Journal of Combinatorial Optimization, 7, 381–384, 2003

A Note on Decoding of Superimposed Codes

TAYUAN HUANG
CHIH-WEN WENG weng@math.netu.edu.tw
Department of Applied Mathematics, National Chiao Tung University, Taiwan, Republic of China

Received September 25, 2003; Accepted November 22, 2003

Abstract. A superimposed code with general distance D can be used to construct a non-adaptive pooling design. It can then be used to identify a few unknown positives from a large set of items by associating naturally an outcome vector u. A simple method for decoding the outcome vector u is given whenever there are at most $\lfloor \frac{D-1}{2} \rfloor$ errors occuring in the outcome vector u. Moreover, another simple method of detecting whether there is any error occuring in the outcome vector u is also given whenever there are at most $D - 1$ errors in u. Our method is a generalization of the classical result of Kautz and Singleton (Nonadaptive binary superimposed codes, *IEEE Trans. Inform. Theory*, vol. 10, pp. 363–377, 1964).

Keywords: superimposed codes

1. Introduction

The notion of superimposed codes was first introduced by Kautz and Singleton (1964) with distance 1 in 1964, and then by D'yachkov et al. (1989) for general distance around 1989. In addition to some applications found in Bassalygo and Pinsker (1983), superimposed codes have become a dominating tool in a recent study of non-adaptive group testings, and have attracted more attentions nowadays due to its recent application to pooling designs in DNA mapping, (see Du and Hwang, 2000 for more details). A uniform way of constructing a class of superimposed codes with distance 1 was given by Macula (1996). Two families of superimposed codes with general distance were found by Ngo and Du (2002). It was soon generalized over a class of ranked posets, called pooling spaces, by Huang and Weng (submitted) to find the superimposed codes with general distance.

A superimposed code with general distance D can be used to construct a non-adaptive pooling design. It can then be used to identify a few positive items from a large set of items by associating naturally an outcome vector u. The purpose of this article is to give a simple method for decoding the outcome vector u to identify those positives correctly whenever there are at most $\lfloor \frac{D-1}{2} \rfloor$ errors occurring in the outcome vector u. Moreover, another simple method of detecting whether there is any error occurring in the outcome vector u is also given whenever there are at most $D - 1$ errors in u. Our method is a generalization of the classical result of Kautz and Singleton (1964).

2. Preliminaries

For a positive integer m, set $[m] := \{1, 2, \ldots, m\}$. Fix four positive integers t, n, D, d with $D \leq t$ and $d \leq n$. A *superimposed code M with length t, volume n, distance D and strength d* is a family $M = \{C_1, C_2, \ldots, C_n\}$ of n subsets of $[t]$ such that for any index subset $S \subseteq [n]$ with $|S| \leq d$ and any $i \in [n]\backslash S$,

$$\left| C_i - \bigcup_{j \in S} C_j \right| \geq D. \tag{2.1}$$

The $t \times n$ incidence matrix of a superimposed code M with length t, volume n, distance D and strength d is called the *(d, e)-disjunct matrix* (or *d^e-disjunct matrix*) of size $t \times n$ where $e = D - 1$, and if $D = 1$ it is called a *d-disjunct matrix* (D'yachkov et al., 1989; Huang and Weng, submitted).

Throughout the note $M = \{C_1, C_2, \ldots, C_n\}$ is a superimposed code with length t, volume n, distance D and strength d. M can be used to construct a non-adaptive group testing design on n items by associating the set $[n]$ with the set of items and the set $[t]$ with the set of tests. If $i \in C_j$ then item j is contained in test i. By a *set of positives* we mean a subset $S \subseteq [n]$ such that $|S| \leq d$. Let S be a set of positives. The *ideal output $o(S)$* of S in M is defined by

$$o(S) := \bigcup_{j \in S} C_j, \tag{2.2}$$

and the *test result* (or *outcome vector*) u of S under M is any subset of $[t]$. The *number of test errors* in the test result u of S under M is the Hamming distance $\partial(u, o(S))$, where

$$\partial(u, o(S)) := |u - o(S)| + |o(S) - u|.$$

Suppose the test result u of S under M does not contain any error, or equivalently $u = o(S)$. Kautz and Singleton showed the set S of positives can be determined by the test result u (Kautz and Singleton, 1964). In next section we will generalize their result to allow the test result u containing a few errors.

3. The decoding method

The methods in decoding and in error detecting of a test result are given in this section. We need a lemma first.

Lemma 3.1. *Let $M = \{C_1, C_2, \ldots, C_n\}$ denote a superimposed code with length t, volume n, distance D and strength d. Let $S, T \subseteq [n]$ be two distinct subsets with each at most d elements. Then the Hamming distance of the ideal outputs $o(S)$, $o(T)$ of S, T respectively under M is at least D.*

Proof: At least one of $S - T$, $T - S$ is nonempty, so assume $S - T \neq \emptyset$. Pick $i \in S - T$. By construction

$$\left| C_i - \bigcup_{j \in T} C_j \right| \geq D.$$

Referring to notation in (2.2), we find $\partial(o(S), o(T)) \geq D$. This proves the lemma. □

The following theorem is the main idea.

