Information-Theoretic Modeling

Lecture 4: Noisy Channel Coding

Teemu Roos

Department of Computer Science, University of Helsinki

Fall 2014

UNIVERSITY OF HELSINKI

Lecture 4: Noisy Channel Coding

1. Noisy Channels
    - Reliable communication
    - Error correcting codes
    - Repetition codes
2. Channel Coding and Shannon's 2nd Theorem
    - Channel capacity
    - Codes and rates
    - Channel coding theorem
3. Hamming Codes
    - Parity Check Codes
    - Hamming (7,4)

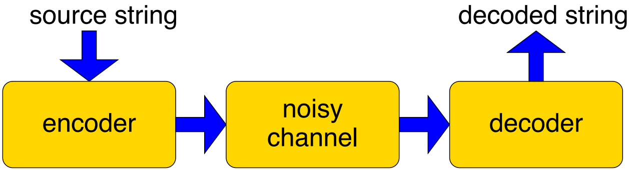
Reliable communication

In practice, most media are not perfect — *noisy channels*:

- Modem line
- Satellite link
- Hard disk

Can we recover the original message (without errors) from a noisy code string?

Error correcting codes



We want to minimize two things:

1. Length of the code string.
2. Probability of error.

Repetition codes

A simple idea: Just repeat the original string many times.

Get it? Get it? Get it? Get it? Get it? Get it? Get it? Ge

Repetition codes

A simple idea: Just repeat the original string many times.

`T R A N S M I S S I O N`

`TTTRRRAAANNNSSSMMMIIISSSSSSIIIOOONNN`

`TTTHRRAAANNBSSSMMMIIISSSSWSPILOOONNG`

`T R A N S M I S S ? O N`

Transmission rate reduced to 1 : 3.

If errors independent and symmetric, probability of error reduced to $3(1-p)p^2 + p^3 \approx 3p^2$, where p is the error rate of the channel.

Channel Capacity: basic intuition

- We are going to define the channel capacity C purely in terms of the probabilistic properties of the channel.
- We consider encoding messages of b bits into codewords of b/R bits, for some rate $0 < R < 1$.
- We say a rate R is achievable using a channel, if there is an encoding such that the probability of error goes to zero as b increases.
- The *Source Coding Theorem*, or *Shannon's Second Theorem*, says rate R is achievable if $R < C$, and not achievable if $R > C$.

Channel Capacity

- Binary symmetric channel (BSC), error rate p:

$$\Pr[y = 1 \mid x = 0] = \Pr[y = 0 \mid x = 1] = p$$

  where x is the transmitted and y the received bit
- We define *channel capacity* as

$$C(p) = 1 - H(p) = 1 - \left[p \log_2 \frac{1}{p} + (1-p) \log_2 \frac{1}{1-p} \right] .$$

- For instance, $C(0.1) \approx 0.53$. Ratio about $1 : 2$.

Channel Capacity

For channels other than BSC, the channel capacity is more generally defined as

$$C = \max_{p_X} I(X,Y) = \max_{p_X} \left(H(Y) - H(Y \mid X)\right)$$

- X is the transmitted and Y the received symbol
- I is calculated with respect to $p_{X,Y}(x,y) = p_X(x)p_{Y|X}(y \mid x)$
- $p_{Y|X}$ is defined by the channed characteristics.

Intuition:

- for a large capacity, we want Y to carry a lot of information
- however, knowing X should remove most of the uncertainty about Y
- we can get a favourable p_X by choosing a suitable coding.

Codes and rates

For simplicity, we consider BSC unless we say otherwise.

- *Messages* we want to send are blocks of b bits.
  Thus, there are $M = 2^b$ possible messages.
- We encode a message into *codewords* of n bits.
  So generally we need $n \geq \log_2 M = b$.
- Notation:
  - $w \in \{1, \ldots, M\}$: (index of) a message
  - $X^n = f(W) \in \{0, 1\}^n$: codeword for message w
  - $Y^n \in \{0, 1\}^n$: received codeword (noisy version of X^n)
  - $g(Y^n) \in \{1, \ldots, M\}$: our guess about what the correct message was.
- The *rate* of the code is $R = (\log_2 M)/n$.

Codes and rates

Let λ_w, for $w \in \{1, \ldots, M\}$, denote the probability that message w was sent but not correctly received.

We can write this as

$$\lambda_w = \sum_{y \,:\, g(y) \neq w} p(y \mid X = f(w)) \ .$$

Average error: $\bar{\lambda} = \frac{1}{M} \sum_w \lambda_w$

Maximum error: $\lambda_{\max} = \max_w \lambda_w$

Channel coding theorem

A rate R is *achievable* if there is a sequence of codes, for increasingly large codeword lengths n, such that as n goes to infinity, the maximum error $\lambda_{\max}$ goes to zero.

Channel Coding Theorem

If $R < C$, where C is the channel capacity, then rate R is achievable.

If $R > C$, then rate R is not achievable.

