Mô hình Markov ẩn

Trịnh Quốc Huy - 20120013
Nguyễn Anh Khoa-20120118
Võ Thị Phước Thảo - 20120191

Ngày 14 tháng 5 năm 2022

Giảng viên hướng dẫn: Nguyễn Đình Thúc - Nguyễn Văn Quang Huy

1 Giới thiệu

Mô hình Markov ẩn (Hidden Markov model) là một mô hình máy học cổ điển thông dụng trong việc xử lý chuỗi (sequence processing). Cụ thể, mô hình này thường được dùng cho các bài toán phân loại các thành phần trong một chuỗi. Có nhiều thể hiện khác nhau của loại bài toán này, bao gồm xác định từ loại của các từ trong một câu (part-of-speech tagging) hoặc nhận dạng tiếng nói (speech recognition).

Trong đồ án này có 3 phần lý thuyết, cài đặt và vận dụng. Bài báo cáo này là để trả lời các câu hỏi trong phần lý thuyết.

2 Lý thuyết

2.1 Câu 1: Các thành phần của một mô hình Markov ẩn là gì? Chúng khác gì với mô hình Markov?

- Mô hình Markov ẩn có 2 thành phần là sự kiện quan sát được và các sự kiện ẩn
- Mô hình Markov chỉ có các sự kiện quan sát được trong quá khứ còn mô hình Markov ẩn có thêm phần sự kiện ẩn. Ngoài ra Markov ẩn còn có thêm phần mạng Bayesian.

2.2 Câu 2: Các giả thiết (assumption) đặt ra cho mô hình Markov ẩn là gì? Tìm ví dụ các bài toán mà các giả thiết này hợp lý và bất hợp lý.

- Các giả thiết đặt ra cho mô hình Markov ẩn gồm:

- Tập các trạng thái Q

$$Q = (q_1, q_2, ..., q_N)$$

- Tập các dữ liệu quan sát được V

$$V = (v_1, v_2, .., v_V)$$

- Tập khả năng ban đầu π

$$\pi = (\pi_1, \pi_2, ..., \pi_n)$$

- Chuỗi các trạng thái S

$$S = (s_1, s_2, ..., s_T)$$

- Chuỗi các sự kiện quan sát được O

$$O = (o_1, o_2, ...o_T)$$

- Tập các phân bố xác suất A với $A_j(k)$ là xác suất dịch chuyển được ở sự kiện V_k trong trạng tháiS_j
- Tập các phân bố xác suất A với $A_j(k)$ là xác suất đầu ra được ở sự kiện V_k trong trạng tháiS_j

• Ví dụ bài toán có giả thiết là hợp lý:
Giả sử bạn bị nhốt trong phòng trong vài ngày và bạn được hỏi về thời tiết bên ngoài Một bằng chứng duy nhất bạn có là liệu người vào phòng có mang theo bữa ăn hàng ngày của bạn có mang theo dù hay không.
Dưới đây là bảng xác suất thời tiết ngày mai dựa vào thời tiết hôm nay:

Thời tiết hôm nay	Thời tiết ngày mai		
	Nắng	Mưa	Mây mù
Nắng	0.8	0.05	0.15
Mưa	0.2	0.6	0.2
Mây mù	0.2	0.3	0.5

Đây là bảng xác suất thời tiết nếu người đó có mang dù hay không mang dù.

Thời tiết	Có mang dù	Không mang dù
Nắng	0.1	0.9
Mưa	0.8	0.2
Mây mù	0.3	0.7

Các tập các trạng thái $Q = (nng, ma, mym)$
Tập quan sát được $V = (cmangd, khngmangd)$

Giả sử hôm qua trời nắng và hôm nay người mang cơm có mang theo dù. Hãy dự đoán hôm nay thời tiết như thế nào?
Xác suất hôm nay trời nắng: $L(s2 = nng|s1 = nng, v2 = cmangd) = P(v2 = cmangd|s2 = nng) * P(s2 = nng|s1 = nng) = 0.1 * 0.8 = 0.08$
Xác suất hôm nay trời mưa: $L(s2 = ma|s1 = nng, v2 = cmangd) = P(v2 = cmangd|s2 = ma) * P(s2 = ma|s1 = nng) = 0.8 * 0.05 = 0.04$
Xác suất hôm nay trời có mây mù: $L(s2 = mym|s1 = nng, v2 = cmangd) = P(v2 = cmangd|s2 = mym) * P(s2 = mym|s1 = nng) = 0.3 * 0.15 = 0.045$
$\rightarrow$ Vậy xác suất để hôm nay trời nắng là cao nhất.

