Fifteenth International Workshop on Algebraic and Combinatorial Coding Theory
June 18-24, 2016, Albena, Bulgaria *pp. 145–150*

On a Hypergraph Approach to Multistage Group Testing Problems [1]

A. G. D'yachkov agd-msu@yandex.ru
Lomonosov Moscow State University, Moscow, Russia
I.V. Vorobyev vorobyev.i.v@yandex.ru
Lomonosov Moscow State University, Moscow, Russia
N.A. Polyanskii nikitapolyansky@gmail.com
Lomonosov Moscow State University, Moscow, Russia
V.Yu. Shchukin vpike@mail.ru
Lomonosov Moscow State University, Moscow, Russia

Abstract. Group testing is a well known search problem that consists in detecting up to s defective elements of the set $[t] = \{1, \ldots, t\}$ by carrying out tests on properly chosen subsets of $[t]$. In classical group testing the goal is to find all defective elements by using the minimal possible number of tests. In this paper we consider multistage group testing. We propose a general idea how to use a hypergraph approach to searching defects. For the case $s = 2$, we design an explicit construction, which makes use of $2\log_2 t(1 + o(1))$ tests in the worst case and consists of 4 stages.

1 Introduction

Group testing is a very natural combinatorial problem that consists in detecting up to s defective elements of the set of objects $[t] = \{1, \ldots, t\}$ by carrying out tests on properly chosen subsets (pools) of $[t]$. The test outcome is positive if the tested pool contains one or more defective elements; otherwise, it is negative.

There are two general types of algorithms. In *adaptive* group testing, at each step the algorithm decides which group to test by observing the responses of the previous tests. In *non-adaptive* algorithm, all tests are carried out in parallel. There is a compromise algorithm between these two types, which is called a *multistage* algorithm. For the multistage algorithm all tests are divided into p sequential stages. The tests inside the same stage are performed simultaneously. The tests of the next stages may depend on the responses of the previous. In this context, a non-adaptive group testing algorithm is reffered to as a one stage algorithm.

[1]The research is supported in part by the Russian Foundation for Basic Research under Grant No. 16-01-00440.

1.1 Previous results

We refer the reader to the monograph [1] for a survey on group testing and its applications. In spite of the fact that the problem of estimating the minimum *average* (the set of defects is chosen randomly) number of tests has been investigated in many papers (for instance, see [2, 3]), in the given paper we concentrate our attention only on the minimal number of test in the *worst case*.

In 1982 [4], Dyachkov and Rykov proved that at least

$$\frac{s^2}{2\log_2(e(s+1)/2)}\log_2 t(1+o(1))$$

tests are needed for non-adaptive group testing algorithm.

If the number of stages is 2, then it was proved that $O(s\log_2 t)$ tests are already sufficient. It was shown by studying random coding bound for disjunctive list-decoding codes [6, 7] and selectors [8]. The recent work [5] has significantly improved the constant factor in the main term of number of tests for two stage group testing procedures. In particular, if $s \to \infty$, then

$$\frac{se}{\log_2 e}\log_2 t(1+o(1))$$

tests are enough for two stage group testing.

As for adaptive strategies, there exist such ones that attain the information theory lower bound $s\log_2 t(1+o(1))$. However, for $s > 1$ the number of stages in well-known optimal strategies is a function of t, and grows to infinity as $t \to \infty$.

1.2 Summary of the results

In the given article we present some explicit algorithms, in which we make a restriction on the number of stages. It will be a function of s. We briefly give necessary notations in section 2. Then, in section 3, we present a general idea of searching defects using a hypergraph approach. In section 4, we describe a 4-stage group testing strategy, which detects 2 defects and uses the asymptotically optimal number of tests $2\log_2 t(1+o(1))$. As far as we know the best result for such a problem was obtained [9] by Damashke et al. in 2013. They provide an exact two stage group testing strategy and use $2.5\log_2 t$ tests. For other constructions for the case of 2 defects, we refer to [10, 11].

2 Preliminaries

Throughout the paper we use t, s, p for the number of elements, defectives, and stages, respectively. Let $\triangleq$ denote the equality by definition, $|A|$ – the cardinality of the set A. The binary entropy function $h(x)$ is defined as usual

$$h(x) = -x\log_2(x) - (1-x)\log_2(1-x).$$

A binary $(N \times t)$-matrix with N rows $\boldsymbol{x}_1, \dots, \boldsymbol{x}_N$ and t columns $\boldsymbol{x}(1), \dots, \boldsymbol{x}(t)$ (codewords)

$$X = \|x_i(j)\|, \quad x_i(j) = 0, 1, \quad i \in [N],\, j \in [t]$$

is called a *binary code of length* N *and size* t. The number of 1's in the codeword $x(j)$, i.e., $|\boldsymbol{x}(j)| \triangleq \sum\limits_{i=1}^{N} x_i(j) = wN$, is called the *weight* of $\boldsymbol{x}(j)$, $j \in [t]$ and parameter w, $0 < w < 1$, is the *relative weight*.