Theorem 3.2. *Let $M = \{C_1, C_2, \ldots, C_n\}$ denote a superimposed code with length t, volume n, distance D and strength d. Suppose $S \subseteq [n]$ with $|S| \leq d$ and $u \subseteq [t]$. Set*

$$T = \left\{ j \mid C_j - u| \leq \left\lfloor \frac{D-1}{2} \right\rfloor \right\}. \tag{3.1}$$

Then the following (i)–(ii) *hold.*
(i) *Suppose $\partial(o(S), u) \leq \lfloor \frac{D-1}{2} \rfloor$. Then $T = S$.*
(ii) *Suppose $\partial(o(S), u) \leq D - 1$ and $|T| \leq d$. Then $o(S) = u$ if and only if $o(T) = u$.*

Proof: (i) ($\supseteq$) Pick $j \in S$. Then $C_j \subseteq o(S)$ by (2.2). Hence

$$\begin{aligned} |C_j - u| &\leq |o(S) - u| \\ &\leq \partial(o(S), u) \\ &\leq \left\lfloor \frac{D-1}{2} \right\rfloor. \end{aligned}$$

Thus $j \in T$ by (3.1).
($\subseteq$) Pick $j \in T$. Suppose $j \notin S$. By the construction of M, there are at least D elements in $C_j - o(S)$. Since $\partial(o(S), u) \leq \lfloor \frac{D-1}{2} \rfloor$, there are at least

$$D - \left\lfloor \frac{D-1}{2} \right\rfloor = \left\lfloor \frac{D-1}{2} \right\rfloor + 1$$

elements in $C_j - u$, a contradiction to (3.1).
(ii) This is clear if $S = T$. Suppose $S \neq T$. Then $\partial(o(S), u) > \lfloor \frac{D-1}{2} \rfloor$ by (i). In particular, $o(S) \neq u$. Applying triangular inequality and using Lemma 3.1 we find

$$\begin{aligned} \partial(o(T), u) &\geq \partial(o(T), o(S)) - \partial(o(S), u) \qquad (3.2) \\ &\geq D - (D - 1) \qquad (3.3) \\ &= 1. \qquad (3.4) \end{aligned}$$

Hence $o(T) \neq u$. □

Remark 3.3. The special case D = 1 in Theorem 3.2 is Kautz and Singleton's result in 1964 (Kautz and Singleton, 1964).

A decoding algorithm

Suppose $[n]$ is the set of items and $S \subseteq [n]$ with $|S| \leq d$ is the set of positives to be identified. A superimposed code M with length t, volume n, distance D and strength d for some positive integers t, D is on hand. Let u be the test result of S under M, and $o(S)$ be the unknown ideal output of S. The Hamming distance $\partial(u, o(S))$ is simplify the number of test errors occuring in the testing procedure. Then do the following:

(i) Determine T first by (3.1) and then determine $o(T)$ by (2.2);
(ii) Suppose there are at most $\lfloor \frac{D-1}{2} \rfloor$ errors in the outcome vector u. Then $T = S$ is concluded by applying Theorem 3.2 (i);
(iii) Check where there is an error in the outcome vector u by applying Theorem 3.2 (ii) whenever there are at most $D - 1$ errors in u;
(iv) If $|T| > d$, there is an error; otherwise an error occurs in u if and only if $o(T) \neq u$.

Acknowledgment

The authors thank an anonymous referee of the preprint (Huang and Weng, submitted) for enlightening the line of study and many valuable ideas.

References

L.A. Bassalygo and M.S. Pinsker, "Limited multiple-access of a non-synchronous channel," *Promlemy Peredachi Informatsii*, vol. 19, no. 8, pp. 92–96, 1983 (in Russian).
D.Z. Du and F.K. Hwang, *Combinatorial Group Testing and its Applications.* Series on Applied Mathematics World Scientific: River Edge, NJ, vol. 12, 2000.
A.G. D'yachkov, A.J. Macula, and P.A. Vilenkin, "Nonadaptive group testing with error-correction d^e-disjunct inclusion matrices," preprint.
A.G. D'yachkov, V.V. Rykov, and A.M. Rashad, "Superimposed distance codes," *Prob. of Control and Inform. Theory*, vol. 18, pp. 237–250, 1989.
T. Huang and C. Weng, "Pooling spaces and non-adaptive pooling designs," *Discrete Math.*, submitted.
W.H. Kautz and R.C. Singleton, "Nonadaptive binary superimposed codes," *IEEE Trans. Inform. Theory*, vol. 10, pp. 363–377, 1964.
A.J. Macula, "A simple construction of d-disjunct matrices with certain constant weights," *Discrete Math.*, vol. 162, pp. 311–312, 1996.
H. Ngo and D. Du, "New constructions of non-adaptive and error-tolerance pooling designs," *Discrete Math.*, vol. 243, pp. 161–170, 2002.