- Ví dụ bài toán có giả thiết không hợp lý:
  Trong dịp lễ 30 tháng 4 và 1 tháng 5, nhóm bạn Huy, Khoa và Thảo rủ nhau đi ăn quán gà lá é Tao Ngộ trên Đà Lạt, nhưng vì bạn Khoa bất ngờ bị COVID nên bạn Khoa đành ở nhà theo dõi bạn Huy và Thảo đi ăn, đồng thời bạn Khoa chỉ biết quán Tao Ngộ có ba chi nhánh. Tuy nhiên, vì là dịp lễ nên quán rất đông, không phải chi nhánh nào cũng đủ chỗ cho Huy và Thảo. Khoa ở nhà và dự đoán dựa trên hình ảnh Huy và Thảo sẽ ăn ở quán nào, có 2 quan sát là ăn và không ăn dựa trên bài đăng Facebook, trạng thái sẽ là các chi nhánh 1 hoặc 2 hoặc 3. Xác suất để đủ chỗ cho Huy và Thảo ở quán đầu là 0.4, nếu quán 1 không còn chỗ thì Huy và Thảo sẽ đến quán 2, tuy vậy xác suất để quán 2 còn chỗ là 0.3, nếu chi nhánh 2 hết chỗ thì Huy và Thảo sẽ di chuyển lên chi nhánh 3, xác suất để chi nhánh 3 còn chỗ là 0.2.Điều không hợp lí ở đây là số lượng khách ở các chi nhánh không liên quan tới nhau nên giả thiết là không hợp lí.

2.3 Câu 3: Cho một mô hình Markov ẩn với các tham số đã biết, thuật toán tiến trước (forward algorithm) được dùng để xác định độ hợp lý (likelihood) của một chuỗi quan sát (observation). Mô tả và đánh giá độ phức tạp của thuật toán tiến trước.

- Mô tả:

  Cho chuỗi $O = (o_1, o_2, ..., o_T)$ và các ma trận xác suất dịch chuyển (transition) A và ma trận xác suất đầu ra(emission) B.

  Ta cần tính $P(O|A, B)$. Ta dùng thuật toán quy hoạch động với từng $t = 1, 2, .., T$ và các trạng thái ẩn s ta có công thức:

$$f[t, s] = P(o_1, o_2, ..., o_t, s_t = s|A, B)$$

nếu có $f[t,s]$ ta tính $P(O|A,B)$ với công thức:

$$P(o_1,o_2,...,o_t|A,B) = \sum_s P(o_1,o_2,..,o_t,s_t=s|A,B).P(o_t|s_t=s)$$
$$= \sum_s f[t,s].B(s,o_t)$$

Mã giả:

1. Khởi tạo:
   - Với từng trạng thái ẩn s:
   $$f[t,s] = P(o_1,s_1=s|A,B) = B(s,o_1).A(s_0,s)$$
2. Lặp $t=2$ đến T:
   - Với từng trạng thái ẩn s:
   $$f[t,s] = \sum_{s'} f[t-1,s'].A(s',s)B(s,o_t)$$
3. Cuối cùng:
   $$P(o_1,o_2,...o_T)|A,B) = \sum_s f[T,s]B(s,o_T)$$

- Độ phức tạp: $O(N^2.T)$ với N là kích thước của tập ẩn.

2.4 Câu 4: Cho một mô hình Markov ẩn với các tham số đã biết, thuật toán Viterbi được dùng để xác định chuỗi trạng thái (state) khả dĩ nhất. Mô tả và đánh giá độ phức tạp của thuật toán Viterbi.

