Optimal non-adaptive two-dimensional group testing with equal group size

Tifaout Almeftah[1], Diego Cattaruzza[2], Martine Labbé[3,1], Frédéric Semet[2]

[1] Inria de l'Université de Lille, INOCS, F-59000 Lille, France
tifaout.almeftah@inria.fr

[2] Université de Lille, CNRS, Centrale Lille, Inria, UMR 9189 - CRIStAL, F-59000 Lille, France
{diego.cattaruzza,frederic.semet}@centralelille.fr

[3] Department of Computer Science, Université libre de Bruxelles, 1050 Brussels, Belgium
martine.labbe@ulb.be

Keywords : *Group testing, Misclassification error, Non-adaptive two-dimensional, Column-dependent-row generation, Branch-and-Price-and-Cut.*

1 Introduction and Motivation

In settings where a large-scale population of individuals has low prevalence to have a binary characteristic (e.g. disease, product defect, system error ...), testing the population individually is costly and may not be a viable strategy. Dorfman [3] proposed the concept of group testing, in which a population is instead tested in groups. If a test on a group is negative, the entire group is classified as negative. If not, at least one individual is positive, and the individuals are tested again either individually or in smaller groups until all the population is classified into positive and negative individuals.

The two fundamental problems of group testing are the determination of the group size and the group composition in order to minimize the classification error. To this end one needs to take into consideration several aspects. A test can lead to incorrect classification, that is, a true negative individual as positive and a true positive individual as negative. The population to be screened is usually heterogeneous, i.e., each individual has a different probability of testing positive. Moreover, large size groups may worsen the ability to detect a positive individual among negative individuals. Finally, a sample can belong to a limited number of tests.

Non-adaptive two-dimensional designs where individuals are assigned to a matrix were proposed in [5]. The individuals in the same row or column represent a group. In settings with perfect tests, the suspicious positive individuals would only appear in the intersection of positive rows and positive columns. Works studying the optimal solution of a two-dimensional group testing considering realistic settings are only reported in [2]. However, they concern solely adaptive design where the groups for the following stages are determined by the results of previous stages. To the best of our knowledge, no work on exact algorithms for non-adaptive two-dimensional group testing with heterogeneous population has been reported.

2 Problem Definition

The problem studied is formulated as follows : let $\mathcal{I} = \{\mathcal{I}_1, \mathcal{I}_2, \ldots, \mathcal{I}_N\}$ a population of N individuals with known positivity risks $\mathbf{p} = \{p^1, p^2, \ldots, p^N\}$. Let us assume we have an imperfect test with a given sensitivity and a specificity. The studied two-dimensional group testing problem aims at determine groups of exactly L individuals so as to minimize a convex combination of the classification error and the expected number of tests and such that:

(i) Each individual $\mathcal{I}_i$ belongs to exactly two distinct groups S_k and S_l;

(ii) Each two groups S_k and S_l share at most one individual $\mathcal{I}_i$.

One important question we raise concerns the values of eligible group size L for which the two constraints ((i)-(ii)) are satisfied. The following proposition holds.

Proposition 1 *For $N > 2$, the non-adaptive two-dimensional group testing problem has at least one feasible solution if on only if the value of group size L is chosen such that $L|2N$ and $2 \leq L \leq \frac{4N}{1+\sqrt{1+8N}}$.*

3 Solution Approach

We propose an integer linear model to formulate the group testing problem where variables are binary and exponential number. In particular there is one variable for each possible group of size L of the individuals. Since the number of variables is exponential, we generate them on the fly with a pricing procedure. However, the generation of new variables implies the generation of new constraints of the model, which make the resolution of the linear relaxation of the model a column-and-row generation procedure [4]. We embed the procedure in a branch-and-bound algorithm to recover integrality. To strengthen the linear relaxation we consider inequalities obtained using the Reformulation-Linearization Technique (RLT, [6]). To recover integrality we develop an ad-hoc branching scheme that takes into account the problem structure.

In order to validate the proposed method, experiments are conducted on random generated instances with up to 100 individuals.

4 Conclusions and Perspectives

We have proposed a generic model handling non-adaptive two-dimensional group testing with equal group size and with different risk profile population. We give the values of eligible group sizes satisfying our design. Finally, we propose a Branch-and-Price-and-Cut algorithm to solve the problem. First, we plan to expand our experiments to validate the efficiency of our Branch-and-Price-and-Cut algorithm on large instances of real SARS-CoV-2 data. Second, we plan to expand our current work in order to consider potentially variable group sizes. Specifically, inspired by the work of [1], we are interested in studying the potential of a hybrid model in which an individual may belong to either one or two groups.

References

[1] Hrayer Aprahamian, Douglas R Bish, and Ebru K Bish. Optimal risk-based group testing. *Management Science*, 65(9):4365–4384, 2019.

[2] Hrayer Aprahamian, Ebru K Bish, and Douglas R Bish. Adaptive risk-based pooling in public health screening. *IISE Transactions*, 50(9):753–766, 2018.

[3] Robert Dorfman. The detection of defective members of large populations. *The Annals of Mathematical Statistics*, 14(4):436–440, 1943.

[4] Ibrahim Muter, Ş İlker Birbil, and Kerem Bülbül. Simultaneous column-and-row generation for large-scale linear programs with column-dependent-rows. *Mathematical Programming*, 142(1):47–82, 2013.

[5] RM Phatarfod and Aidan Sudbury. The use of a square array scheme in blood testing. *Statistics in medicine*, 13(22):2337–2343, 1994.

[6] H.D. Sherali and W.P. Adams. A reformulation-linearization technique (rlt) for semi-infinite and convex programs under mixed 0-1 and general discrete restrictions. *Discrete Applied Mathematics*, 17(6):1319–1333, 2009.