Lecture 15: Hamming codes and Viterbi algorithm

- Hamming codes
- Viterbi algorithm

Why reliable communication is possible?

- After shuffling a deck of cards, dealer hands player-A 5 cards
- player-A randomly picks 1 card, and gives the other 4 cards to player-B
- is it possible for player-A to hint to player-B which cards has been kept, using the four cards given to player-B?

- The channel coding theorem promises the existence of block codes with rate below capacity and arbitrarily small P_e, when block length is large

- Since Shannon's original paper, people have been searching for capacity achieving code

- Goal: capacity achieving, encoding and decoding are simple

Naive idea: repetition code

- Introduce redundancy so if some bits are corrupted, still be able to recover the message

- Repeat bits:

$$1 \to 11111$$

$$0 \to 00000$$

- decoding scheme: majority vote

- error occurs if more than 3 bits are corrupted

- Not efficient:
  rate $= 1/5$ bit per symbol

Quest for capacity-achieving codes ...

- Block codes: map a block of information bits onto a codeword, no dependence on past information bits
  - Hamming codes (1950)
  - simplest, illustrates basic ideas underlying most codes
- Convolutional codes (past 40 years)
  - Each output block depends also on some of the past inputs
- Turbo codes and Low-density-parity-check (LDPC) code (90s)
  - Using iterative message-passing algorithm decoding can achieve channel capacity
- Polar codes
  - A novel channel coding scheme (E. Arikan, 2009)
  - allow a transmission approaching capacity for large block sizes
  - first capacity-achieving codes that can be successively decoded

Hamming code

- Richard Hamming (1915 - 1988)
- Basic idea: combine bits in an intelligent fashion so that each extra bit checks whether there is an error in a subset of information bits

- Detecting odd number of error
  for a block with $n-1$ information bits, add one extra bit so that parity of the entire block is 0 (the number of 1's in the block is even)

- Parity check code:
  if we use multiple parity check bits. Hamming code is one example.

Hamming code example

$$H = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

- the set of vectors of length 7 in the null space $\mathcal{N}(H)$: $Hb = 0$: since H has rank 3, null space of $\mathcal{N}(H)$ has dimension 4, there are $2^4 = 16$ codewords in $\mathcal{N}(H)$

$$\begin{array}{cccc} 0000000 & 0100101 & 1000011 & 1100110 \\ 0001111 & 0101010 & 1001100 & 1101001 \\ 0010110 & 0110011 & 1010101 & 1110000 \\ 0011001 & 0111100 & 1011010 & 1111111 \end{array}$$

0000000	0100101	1000011	1100110
0001111	0101010	1001100	1101001
0010110	0110011	1010101	1110000
0011001	0111100	1011010	1111111

- Property of null space
  - Null space is *linear* : sum of any two codewords is also a codeword
  - Minimum number of 1's in any codeword is 3: "minimum weight" of the code
  - Difference any two codewords has 3 ones
  - Minimum distance ≥ 3: distinguishability of codewords

- **Hamming distance**: number of positions at which corresponding symbols are different

