing algorithms can be represented by a binary matrix, where the columns of the matrix correspond to each unit in the population and the rows correspond to a test and determine which units of the population are to be tested. The binary matrix has constraints imposed on the component-wise Boolean sum of any s (or up to s) columns. Such matrices fall under Superimposed Coding theory and are known as superimposed codes [1]. In this paper, we focus on testing a method of construct a superimposed code, where a code is constructed by trivial substitution from a Reed Solomon matrix. We will particular measure the rate of covering of the code.

I. Notations and definitions

Superimposed Coding Theory

Definition 1.[2] Let $1 \leq s < t$ and $N \geq 1$ and $u(j) = (u_1(j), u_2(j), \dots, u_N(j))$, $j = \overline{1, s}$ denote the binary columns (of 0 and 1) of length N . The *Boolean sum* $\mathbf{u} = \mathbf{u}(1) \vee \mathbf{u}(2) \vee \dots \vee \mathbf{u}(s)$ of columns $\mathbf{u}(1), \dots, \mathbf{u}(s)$ is the binary column $\mathbf{u} = (u_1, u_2, \dots, u_N)$ with components:

$$u_i = \begin{cases} 0 & \text{if } u_i(1) = u_i(2) = \dots = u_i(s) = 0 \\ 1 & \text{otherwise} \end{cases}$$

Definition 2.[2] Let us say that column \mathbf{u} covers column \mathbf{v} iff $\mathbf{u} \vee \mathbf{v} = \mathbf{u}$.

Definition 3. (Superimposed code [1]). Let X be a code of length N and size t . Code X has strength s if and only if the Boolean sum of any s code packets does not cover any other code packet (not in the s -set) in X . Code X is then a *superimposed* code of strength s . The matrix X is also called an *s-disjunct* matrix.

Definition 3. The weight: $w(x)$ of codeword x is the number of non-zero elements in the codeword.

Definition 4. Let $x(u)$ and $x(v)$ be two codeword of length N . The intersection $\lambda(x(u), x(v))$ is defined as the number of place where both $x(u)$ and $x(v)$ has equal elements.

Let $F_q = \{0, 1, \dots, q-1\}$ be a set of q distinct elements. F_q is known as the *alphabet* and is often taken to be Z_q (integers by mod q) when q is a prime number. C is the set consisting of vectors of length n that are built from the elements of the alphabet F_q .

Definition 5. The *Hamming distance* between two vectors $\mathbf{x}, \mathbf{y} \in F_q^n$ is the number of places in which they are differ and is denoted by $d(\mathbf{x}, \mathbf{y})$.

Definition 6. A q -nary code C , of length n , size t , and minimum distance d , is a subset of F_q^n such that $d = \min\{d(\mathbf{x}, \mathbf{y}) \mid \mathbf{x}, \mathbf{y} \in F_q^n, \mathbf{x} \neq \mathbf{y}\}$. C is referred to as an (n, t, d) code.

Assume F_q is the *Galois field* $GF(q)$, the set of elements endowed with two operations (addition $+$ and multiplication \cdot), containing 0 and 1, and where any equation of the form $a \cdot x + b = c, a, b, c \in GF(q)$, has solution. Then q has to be prime power and we regard F_q^n as the vector space $V(n, q)$.

In this paper we suggest the follow construction of with trivial substitution from the matrix H .

$$H = \begin{bmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_{q^k-1} \\ \alpha_1^r & \alpha_2^r & \dots & \alpha_{q^k-1}^r \\ \alpha_1^t & \alpha_2^t & \dots & \alpha_{q^k-1}^t \end{bmatrix}$$

Where q is the prime with $q \geq 2$, $k \geq 3$ and $k \in N$, and r, t are the two least relatively prime to q . $\alpha_1, \alpha_2, \dots, \alpha_{q^k-1}$ are column vector of coefficient of all possible polynomial with degree of k over the $GF(q^k)$. Then H can be transformed into a binary superimposed code by applying the transformation where each symbol of $GF(q)$ is associated with a binary column vector of length q and weight 1 i.e.

$$[0 \ 1 \ 2 \ 3 \ \dots \ q-1] = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

Each symbol in C is then replaced by the binary column associated with it. This transformation produces a binary $(k.q.3) \times (q^k - 1)$ matrix X, which is the superimposed code of size $t = q^k - 1$, with codeword length of $(k.q.3)$.

II. Program

The program will generate all possible *Boolean sum* of the 3 code words and then check if the sum covers any of other code word (see appendix A for actual Maple code). Then we calculate the rate as following:

$$Rate = 1 - \left(\frac{\text{number of times a word is covered}}{\binom{p^k - 1}{3} (p^k - 1 - 3)} \right)$$

III. Result Obtained

p, k	3,3	4,3	5,3	6,3	7,3
Rate	—	0.0298	0.1462	—(*)	0.179
p, k	3, 4	4,4			
Rate	0.1033	—(*)			

*Result is still being calculated.

IV. Conclusion

As when we measured the rate which is the ratio of number of times a code word is covered by the Boolean sum of 3 code word in the code, we know how the strength of the code. If the rate is 0 means that there is no code word got covered then the code will have strength of 3. On the other hand, we also only interest in those has number of columns are more than number of rows (number of tests less than number of testing units). Therefore we need:

$$q^k - 1 > k.q.3$$

So far, we haven't found any code that has perfect rate of 1. However, as we increased p and k the rate increased.

Reference:

[1] Vyacheslav V. Rykov, Vladimir V. Ufimtsev, "Group testing and its Application to Multiple Access information Transmission Models". *Final report*, University of Nebraska At Omaha, March 27, 2006.

[2] A.G D'yachkov, V.V. Rykov, "A survey of Superimposed Code Theory," *Problems of control and Inform. Theory*, vol. 12, no. 4, pp. 229-242, 1983.