One can see that the binary code X can be associated with N tests. A column $\boldsymbol{x}(j)$ corresponds to the j-th sample; a row $\boldsymbol{x}_i$ corresponds to the i-th test. Let $\mathbf{u} \bigvee \mathbf{v}$ denote the disjunctive sum of binary columns $\mathbf{u}, \mathbf{v} \in \{0,1\}^N$. For any subset $\mathcal{S} \subset [t]$ define the binary vector

$$r(X, \mathcal{S}) = \bigvee_{j \in \mathcal{S}} \boldsymbol{x}(j),$$

which later will be called the *outcome vector*.

By $\mathcal{S}_{un}$, $|\mathcal{S}_{un}| \leqslant s$, denote an unknown set of defects. Suppose there is a p-stage group testing strategy $\mathfrak{S}$ which finds up to s defects. It means that for any $\mathcal{S}_{un} \subset [t]$, $|\mathcal{S}_{un}| \leqslant s$, according to $\mathfrak{S}$:

1. we are given with a code X_1 assigned for the first stage of group testing;

2. we can design a code X_{i+1} for the i-th stage of group testing, based on the outcome vectors of the previous stages $r(X_1, \mathcal{S}_{un})$, $r(X_2, \mathcal{S}_{un})$, $\dots$, $r(X_i, \mathcal{S}_{un})$;

3. we can identify all defects $\mathcal{S}_{un}$ using $r(X_1, \mathcal{S}_{un}), r(X_2, \mathcal{S}_{un}), \dots, r(X_p, \mathcal{S}_{un})$.

Let N_i be the number of test used on the i-th stage and

$$N_T(\mathfrak{S}) = \sum_{i=1}^{p} N_i$$

be the maximal total number of tests used for the strategy $\mathfrak{S}$. We define $N_p(t, s)$ to be the minimal worst-case total number of tests needed for group testing for t elements, up to s defectives, and at most p stages.

3 Hypergraph approach to searching defects

Let us introduce a hypergraph approach to searching defects. Suppose a set of vertices V is associated with the set of samples $[t]$, i.e. $V = \{1, 2 \ldots, t\}$.

First stage: Let X_1 be the code corresponding to the first stage of group testing. For the outcome vector $r = r(X_1, \mathcal{S}_{un})$ let $E(r, s)$ be the set of subsets of $\mathcal{S} \subset V$ of size at most s such that $r(X, \mathcal{S}) = r(X, \mathcal{S}_{un})$. So, the pair $(V, E(r, s))$ forms the hypergraph $H = H(X_1)$. We will call two vertices *adjacent* if they are included in some hyperedge of H. Suppose there exist a *good* vertex coloring of H in k colours, i.e., assignment of colours to vertices of H such that no two adjacent vertices share the same colour. By $V_i \subset V$, $1 \leqslant i \leqslant k$, denote vertices corresponding to the i-th colour. One can see that all these sets are pairwise disjoint.

Second stage:

Now we can perform k tests to check which of monochromatic sets V_i contain a defect. Here we find the cardinality of set $\mathcal{S}_{un}$ and $|\mathcal{S}_{un}|$ sets $\{V_{i_1}, \ldots, V_{i_{|\mathcal{S}_{un}|}}\}$, each of which contains exactly one defective element.

Third stage:

Carrying out $\lceil \log_2 |V_{i_1}| \rceil$ tests we can find a vertex v, corresponding to the defect, in the suspicious set V_{i_1}. Observe that actually by performing $\sum\limits_{j=1}^{\mathcal{S}_{un}} \lceil \log_2 |V_{i_j}| \rceil$ tests we could identify all defects $\mathcal{S}_{un}$ on this stage.

Fourth stage:

Consider all hyperedges $e \in E(r, s)$, such that e contains the found vertex v and consists of vertices of $v \cup V_{i_2} \cup \ldots \cup V_{i_{|\mathcal{S}_{un}|}}$. At this stage we know that the unknown set of defects coincides with one of this hyperedges. To check if the hyperedge e is the set of defects we need to test the set $[t] \backslash e$. Hence, the number of test at fourth stage is equal to degree of the vertex v.

4 Optimal searching of 2 defects

Now we consider a specific construction of 4-stage group testing. Then we upper bound number of tests N_i at each stage.

