ELE539: Optimization of Communication Systems
Lecture 22: DP Applications:
Bellman Ford Routing and Viterbi Decoding

Professor M. Chiang
Electrical Engineering Department, Princeton University

April 16, 2004
Lecture Outline

• Shortest path routing
• Bellman Ford algorithm
• IP routing
• Hidden Markov model
• Viterbi algorithm
• Convolutional code decoding
**Shortest Path Routing**

Given a directed graph with vertex set \( V \) and edge set \( E \).

Each edge \((i, j)\) has cost or length \( c_{ij} \).

Allow negative length edges, but no negative length cycles.

Consider all-to-one and one-to-all shortest path routing.

Special case of network flow problems in Lecture 5, and Bellman’s equation can be obtained by duality arguments.

Our development follows DP algorithm.
Bellman Ford Algorithm

Consider all-to-one shortest path routing with destination vertex $n$

Let $p_i(t)$ be length of shortest path from $i$ to $n$ using at most $t$ edges, with $p_i(t) = \infty$ if no such path exists

Let $p_n(t) = 0$, $\forall t$ and $p_i(0) = \infty$, $\forall i \neq n$

$p_i(t + 1)$ consists of two parts:

- cost of getting from $i$ to a neighboring $k$
- cost of getting from $k$ to destination $n$

Pick the minimum total cost:

$$p_i(t + 1) = \min_{k \in \mathcal{O}(i)} \{c_{ik} + p_k(t)\}$$
Example
Example
Example

A directed graph with labeled edges:

- From 0 to 6: 8
- From 0 to 7: 7
- From 6 to 0: 5
- From 6 to 7: -2
- From 6 to 4: -3
- From 7 to 2: -4
- From 7 to 6: 2
- From 4 to 7: 7
- From 4 to 2: 9
Example
Example
IP Routing

Basic versions:
- **RIP**: distance-vector based
- **OSPF**: link-state based
- **BGP**: across Autonomous Systems

Extensions:
- Multicast routing
- Mobile IP
- Mobile wireless ad hoc routing
- QoS routing
RIP Routing

Simple example:

Practical concerns:
- Loop avoidance
- Stability
- Speed of convergence
- Scalability
Hidden Markov Models

Markov chain with finite number of states and given transition probabilities \( p_{ij} \)

During transition, exact states are hidden but a related observation is obtained with probability \( r(z; i, j) \) for taking value \( z \) during transition \( i \to j \)

Memoryless property: \( r \) only depends on current transition

Estimate states based on a sequence of \( N \) observations: maximize conditional probability \( \text{Prob}(x^N | z^N) \) over all \( x^N \)

Equivalent to maximize \( \text{Prob}(x^N, z^N) \), which can be written as

\[
\text{Prob}(x^N, z^N) = \pi_{x_0} \prod_{k=1}^{N} p_{x_{k-1} x_k} r(z_k; x_{k-1}, x_k)
\]
Trellis Diagram

Estimation problem equivalent to minimizing

\[- \log(\pi_{x_0}) - \sum_{k=1}^{N} \log(p_{x_{k-1}x_k} r(z_k; x_{k-1}, x_k))\]

over all possible sequences \(\{x_0, \ldots, x_N\}\)

Concatenate \(N + 1\) copies of state space with dummy nodes \(s\) and \(t\)

An edge connects \(x_{k-1}\) and \(x_k\) if \(p_{x_{k-1}x_k} > 0\)
Viterbi Algorithm

Length of edge \((s, x_0)\): \(-\log(\pi x_0)\)

Length of edge \((x_{k-1}, x_k)\): \(-\log(p_{x_{k-1}x_k}r(z_k; x_{k-1}, x_k))\)

Forward DP algorithm:

\[
D_{k+1}(x_{k+1}) = \min_{x_k : p_{x_{k}x_{k-1}} > 0} \left[ D_k(x_k) - \log(p_{x_{k}x_{k+1}}r(z_{k+1}; x_k, x_{k+1})) \right]
\]

with initial condition \(D_0(x_0) = -\log(\pi x_0)\)

Find final state \(\hat{x}_N\) that minimizes \(D_N(x_N)\) over all possible \(x_N\)

Final estimated state sequence \(\hat{x}^N\) is the shortest path from \(s\) to \(\hat{x}_N\)
Convolutional Codes

Binary Data → Encoder → Transmit Codeword → Channel → Receive Codeword → Decoder → Decoded Data
Convolutional Codes

Convolutional codes: given $x_0$,

$$y_k = C x_{k-1} + dw_k, \quad k = 1, \ldots$$

$$x_k = A x_{k-1} + bw_k, \quad k = 1, \ldots$$

where $x_k \in \mathbb{R}^{m \times 1}$ with binary coordinates,

$C \in \mathbb{R}^{n \times m}, d \in \mathbb{R}^{n \times 1}, A \in \mathbb{R}^{m \times m}, b \in \mathbb{R}^{m \times 1}$

Products and sums are mod-2.

Example: let

$$C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}; \quad A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}; \quad d = \begin{bmatrix} 1 \\ 1 \end{bmatrix}; \quad b = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Then given initial state $x_0 = 00$ and input data $w = \{1, 0, 0, 1\}$

- State sequence: $\{00, 01, 11, 10, 00\}$
- Codeword sequence: $\{111, 011, 111, 011\}$
Viterbi Decoding

Given discrete memoryless channel \( \text{Prob}(z^N|y^N) = \prod_{k=1}^{N} \text{Prob}(z_k|y_k) \)

**ML sequence estimation:** find \( \hat{y}^N \) such that

\[
\text{Prob}(z^N|\hat{y}^N) = \max_{y^N} \text{Prob}(z^N|y^N)
\]

Equivalent to shortest path in trellis diagram or forward DP algorithm:

Minimize

\[
\sum_{k=1}^{N} - \log(\text{Prob}(z_k|y_k))
\]

over all possible binary sequences \( \{y_1, \ldots, y_N\} \)

Using Viterbi algorithm (DP recursion):

\[
D_{k+1}(x_{k+1}) = \min_{x_k: p_{x_k x_{k-1}} > 0} [D_k(x_k) - \log(\text{Prob}(z_{k+1}|y_{k+1}))]
\]

and backtrace to obtain \( \hat{x}^N \)

Using trellis diagram, obtain \( \hat{y}^N \)
Lecture Summary

• Simplest case of DP: deterministic and finite state

• Two most widely-used applications: shortest path routing in networking and convolutional decoding in communications (and Hidden Markov Models in signal processing and applied probability)

• Two simple and powerful algorithms: Bell Ford algorithm and Viterbi algorithm, are instances of DP recursions

Readings: Chapters 2 in Bertsekas *Dynamic Programming and Optimal Control, vol. 1*, and Section 7.9 in Bertsimas and Tsitsiklis *Introduction to Linear Optimization*