- Mô tả

  Cho chuỗi $O=(o_1,o_2,..,o_T)$ và các ma trận xác suất dịch chuyển(transition) và ma trận xác suất đầu ra A. Vói từng trạng thái ẩn S, đẻ tính trường hợp khả dĩ nhất mà cả O và S xảy ra là:

  $$P(O,S|A,B) = P(O|S,A,B).P(S|A,B)$$
  $$= \prod_{i=1}^{T} P(o_t|s_T).\prod_{i=1}^{T} P(s_T|s_{t-1})$$
  $$= \prod_{i=1}^{T} B(s_t,o_t).\prod_{i=1}^{T} A(s_{t-1},s_t)$$

  Ta cần tính chuỗi S sao cho biểu thức trên đạt giá trị lớn nhất, hay

  $$S^* = argmax_S P(O,S|A,B)$$

. Thuật toán Viterbi được dùng để xác định chuỗi trạng thái khả dĩ nhất. Ta có:

$$\begin{aligned} g[t,s] &= \max_{s_1 \dots s_{t-1}} P(o_1, o_2, \dots_t, s_t = s|A,B) \\ &= \max_{s'} P(o_1, o_2, ..o_t, s_{t-1} = s'|A,B).P(s_t = s|s_{t-1} = s')P(o_t|s) \\ &= \max_{s'} g[t-1, s'].A(s', s)B(s, o_t) \end{aligned}$$

Vì g[t,s] đã biết nên để tính xác suất khả dĩ P(O,S|A,B) ta có:

$$\max_{s_1 \dots s_{t-1}} P(o_1, o_2, .., o_t|A,B) = \max_{s} P(o_1, o_2, ...o_t, s_t = s|A,B).P(o_t|s_t = s)$$

$$= \max_{s} g[t,s].B(s, o_t)$$

Để tính giá trị argmax(.) ta tạo 1 mảng để lưu vết.Cụ thể khi tính được giá trị s' làm cho $g[t-1, s'].A(s', s)B(s, o_t)$, ta lưu s' vào mảng $h[t,s]$. Sau khi tính hết các giá trị đến $g[T,s]$, ta sẽ truy vết những trạng thái ẩn để giá trị của $P(O,S|A,B)$ lớn nhất thì đso là argmax.

*Mã giả:

1. Khởi tạo
   - Với từng trạng thái ẩn s:

$$g[1,s] = B(s, o_1).A(s_0, s)$$

2. Lặp t từ 2 đến T:
   - Với từng trạng thái ẩn s:

$$g[t,s] = \max_{s'} g[t-1, s'].A(s', s)B(s, o_t)$$

$$h[t,s] = \operatorname*{argmax}_{s'} g[t-1, s'].A(s', s)B(s, o_t)$$

3. Từ mảng h[t,s] tìm s_T^*,s_{T-1}^*,...,s_1^*. Từ $t = T$

$$\begin{aligned} &s_T^* = argmax g[T,s] \\ &s_T^* = h[t+1, s_{t+1}^*] \qquad for \quad t = T-1, T-2, ..., 1 \end{aligned}$$

- Đánh giá độ phức tạp: $O(N^2.T)$

2.5 Câu 5: Cho một chuỗi quan sát, giả sử ta cho rằng chuỗi quan sát này được sinh ra từ một mo hình Markov ẩn với tham số chưa biết, thuật toán Baum-Welch được dùng để ước luọng các tham số này. Thuật toán Baum-Welch là trường hợp đặt biệt của thuật toán Kỳ vọng-Tối ưu (Expectation-Maximization, hay EM). Thuật toán này gồm 2 bước: bước E(Expectation, hay kỳ vọng) và bước M (Maximization, hay tối ưu)

1. Mô tả thuật toán kỳ vọng - tối ưu tổng quát

   - Mô tả - Giả sử ta quan sát được biến x mà không quan sát được biến y (ẩn).
     - Cho mẫu x, ước lượng mật độ $p(x, y)$ với lớp hàm phân bố $p(x, y; \theta)$
     - Ước lượng hợp lý cực đại (MLE)

     $$\theta^{MLE} = \underset{\theta}{argmax} logp(x, \theta$$

     $$= \underset{\theta}{argmax} log[\int_y p(x, y; \theta)dy]$$

     - Bất đẳng thức biến phân cho EM:
     Xét một phân bố bất kì $q(y)$, ta có :

     $$logp(x, \theta) = \int_y logp(x, \theta)q(y)dy]$$

     $$= \int_y lfp(x, y; \theta)q(y)dy - \int_y logp(y|x; \theta)q(y)dy$$

     Trong đó:

     $$\int_y logp(y|x; \theta)q(y)d(y) = E_q[logp(x, y; \theta)] - \int_y logp(y|x; \theta)q(y)d(y)$$