- Idea: use these null space vectors as codewords

$$\begin{bmatrix}0 & 0 & 0 & 1 & 1 & 1 & 1\end{bmatrix}$$

first 4: information bits, last 3: parity check bits

- $(7, 4, 3)$ Hamming code

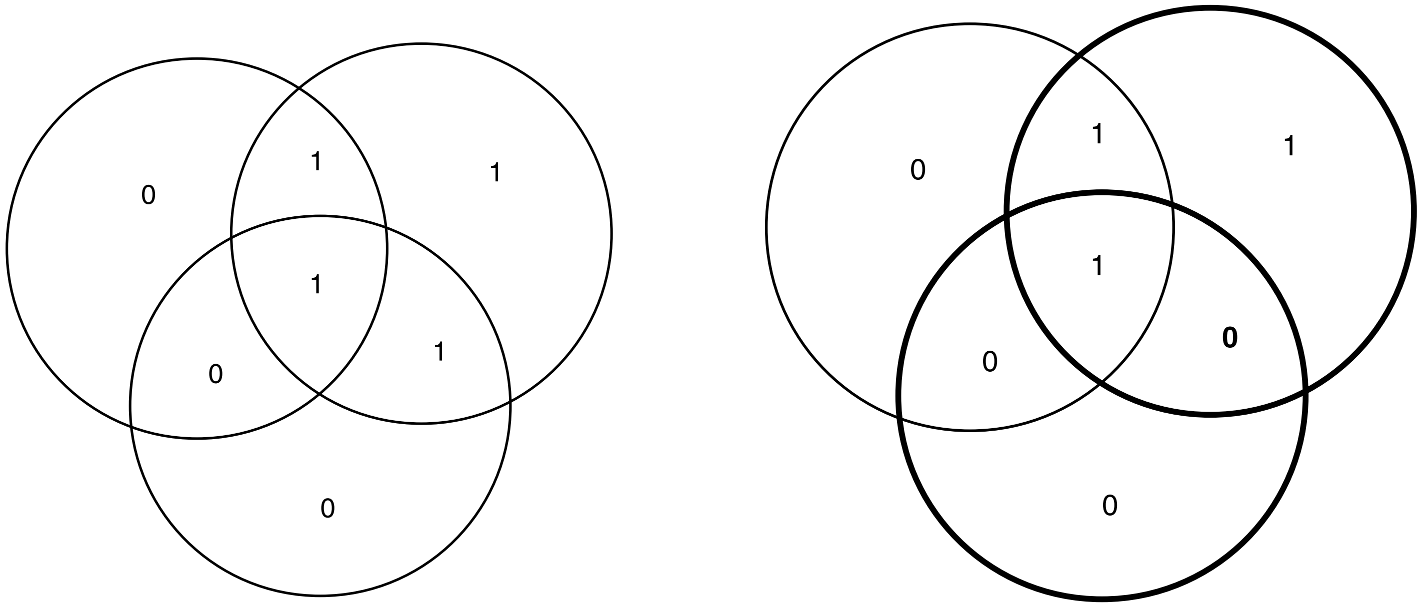
- c: a codeword, is corrupted in only one place, we can detect the location of the error

- if $r = c + e_i$, $e_i = \begin{bmatrix}0 \ldots & 1 \ldots & 0\end{bmatrix}$

$$Hr = H(c + e_i) = Hc + He_i = He_i$$

He_i is the i-th column of H

Venn diagram



- Hamming code can correct one error
- Reed and Solomon code (early 1950s), multiple error-correcting codes
- Bose and Ray-Chaudhuri and Hocquenghem (BCH) (late 1950s) codes, multiple error-correcting codes using Galois field theory
- All compact disc (CD) players include error-correction circuitry using Reed-Solomon codes correct bursts of up to 4000 errors

Viterbi algorithm

- Developed by Andrew Viterbi, 1966
- A version of forward dynamic programming
- Exploit structure of the problem to beat "curse-of-dimensionality"
- Widely used in: wireless and satellite communications, DNA analysis, speech recognition

Detective

- Catch a suspect making transition at RDU airport
- you know during this period 4 domestic flights arrived from 4 cities connected to departing flights to 18 others
- one way to catch the suspect would be search all 18 gates
- alternative: investigate only departing flights with connections to 4 arriving flights

Derivations

- $X_i \in \{1, 2, \cdots, M\}$, size of the alphabet is M
- $X^n = [X_0, X_1, \ldots, X_n]$, n codewords
- $Y^n = [Y_1, \ldots, Y_n]$, received codewords
- Assume codewords form first order Markov chain

$$p(X_0, X_1, \ldots, X_n) = \prod_{i=1}^{n} p(X_i | X_{i-1})$$

- Discrete Memoryless Channel (DMC): $p(Y^n | X^n) = \prod_{i=1}^{n} p(Y_i | X_i)$

