Hidden Markov Model

20090721

Han-eol Kim
(GIST, M.S. student)

광주과학기술원
Gwangju Institute of Science and Technology

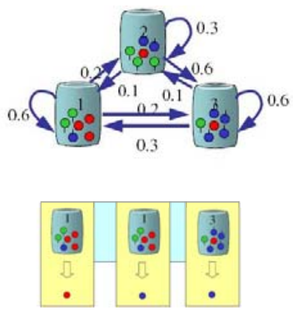
Introduction

- About Hidden Markov Models,
  - Background
  - Hidden Markov Models (Definition, Problem, Solution, Application)
  - Discussion

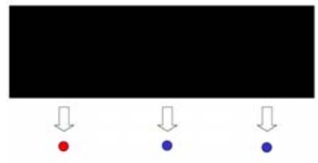
  - References
    - An introduction to Hidden Markov Models (L.R.Rabiner, 1986)
    - Dynamic Alignment Kernels (CJCH Watkins, 1999)

Markov chains vs. Hidden Markov Models

- Markov Chains : current state is observable
- Hidden Markov Models : current state is non-observable



Markov Process : {q(t)}
Output Process : {f(x|q)}



Which urn is selected at a time??

Background

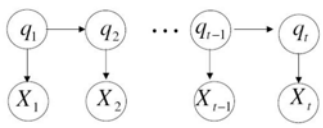
- 1st order Markov Assumption of transition

$$P(q_t \mid q_1, q_2, \cdots, q_{t-1}) = P(q_t \mid q_{t-1})$$

- Conditional independency of observation parameter

$$P(X_t \mid q_t, X_1, \cdots, X_{t-1}, q_1, \cdots, q_{t-1}) = P(X_t \mid q_t)$$

  - This is following Bayesian network representation



Definition of Hidden Markov Models

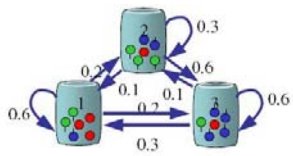
- Definition : A variant of a finite state machine having a set of states(Q), an output alphabet(O), transition probabilities(A), output probabilities(B), and initial state probabilities(Π). The current state is not observable. Instead, each state produces an output with a certain probability (B). Usually the states(Q) and outputs(O) are understood, so an HMM is said to be a triple, (A, B, Π).

- Notation : $\lambda = (A, B, \Pi)$
  - T : length of the observation sequence (total number of clock times)
  - N : number of states in the model (ex: number of urns)
  - M : number of observation symbols (ex: number of colors, RGB)
  - Q : states (ex: urns) $Q = \{q_1, q_2, \dots, q_N\}$
  - A : state transition probability distribution $A = \{a_{ij}\}$, $a_{ij} = \Pr(q_{j\ at\ t+1} | q_{i\ at\ t})$
  - B : observation symbol probability distribution $B = \{b_j(v_k)\}$, $b_j(k) = \Pr(v_{k\ at\ t} | q_{j\ at\ t})$
  - Π : initial state distribution $\pi = \{\pi_j\}$, $\pi_i = P(q_1 = i)$
  - O : output alphabet $O = \{O_1, O_2, \dots, O_N\}$

Example of HMM

- Number of states : N=3
- Number of observation : M=3, V={R,G,B}
- Initial state distribution $\pi = \{\pi_j\}$ = [1 , 0 , 0]
- State transition probability distribution

$$A = \{a_{ij}\} = \begin{bmatrix} 0.6 & 0.2 & 0.2 \\ 0.1 & 0.3 & 0.6 \\ 0.3 & 0.1 & 0.6 \end{bmatrix}$$



- Observation symbol probability distribution

$$B = \{b_i(v_k)\} = \begin{bmatrix} 3/6 & 2/6 & 1/6 \\ 1/6 & 3/6 & 2/6 \\ 1/6 & 1/6 & 4/6 \end{bmatrix}$$

Problems

- There are three key problems

[Problem 1] Given the observation sequence $O = \{O_1, O_2, \ldots, O_M\}$ and the model $\lambda = (A, B, \Pi)$, how we compute $Pr(O|\lambda)$, the probability of the observation sequence.
=> **Probability estimation problem**