In other words, for any given $\epsilon > 0$ and $R < C$, for large enough b we can encode messages of b bits into codewords of $n = b/R$ bits so that the probability of error is at most ϵ.

This is also known as Shannon's Second Theorem (the first one being the Source Coding Theorem).

Channel coding theorem

Channel Coding Theorem—So what?

Assume you want to transmit data with probability of error 10^{-15} over a BSC, $p = 0.1$. Using a repetition code, we need to make the message **63** times as long as the source string.
(Exercise: Check the math. Hint: binomial distribution.)

Shannon's result says twice as long is enough.

If you want probability of error 10^{-100}, Shannon's result still says that twice is enough!

Channel coding theorem

- The proof of Channel Coding Theorem (which we won't cover) is based on choosing M codewords, each n bits long, completely at random.
- To decode y, just pick w for which $f(w)$ is closest to y.
- If $\log_2 M < nR$, then the expected error rate, over random choice of code books, is very small. This is the tricky part.
- If random code books are good on average, then surely the best single code book is at least as good.
- However, in practice we need specific codes that have high rates and *are easy to compute*. Finding such is difficult and out of scope for this course. We will next give a simple example to illustrate the basic idea.

Hamming Codes

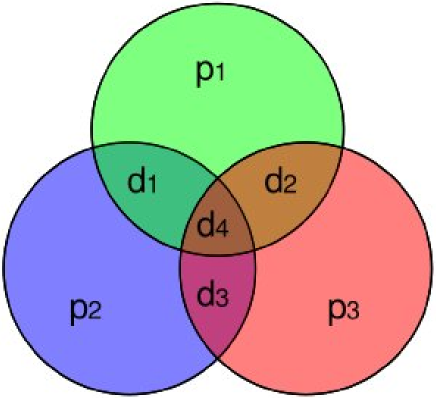
Richard W. Hamming (11.2.1915–7.1.1998)

Parity Check Codes

One way to detect and correct errors is to add *parity checks* to the codewords:

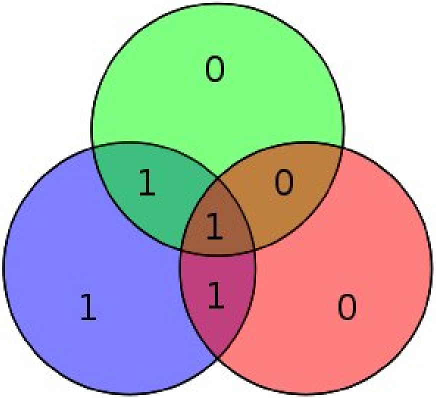
- If we add a parity check bit at the end of each codeword we can detect one (but not more) error per codeword.
- By clever use of more than one parity bits, we can actually identify where the error occurred and thus also *correct errors*.
- Designing ways to add as few parity bits as possible to correct and detect errors is a *really* hard problem.

Hamming (7,4)



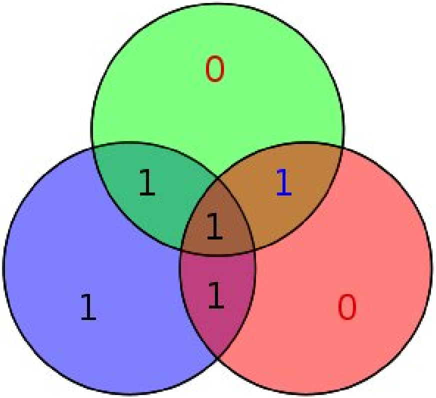
4 data bits (d_1, d_2, d_3, d_4), 3 parity bits (p_1, p_2, p_3)

Hamming (7,4)



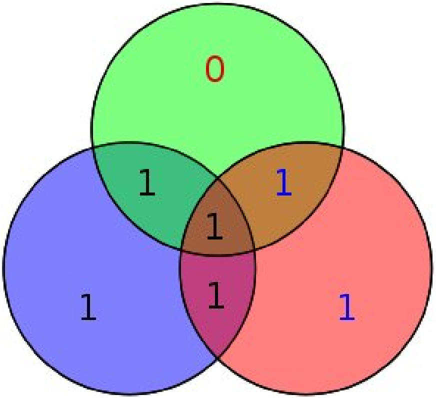
source string 1011, parity bits 010

Hamming (7,4)



error in data bit d_2 $(0 \mapsto 1)$ is identified and corrected

Hamming (7,4)



two errors can be detected but not corrected

Advanced Error Correcting Codes

The Hamming (7,4) code is an example of a code that can detect and correct errors at rate greater than $1 : 2$.

More complex Hamming codes, like Hamming (8,4), Hamming (11,7), etc. can correct and/or detect more errors.

The present state-of-the-art is based on so called *low-density parity-check* (LDPC) codes, which likewise include a number of parity check bits.

Massive research effort: At ISIT-09 conference, 12 sessions (4 talks in each) about LDPC codes.

Next topics

Back to noiseless source coding

- prefix codes and Kraft Inequality
- coding algorithms: Shannon coding, Huffman coding, arithmetic coding