Biostatistics 615/815 Lecture 9: Dynamic Programming and Hidden Markov Models

Hyun Min Kang

October 2nd, 2012

Minimum edit distance problem

Edit distance
Minimum number of letter insertions, deletions, substitutions required to transform one word into another

An example

FOOD → MOOD → MOND → MONED → MONEY

Edit distance is 4 in the example above

More examples of edit distance

FOOD
MONEY
ALGORITHM

- Similar representation to DNA sequence alignment
- Does the above alignment provides an optimal edit distance?
Recursively formulating the problem

- Input strings are \( x[1, \cdots, m] \) and \( y[1, \cdots, n] \).
- Let \( x_i = x[1, \cdots, i] \) and \( y_j = y[1, \cdots, j] \) be substrings of \( x \) and \( y \).
- Edit distance \( d(x, y) \) can be recursively defined as follows

\[
d(x_i, y_j) = \begin{cases} 
  i & j = 0 \\
  j & i = 0 \\
  \min \{ d(x_{i-1}, y_j) + 1, d(x_i, y_{j-1}) + 1, d(x_{i-1}, y_{j-1}) + I(x_i \neq y_j) \} & \text{otherwise}
\end{cases}
\]

- Similar to the Manhattan tourist problem, but with 3-way choice.
- Time complexity is \( \Theta(mn) \).

**editDistance() algorithm**

```cpp
// note to declare the function before main()
int editDistance(std::string& s1, std::string& s2, Matrix615<int>& cost,
                 Matrix615<int>& move, int r, int c)
{
    int iCost = 1, dCost = 1, cCost = 1; // insertion, deletion, mismatch cost

    if ( cost.data[r][c] == INT_MAX ) {
        if ( r == 0 & c == 0 ) { cost.data[r][c] = 0; }
        else if ( r == 0 ) {
            move.data[r][c] = 0; // only insertion is possible
            cost.data[r][c] = editDistance(s1, s2, cost, move, r, c-1) + iCost;
        }
        else if ( c == 0 ) {
            move.data[r][c] = 1; // only deletion is possible
            cost.data[r][c] = editDistance(s1, s2, cost, move, r-1, c) + dCost;
        }
        else {move.data[r][c] = cost.data[r][c];}
    }
```

**editDistance() algorithm**

```cpp
else { // compare 3 different possible moves and take the optimal one
    int iDist = editDistance(s1, s2, cost, move, r, c-1) + iCost;
    int dDist = editDistance(s1, s2, cost, move, r-1, c) + dCost;
    int mDist = editDistance(s1, s2, cost, move, r-1, c-1) + (s1[r-1] == s2[c-1] ? 0 : mCost);
    if ( iDist < dDist ) {
        if ( iDist < mDist ) { // insertion is optimal
            move.data[r][c] = 0;
            cost.data[r][c] = iDist;
        }
        else {
            move.data[r][c] = 1; // match is optimal
            cost.data[r][c] = mDist;
        }
    }
}
```
editDistance() algorithm

```cpp
editDistance.cpp

else {
    if ( dDist < mDist ) {
        move.data[r][c] = 1; // deletion is optimal
        cost.data[r][c] = dDist;
    } else {
        move.data[r][c] = 2; // match is optimal
        cost.data[r][c] = mDist;
    }
}
return cost.data[r][c];
}
```

Running example

$ ./editDistance FOOD MONEY
EditDistance is 4

`-I**
F0-0D
MONEY

Graphical Model 101

- Graphical model is marriage between probability theory and graph theory (Michael I. Jordan)
- Each random variable is represented as vertex
- Dependency between random variables is modeled as edge
  - Directed edge: conditional distribution
  - Undirected edge: joint distribution
- Unconnected pair of vertices (without path from one to another) is independent
- An effective tool to represent complex structure of dependence / independence between random variables.
An example graphical model

- Are $H$ and $P$ independent?
- Are $H$ and $P$ independent given $S$?

Example probability distribution

$\Pr(H)$

<table>
<thead>
<tr>
<th>Value (H)</th>
<th>Description (H)</th>
<th>$\Pr(H)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Low</td>
<td>0.3</td>
</tr>
<tr>
<td>1</td>
<td>High</td>
<td>0.7</td>
</tr>
</tbody>
</table>

$\Pr(S|H)$

| S  | Description (S) | H  | Description (H) | $\Pr(S|H)$ |
|----|-----------------|----|-----------------|----------|
| 0  | Cloudy          | 0  | Low             | 0.7      |
| 1  | Sunny           | 0  | Low             | 0.3      |
| 0  | Cloudy          | 1  | High            | 0.1      |
| 1  | Sunny           | 1  | High            | 0.9      |

Full joint distribution

$\Pr(H, S, P)$

<table>
<thead>
<tr>
<th>H</th>
<th>S</th>
<th>P</th>
<th>$\Pr(H, S, P)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.105</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.105</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.009</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.081</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.035</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.035</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.063</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.567</td>
</tr>
</tbody>
</table>