[Problem 2] Given the observation sequence $O = \{O_1, O_2, \ldots, O_M\}$, how we choose a state sequence $I = i_1, i_2, \ldots, i_N$ which is optimal in some meaningful sense.
=> **Optimal sequence problem**

[Problem 3] How we adjust the model parameters $\lambda = (A, B, \Pi)$ to maximize $Pr(O|\lambda)$
=> **Parameter estimation problem**

Probability estimation problem

- Forward algorithm, Backward algorithm

Induction

1. Initialization:

$$\alpha_1(i) = \pi_i b_i(o_1)$$

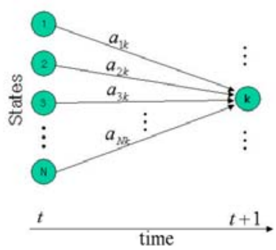
2. Induction:

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij}\right] b_j(o_{t+1})$$

3. Termination:

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i)$$

[Forward Algorithm]



By induction, t=1 to t=T

Optimal sequence problem

- Viterbi algorithm

Initialization

$$\delta_1(i) = \pi_i b_i(o_1), \quad 1 \le i \le N$$

$$\phi_1(i) = 0$$

Recursion

$$\delta_t(j) = \max_{1 \le i \le N} [\delta_{t-1}(i) a_{ij}] b_j(o_t)$$

$$\phi_t(j) = \arg\max_{1 \le i \le N} [\delta_{t-1}(i) a_{ij}]$$

$$2 \le t \le T, \quad 1 \le j \le N$$

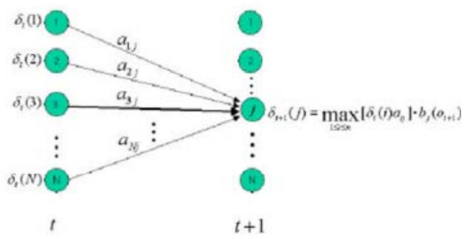
Termination

$$P^* = \max_{1 \le i \le N} [\delta_T(i)]$$

$$q_T^* = \arg\max_{1 \le i \le N} [\delta_T(i)]$$

Path (state sequence) backtracking

$$q_t^* = \phi_{t+1}(q_{t+1}^*), \quad t = T-1, T-2, \cdots, 1$$



Introduction Background **Hidden Markov Models** Discussion QnA

Parameter estimation problem

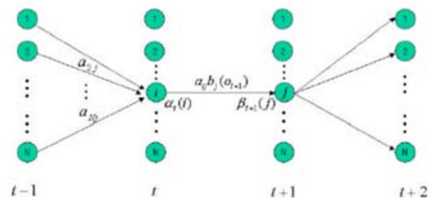
- There is EM(Baum-Welch) algorithm

Step 1. Let initial model be λ_0.

Step 2. Compute new λ based on λ_0 and observation O.

Step 3. If $\log P(O|\lambda) - \log P(O|\lambda_0) < DELTA$ then stop.

Step 4. Else set $\lambda_0 \leftarrow \lambda$ and goto step 2.



Applications of HMM

- Alignment of bio-sequences (Leek, 1997)

Ex)

```
AAB24882        TYHMCQFHCRYVNNHSGEKLYECNERSKAFSCPSHLQCHKRRQIGEKTHEHNQCGKAFPT 60
AAB24881        --------------------YECNQCGKAFAQHSSLKCHYRTHIGEKPYECNQCGKAFSK 40
                                    ****: .***:  * *:** * :****.:* *******..

AAB24882        PSHLQYHERTHTGEKPYECHQCGQAFKKCSLLQRHKRTHTGEKPYE-CNQCGKAFAQ- 116
AAB24881        HSHLQCHKRTHTGEKPYECNQCGKAFSQHGLLQRHKRTHTGEKPYMNVINMVKPLHNS 98
                 **** *:**********:***:**.: .**************    :  *.: :
```

- Pattern recognition
  - Speech recognition (Rabiner, 1989)
  - Character recognition (Freitag. 1998)

Discussion

- Applying HMM on our research,
  - From output symbol sequence, we can model something
    - We can model movement patterns of obstacles, insects, humans
  - Using HMM concept on Reinforcement learning, we can use HMM like as predictor
    - Overcoming incomplete perception with utile distinction memory (R.A. McCallum, 1993)

Any Questions?