COMPSCI 690T Coding Theory and Applications Mar 20, 2017

Lecture 13

Instructor: Arya Mazumdar *Scribe: Arya Mazumdar*

1 Concatenated Codes

A concatenated code comprises of two codes. One q-ary outer code, and a binary inner code.

The outer (linear) code $\mathcal{C}_o$ is of length n_o, dimension k_o and minimum distance d_o. The inner code $\mathcal{C}_i$ is of length n_i, dimension k_i and minimum distance d_i. The codes are chosen such that $q = |\mathcal{C}_i| = 2^{k_i}$.

The concatenated code is a binary linear code and is defined as follows. There is a fixed bijective map $f : F_q \to \mathcal{C}_i$. Every symbol of every codeword of the outer code is replaced by a binary sequence according to this map. As a result, we have $|\mathcal{C}_o|$ binary vectors of length $n_i n_o$ each. These constitute the concatenated code $\mathcal{C}$.

The length of $\mathcal{C}$ is $n = n_i n_o$. The size of the code is $|\mathcal{C}_o| = q^{k_o} = 2^{k_i k_o}$. That means the dimension of the code is $k = k_o k_i$. The minimum distance of the code is at least $d = d_o d_i$. To see this, note that the distance of the outer code is d_o, and in each of these d_o positions, upon applying the map f, there will be a disagreement of d_i bits.

1.1 Zyablov bound

We show the best possible rate-distance trade-off achievable with binary codes that are explicitly (deterministically) constructible.

Consider the outer code to be a q-ary Reed-Solomon code of length $n_o = q$, and the inner code to be a binary linear code that achieves the Gilbert-Varshamov (GV) bound. We must have $n_i = k_i/R_i = \log_2 q/R_i$, where R_i is the rate of the inner code.

We have, $n = n_o n_i = q \log_q /R_i$. This is how we choose q.

Since the inner code has dimension $\log_2 q < \log_2 n$, it is possible to find a GV code (code that achieves the GV bound) in polynomial time.

We have $k_o = n_o - d_o + 1$ or $R_o = n_o(1 - d_o/n_o + 1/q)$, where R_o is the rate of the outer code. Also $R_i = n_i(1 - h(d_i/n_i))$, since the inner code achieves the GV bound. If $d = \delta n$ be the distance of the overall code then

$$\delta = \frac{d}{n} = \frac{d_i d_o}{n_i n_o}.$$

Therefore, the rate of the overall code is

$$R = R_o R_i \geq (1 - d_o/n_o)(1 - h(d_i/n_i)) = (1 - \delta/x)(1 - h(x)),$$

where x is the relative distance of the inner code. Since we are free to choose that, for the best such code we will have,

$$R \geq \max_{0 \leq x \leq 1} (1 - \delta/x)(1 - h(x)).$$

This is known as the Zyablov bound.

Now we will show an application of concatenated code to something called *Group testing*.

2 Group Testing

Suppose there are n items and among them at most t are defective. We are allowed to *test* the elements for defectiveness in pools or *groups*. For any group of items, a test tells us whether there exists a defective item in the group or not. How many such tests are necessary to find out all the defective items? (We mentioned couple of applications: soldiers with syphilis and finding failed links in a network).

If there is only up to 1 defective, then we may proceed to do a binary search, testing a group of size $\sim n/2$ for the first test, and a group of size $\sim n/4$ in the second and so on. In this way within $\log n$ tests we will be able to identify the defective element.

If we do the above procedure t times, each time removing a found defective element, then we will be able to identify up to t defective elements. This will take a total of $t \log n$ tests.

On the other hand, since each test reveal at most 1 bit of information and there are $\sum i = 0^t \binom{n}{i}$ possibilities for the defective sets, we need at least $\log_2 \sum i = 0^t \binom{n}{i} > t \log_2 \frac{n}{t}$ tests. Therefore the aforementioned adaptive scheme is very close to being optimal.

Hence the main challenge lies in designing a nonadaptive scheme for group testing, where all the tests are simultaneously performed. This also has a lot of applications.

2.1 Nonadaptive testing: a testing matrix

A nonadaptive testing scheme with m tests can be represented using a binary $m \times n$ testing matrix $A = (a_{i,j})$. In this matrix, the (i, j)th entry is 1 if and only if the ith test include the jth item (otherwise it is 0).

What is the property that this matrix must satisfy such that it is possible to recover any up to t defective items?

Let $\boldsymbol{x} \in \{0, 1\}^n$ be the *defective indicator vector*, i.e., $x_i = 1$ if and only if the ith item is defective. Then the testing procedure and the results can be summarized as,

$$A \wedge \boldsymbol{x} = \boldsymbol{y},$$

where $\boldsymbol{y} \in \{0, 1\}^m$ is the results of the tests. The operation $\wedge$ signifies a Boolean OR operation. Indeed, $\boldsymbol{y}$ is the point wise Boolean OR of the columns of A that correspond to the entries of $\boldsymbol{x}$ that are 1. One can also think of this as a matrix-vector multiplication over reals, followed by a quantization operation where anything nonzero is substituted by 1.

This problem is philosophically similar to the compressed sensing problem (which is over reals and such quantization is not done) and the decoding of linear codes via parity check matrix (where Boolean XOR instead of OR is used).

Therefore the aim is to construct a matrix A such that the Boolean OR of any up to t columns of the matrix must be different. Such matrices are called t-separable. In other words, we need a matrix such that union of supports of any up to t columns are different. Given n, t, the aim is to come up with a matrix with minimal number of rows such that this is possible.