     $$= \int_y log\frac{q(y)}{p(y|x; \theta)}q(y)d(y) - \int_y logq(y)q(y)dy$$

     Với $E_q[logp(x, y; \theta)]$ là kì vọng theo phân bố q(y) $D_{KL}[q||p(y|x; \theta] = \int_y log\frac{p(y|x;\theta}{q(y)}q(y)d(y)$ là khoảng cách Kullback-Leibler
     $\epsilon[q]$ là entropy của phân bố $q(y)$
     $logp(x; \theta$ là sự hợp lý bị chặn dưới bởi kì vọng + entropy
     - Cận dưới này chặt nhất(dấu bằng xảy ra) khi

     $$q(y) = p(y|x; \theta)$$

     - Nếu cố định $q(y)$ như trên thì chỉ cần cực đại hóa kì vọng

     $$E_{p(y|x;\theta}[logp(x, y; \theta)]$$

- Các bước thuật toán
  while Chưa hội tụ do

  - Bước E: tính phân bố hạ nghiệm$p(y|x;\theta^k$ và biểu thức kì vong

  $$Q(\theta|\theta^{(k)}) = E_{p(y|x;\theta^{(k)})}[logp(x, y; \theta)]$$

  - Bước M: cực đại hóa kì vọng

  $$\theta^{(k+1)} = \underset{\theta}{argmax} Q(\theta|\theta^{(k)})$$

  - k = k + 1

  end while

2. Mô tả và đánh giá độ phức tạp của bước E và bước M của thuật toán Baum-Welch

   (a) Bước kỳ vọng (Expectation)

   - Mô tả: Giả sử ta đã biết A, B và cần tính

   $$\gamma[t, s] = P(s_t = s|O, A, B)$$
   $$\xi[t, s', s] = P(s_{t-1} = s', s_t = s|O, A, B)$$

   Trong đó, $\gamma[t, s]$ đếm số lần trạng thái ẩn thứ t bằng s và $\xi[t, s, s']$ đếm số lần cặp (s', s)xảy ra ở bước thứ $t-1$ và bước thứ t trong chuỗi trạng thái ẩn. Chúng đều được chuẩn hóa và là xác suất có điều kiện trên O.

   $$\begin{aligned}\alpha[t, s] &= P(o_1, o_2, ..., o_t, s_t = s|A, B) \\ &= \sum_{s'} \alpha[t-1, s'].A(s', s).B(s, o_t) = B(s, o_t) \sum_{s'} \alpha[t-1, s'].A(s', s) \\ \beta[t, s] &= P(o_{t+1}, o_{t+2}, ..., o_T|s_t = s, A, B) \\ &= \sum_{s'} \beta[t+1, s'].A(s, s').B(s', o_{t+1})\end{aligned}$$

   Với $\alpha[t, s]$ và $\beta[t, s]$ đã biết, dưới dây là công thức tính $\gamma[t, s]$ và $\xi[t, s', s]$:

   $$\begin{aligned}\gamma[t, s] &= P(s_t = s|O, A, B) = \frac{P(s_t = s, O|A, B)}{P(O|A, B)} \\ &= \frac{\alpha[t, s]\beta[t, s]}{\sum_{s'} \alpha[t, s'].\beta[t, s']} \\ \xi[t, s', s] &= P(s_{t-1} = s', s_t = s|O, A, B) = \frac{P(s_{t-1} = s', s_t = s, O|A, B)}{P(O|A, B)} \\ &= \frac{\alpha[t-1, s].A(s', s).\beta[t, s]}{\sum_{s'} \alpha[t, s'].\beta[t, s']}\end{aligned}$$

- Đánh giá độ phức tạp:$O(T * N^2)$

(b) Bước tối đa (Maximization):

- Mô tả: Khi có được các giá trị γ và ξ, trong bước này, ta có thể đối đa ước lượng độ hợp lí để lấy các giá trị cập nhật cho A và B

$$\hat{A}(s', s) = \frac{\sum_{t=1}^{T-1} \xi[t, s', s]}{\sum_{s''} \sum_{t=1}^{T-1} \xi[t, s', s'']}$$
$$\hat{B}(s, o) = \frac{\sum_{t=1}^{T} 1[o_t = o]\gamma[t, s]}{\sum_{t=1}^{T} \gamma[t, s]}$$

- Độ phức tạp: $O(N^2)$

3 Reference

[1] http://www.cs.cmu.edu/~tbergkir/11711fa17/recitation4_notes.pdf