- Maximum a-Posterior (MAP) decoding

$$\max_{X^n} p(X^n|Y^n) = \frac{p(Y^n|X^n)p(X^n)}{p(Y^n)}$$

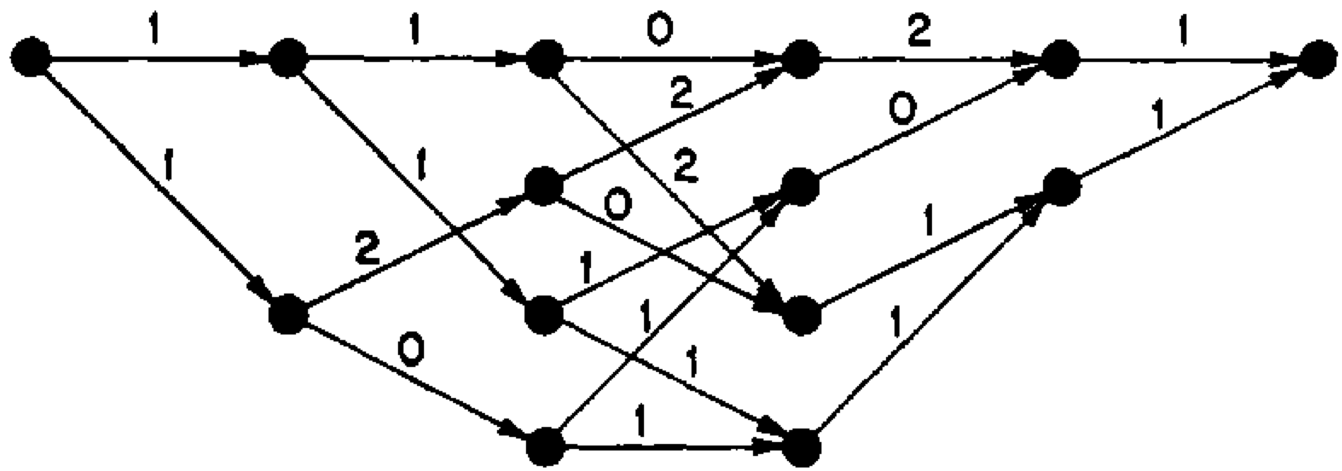
- Using assumptions above

$$\log p(Y^n|X^n)p(X^n) = \sum_{i=1}^{n} [\log p(Y_i|X_i) + \log p(X_i|X_{i-1})]$$

- MAP $=$ finding the shortest path through a graph

- path length $\propto -\log p(Y_i|X_i) - \log p(X_i|X_{i-1})$

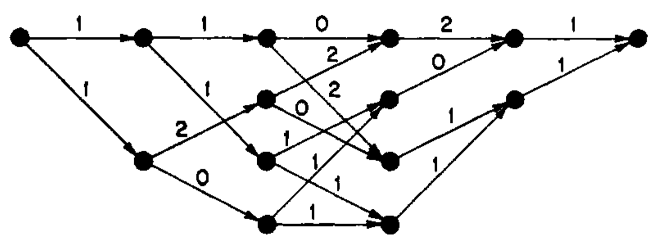
An example Trellis



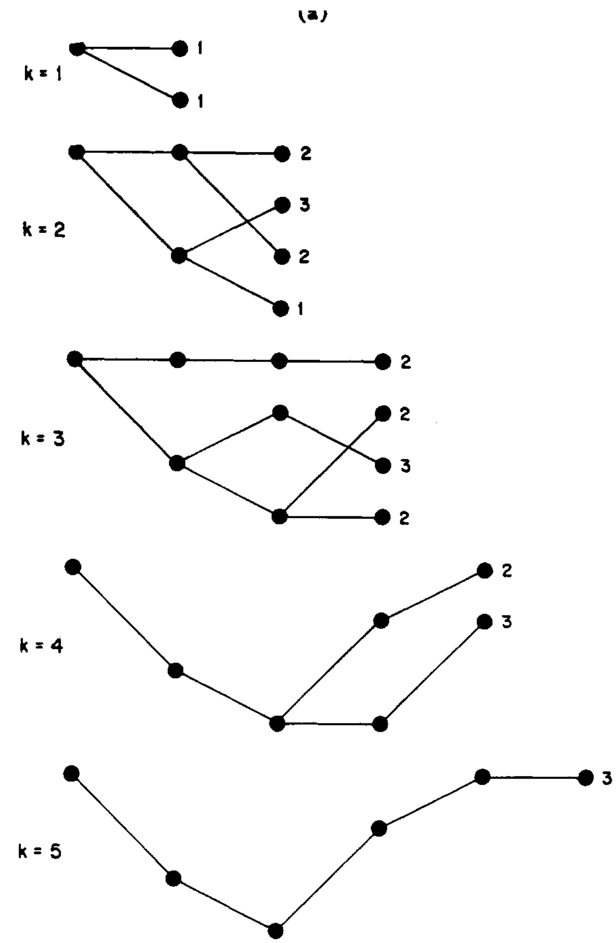
(a)