First stage:

Let $C = \{0, 1, \ldots q-1\}^{\hat{N}}$ be the q-ary code, consisting of all q-ary words of length $\hat{N}$ and having size $t = q^{\hat{N}}$. Let D be the set of all binary words with length N' such that the weight of each codeword is fixed and equals wN', $0 < w < 1$, and the size of D is at least q, i.e., $q \leqslant \binom{N'}{wN'}$. On the first stage we use the concatenated binary code X_1 of length $N_1 = \hat{N} \cdot N'$ and size $t = q^{\hat{N}}$, where the inner code is D, and the outer code is C. We will say X_1 consists of $\hat{N}$ layers. Observe that we can split up the outcome vector $r(X_1, \mathcal{S}_{un})$ into $\hat{N}$ subvectors of lengths N'. So let $r_j(X_1, \mathcal{S}_{un})$ correspond to $r(X_1, \mathcal{S}_{un})$ restricted to the j-th layer. Let w_j, $j \in [\hat{N}]$, be the relative weight of $r_j(X_1, \mathcal{S}_{un})$, i.e., $|r_j(X_1, \mathcal{S}_{un})| = w_j N'$ is the weight of the j-th subvector of $r(X_1, \mathcal{S}_{un})$.

If $w_j = w$ for all $j \in [\hat{N}]$, then we can say that $\mathcal{S}_{un}$ consists of 1 element and easily find it.

If there are at least two defects, then suppose for simplicity that $\mathcal{S}_{un} = \{1, 2\}$. The two corresponding codewords of C are c_1 and c_2. There exists a coordinate $i, 1 \leqslant i \leqslant \hat{N}$, in which they differs, i.e., $c_1(i) \neq c_2(i)$. Notice that the relative weight w_i is bigger than w.

For any $i \in [\hat{N}]$ such that $w_i > w$, we can colour all vertices V in q colours, where the colour of j-th vertex is determined by the corresponding q-nary symbol $c_i(j)$ of code C.

One can check that such a coloring is a good vertex coloring.

Second stage:

We perform q tests to find which coloured group contain 1 defect.

Third stage:

Let us upper bound the size $\hat{t}$ of one of such suspicious group:

$$\hat{t} \leqslant \binom{w_1 N'}{wN'} \cdot \ldots \cdot \binom{w_{\hat{N}} N'}{wN'}.$$

In order to find one defect in the group we may perform $\lceil \log_2 \hat{t} \rceil$ tests.

Fourth stage:

On the final step, we have to bound the degree of the found vertex $v \in V$ in the graph. The degree $\deg(v)$ is bounded as

$$\deg(v) \leqslant \binom{wN'}{(2w - w_1)N'} \cdot \ldots \cdot \binom{wN'}{(2w - w_{\hat{N}})N'}.$$

We know that the second defect corresponds to one of the adjacent to v vertices. Therefore, to identify it we have to make $\lceil \log_2 \deg(v) \rceil$ tests.

The optimal choice of the parameter w gives the procedure with total number of tests equals $2 \log_2 t(1 + o(1))$.

References

[1] *Du D.Z.*, *Hwang F.K.*, Combinatorial Group Testing and Its Applications, 2nd ed., *Series on Applied Mathematics*, vol. 12, 2000.

[2] *Damaschke P.*, *Sheikh Muhammad A.*, *Triesch E.*, Two new perspectives on multi-stage group testing, *Algorithmica*, vol. 67, no. 3, pp. 324-354, 2013.

[3] *M?zard M.*, *Toninelli, C.*, Group testing with random pools: Optimal two-stage algorithms, *Information Theory, IEEE Transactions on*, vol. 57, no. 3, pp. 1736-1745, (2011).

[4] *D'yachkov A.G.*, *Rykov V.V.*, Bounds on the Length of Disjunctive Codes, // *Problems of Information Transmission*, vol. 18. no 3. pp. 166-171, 1982.

[5] *D'yachkov A.G.*, *Vorobyev I.V.*, *Polyanskii N.A.*, *Shchukin V.Yu.*, Bounds on the Rate of Disjunctive Codes, *Problems of Information Transmission*, vol. 50, no. 1, pp. 27-56, 2014.

[6] *Rashad A.M.*, Random Coding Bounds on the Rate for List-Decoding Superimposed Codes. *Problems of Control and Inform. Theory.*, vol. 19, no 2, pp. 141-149, 1990.

[7] *D'yachkov A.G.*, Lectures on Designing Screening Experiments, *Lecture Note Series 10*, Combinatorial and Computational Mathematics Center, Pohang University of Science and Technology (POSTECH), Korea Republic, Feb. 2003, (survey, 112 pages).

[8] *De Bonis A.*, *Gasieniec L.*, *Vaccaro U.*, Optimal two-stage algorithms for group testing problems, *SIAM J. Comp.*, vol. 34, no. 5 pp. 1253-1270, 2005.

[9] *Damaschke P.*, *Sheikh Muhammad A.*, *Wiener G.* Strict group testing and the set basis problem. *Journal of Combinatorial Theory, Series A*, vol. 126, pp. 70-91, August 2014.

[10] *Macula A.J.*, *Reuter G.R.*, Simplified searching for two defects, *Journal of statistical planning and inference*, vol. 66, no. 1, pp 77-82, 1998.

[11] *Deppe C.*, *Lebedev V.S.*, Group testing problem with two defects, *Problems of Information Transmission*, vol. 49, no. 4, pp. 375-381, 2013.