- With a full joint distribution, any type of inference is possible
- As the number of variables grows, the size of full distribution table increases exponentially
\[ \Pr(H, P|S) = \Pr(H|S) \Pr(P|S) \]

| H | P | S | \Pr(H, P|S) |
|---|---|---|-------------|
| 0 | 0 | 0 | 0.3750     |
| 0 | 1 | 0 | 0.3750     |
| 1 | 0 | 0 | 0.1250     |
| 1 | 1 | 0 | 0.1250     |
| 0 | 0 | 1 | 0.0125     |
| 0 | 1 | 1 | 0.1125     |
| 1 | 0 | 1 | 0.0875     |
| 1 | 1 | 1 | 0.7875     |

| H | \Pr(H|S) |
|---|-----------|
| 0 | 0.750     |
| 1 | 0.250     |
| 0 | 0.125     |
| 1 | 0.875     |

| P | \Pr(P|S) |
|---|-----------|
| 0 | 0.500     |
| 1 | 0.500     |
| 0 | 0.100     |
| 1 | 0.900     |

- \( H \) and \( P \) are conditionally independent given \( S \)
- \( H \) and \( P \) do not have direct path one from another
- All path from \( H \) to \( P \) is connected thru \( S \).
- Conditioning on \( S \) separates \( H \) and \( P \)

Conditional independence in graphical models

\[ \Pr(A, C, D, E|B) = \Pr(A|B) \Pr(C|B) \Pr(D|B) \Pr(E|B) \]

Markov Blanket

- If conditioned on the variables in the gray area (variables with direct dependency), \( A \) is independent of all the other nodes.
- \( A \perp (U - A - \pi_A)|\pi_A \)
Markov Process: An example

Mathematical representation of a Markov Process

\[
\pi = \begin{pmatrix}
\Pr(q_1 = S_1 = \text{Sunny}) \\
\Pr(q_1 = S_2 = \text{Cloudy}) \\
\Pr(q_1 = S_3 = \text{Rainy})
\end{pmatrix}
= \begin{pmatrix}
0.7 \\
0.2 \\
0.1
\end{pmatrix}
\]

\[
A_{ij} = \Pr(q_{t+1} = S_j | q_t = S_i)
\]

\[
A = \begin{pmatrix}
0.5 & 0.3 & 0.2 \\
0.4 & 0.3 & 0.3 \\
0.1 & 0.5 & 0.4
\end{pmatrix}
\]

Example questions in Markov Process

What is the chance of rain in the day 2?

\[
\Pr(q_2 = S_3) = (A^T\pi)_3 = 0.24
\]

If it rains today, what is the chance of rain on the day after tomorrow?

\[
\Pr(q_3 = S_3 | q_1 = S_3) = (A^T)^2 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = 0.33
\]

Stationary distribution

\[
p = A^Tp
\]

\[
p = (0.346, 0.359, 0.295)^T
\]

Markov process is only dependent on the previous state

If it rains today, what is the chance of rain on the day after tomorrow?

\[
\Pr(q_3 = S_3 | q_1 = S_3) = (A^T)^2 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = 0.33
\]

If it has rained for the past three days, what is the chance of rain on the day after tomorrow?

\[
\Pr(q_5 = S_3 | q_1 = q_2 = q_3 = S_3) = \Pr(q_5 = S_3 | q_3 = S_3) = 0.33
\]
**Hidden Markov Models (HMMs)**

- A Markov model where actual state is unobserved
  - Transition between states are probabilistically modeled just like the Markov process
- Typically there are observable outputs associated with hidden states
  - The probability distribution of observable outputs given an hidden states can be obtained.

### Mathematical representation of the HMM example

**States** $S = \{S_1, S_2\} = \text{(HIGH, LOW)}$

**Outcomes** $O = \{O_1, O_2, O_3\} = \text{(SUNNY, CLOUDY, RAINY)}$

**Initial States** $\pi_i = \Pr(q_1 = S_i)$, $\pi = \{0.7, 0.3\}$

**Transition** $A_{ij} = \Pr(q_{t+1} = S_j | q_t = S_i)$

$$A = \begin{pmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{pmatrix}$$

**Emission** $B_{ij} = b_{q_t}(O_j) = \Pr(O_j = q_t) = S_i$

$$B = \begin{pmatrix} 0.88 & 0.10 & 0.02 \\ 0.10 & 0.60 & 0.30 \end{pmatrix}$$

### An example of HMM

![HMM example diagram](attachment:image.png)

- **Direct Observation**: (SUNNY, CLOUDY, RAINY)
- **Hidden States**: (HIGH, LOW)

### Unconditional marginal probabilities

**What is the chance of rain in the day 4?**

$$f(q_4) = \begin{pmatrix} \Pr(q_4 = S_1) \\ \Pr(q_4 = S_2) \end{pmatrix} = (A^T)^3 \pi = \begin{pmatrix} 0.669 \\ 0.331 \end{pmatrix}$$

$$g(o_4) = \begin{pmatrix} \Pr(o_4 = O_1) \\ \Pr(o_4 = O_2) \\ \Pr(o_4 = O_3) \end{pmatrix} = B^T f(q_4) = \begin{pmatrix} 0.621 \\ 0.266 \\ 0.233 \end{pmatrix}$$

The chance of rain in day 4 is 23.3%
## Summary

### Edit Distance
- Alignment between two strings
- Can be converted to a problem similar to MTP

### Hidden Markov Models
- Graphical models
- Conditional independence and Markov blankets
- Markov process
- Introduction to hidden Markov models

### Next lectures
- More hidden Markov Models