From "The Viterbi Algorithm", by D. Forney, 1973

- Shortest path segment is called the *survivor* for node c_k
- Important observation: the shortest complete path must begin with one of the survivors
- in this stage, we only need to store M survivor paths
- this greatly reduces storage down to manageable size
- Example: decoding using Viterbi algorithm

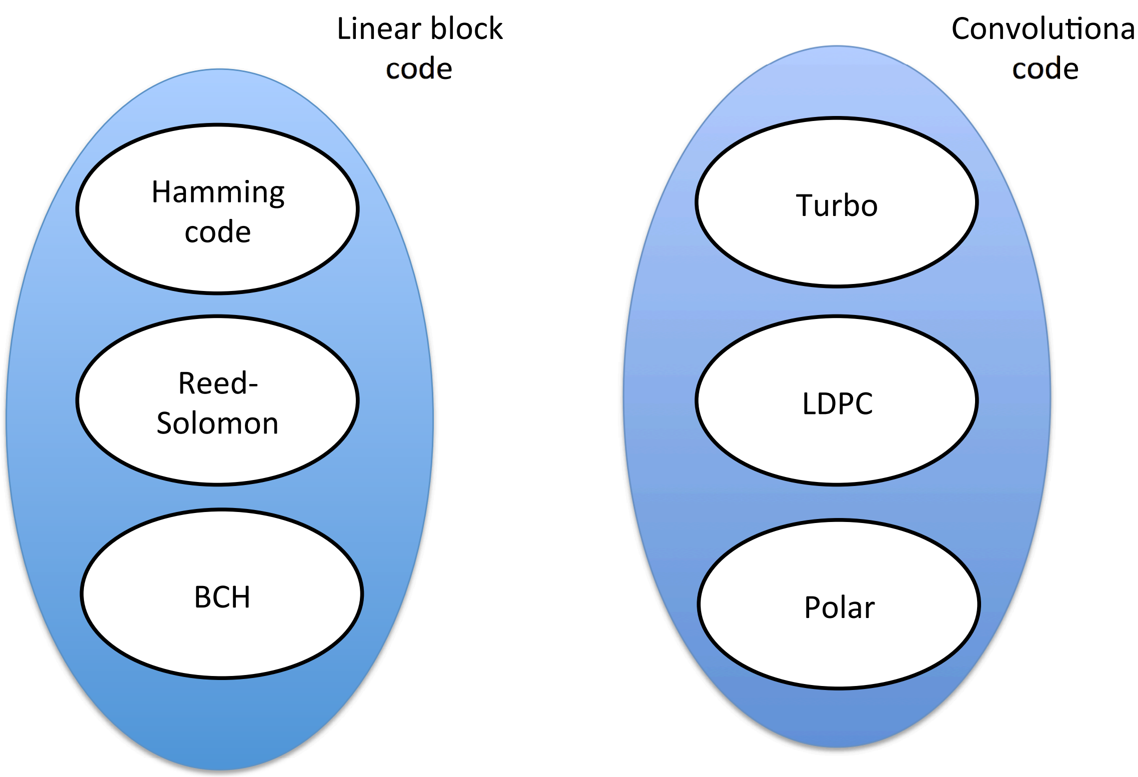


(a)



(b)

Summary



A (partial) diagram