Dynamic Time Warping and Hidden Markov Model

Pattern Recognition, Jul-Nov 2019

Indian Institute of Technology Madras

October 7, 2019

DTW

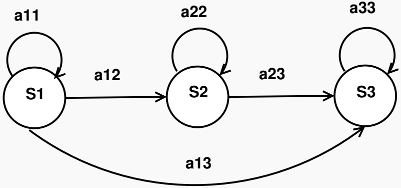
- Synchronous DTW - a non-model based algorithm

$$D(i,j) = min(D(i-1,k) + d(k,j)$$

  Here $d(k,j)$ is a distance metric that is positive.
  If test feature vector is of length m, the algorithm will have m steps since vertical movements are not allowed.
- Different variants of synchronous DTW
  - Constrained end point DTW - end points are anchored
  - Unconstrained end point DTW (UE-DTW) - no anchoring
  - Rough LCS (RLCS) algorithm
  - Longest common segment set algorithm

Hidden Markov Model (HMM)

- A generative model for sequential pattern recognition
- System being modeled is assumed to be a Markov process
- It is known as hidden Markov models because the states are hidden



- $S1, S2, S3$ are the states
- $a11, a12, \ldots$ are state transition probabilities

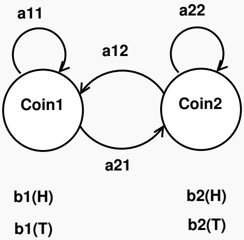
- The model is characterised by $\lambda = (N, M, \pi, A, B)$
  $N \implies$ No. of states
  $M \implies$ No. of observation symbols
  $\pi \implies$ Initial state probability
  $A \implies$ Transition matrix
  $B \implies$ Emission matrix

N, M are chosen emperically
An HMM is generally represented as:

$$\lambda = (\pi, A, B)$$

- Each element of A represents the probability of transition from one state to the next state. Eg: a_{12} represents the probability of transition from state 1 to state 2
  - Size of $A = NxN$
- Each element of B represents the probability of emitting an observation sequence in a particular state.
  - Size of $B = NxM$

Eg:Coin toss experiment



- Observation sequence is given, say, $HTTHHT$
- No. of symbols = 2 (H and T)
- No. of states = 2 ($Coin1$ and $Coin2$)
- State sequence is hidden. In this case, the sequence in which the coins are tossed to generate the given observation sequence is hidden.

HMM

Three basic HMM problems

- **Testing**
  What is the probability that a given model generates a given observation sequence?
- **Optimal state sequence**
  What is the most likely state sequence (path) through the given model that produced the given output sequence?
- **Training**
  What should the model parameters be so that it has a high probability of generating a set of given observed sequences?

Testing

- Find the probability that a given model generates a given observation sequence
  - Should consider all possible models and find the most likely one

$$\arg\max_{l} P(O|\lambda_l) where \lambda_l = (A_l, B_l, \pi_l)$$

That is, consider $\sum_{allQ} P(O|\lambda)$
Q - all possible state sequences that could have generated the observation sequence
$\sum_{allQ} P(O,Q|\lambda) = \sum_{allQ} P(O|Q,\lambda)P(Q|\lambda)$

Example:

Let O = HTH

Let the states be q_1, q_2, q_3

Possible state sequences:-

q_1	q_2	q_3	Probability of state sequence
1	1	1	$\pi_1 a_{11} a_{11} b_1(H) b_1(H) b_1(T)$
1	1	2	$\pi_1 a_{11} a_{12} b_1(H) b_1(T) b_2(H)$
1	2	1	
1	2	2	
2	1	1	
2	1	2	
2	2	1	
2	2	2	

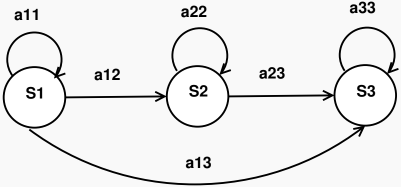
Saving the partial probability of generating partial observation sequence and being in state i at time t can reduce the computational cost

Forward Method

Two approaches for solving testing problem:

- Forward Method
- Backward Method

Forward Method



Forward Method

$$\alpha_t(i) = P(o_1, o_2, ...o_t, q_t = i|\lambda)$$

Initialization:

$$\alpha_1(i) = P(o_1, q_1 = i|\lambda) = \pi_i b_i(o_1); i = 1, 2, ..., N$$

$$\begin{aligned}\alpha_2(j) = P(o_1, o_2, q_2 = j|\lambda) &= \Big[\sum_{i=1}^{N} \pi_i b_i(o_1)\Big] a_{ij} b_j(o_2); j = 1, 2, ..., N \\ &= \alpha_1(i) a_{ij} b_j(o_2)\end{aligned}$$

Induction step

$$\alpha_{t+1}(j) = \sum_{i=1}^{N} \alpha_t(i) a_{ij} b_j(o_t); j = 1, 2, 3, ..., N$$

Termination

$$\alpha_T(i) = P(o_1, o_2, ...o_T, q_T = i|\lambda)$$
$$= \sum_{i=1}^{N} \alpha_T(i)$$

Backward method

Initialization

$$\beta_T(j) = 1; j = 1, 2, 3, ..., N$$

Induction

$$\beta_t(i) = P(o_{t-1}, o_{t-2}, ..., o_T | q_t = i, \lambda)$$

$$\beta_{t-1}(i) = P(o_T | q_{T-1} = i, \lambda)$$

$$= \sum_{j=1}^{N} a_{ij} b_j(o_T) \beta_T(j)$$

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(o_{t+1}) \beta_{t+1}(j); i = 1, 2, ..., N$$

Termination

$$\beta_1(i) = P(o_{t-1}, o_{t-2}, ..., o_T | q_1 = i, \lambda)$$
$$= P(O|\lambda) = \sum_{i=1}^{N} \pi_i b_i(o_1) \beta_1(t)$$

best model $= \arg\max_{l} P(O|\lambda_l)$

Optimal state sequence

Given, $O = o_1, o_2, o_3, ..., o_T$, find the best state sequence that explain the data.

$\delta_t(i) = \max_{q_1, q_2, ..., q_{t+1}} P(q_1, q_2, ..., q_{t+1}, q_t = i | o_1, o_2, ..., o_t, \lambda)$

Algorithm

1.

$$\delta_1(i) = \pi_i b_i(o_1); 1 \leq i \leq N$$

$$\psi_1(i) = 0$$

2. Recurssion:

$$\delta_t(j) = \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] b_j(o_t); 2 \leq t \leq T, 1 \leq j \leq N$$

$$\psi_t(j) = \arg\max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}]; 2 \leq t \leq T, 1 \leq j \leq N$$

3. Termination

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

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

$$q_t^* = \psi_{t+1}(q_{t+1}^*) t = T - 1$$

- An HMM becomes a Markov process after this procedure. The parameters can be re-estimated after this.