BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC MỎ ĐỊA CHẤT

BÁO CÁO HỌC THUẬT

XÂY DỰNG THUẬT TOÁN TIẾN CHO HMM KHÔNG THUẦN NHẤT

TS. Nguyễn Thị Hằng

Hà Nội - 06/2022

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC MỎ ĐỊA CHẤT

BÁO CÁO HỌC THUẬT

XÂY DỰNG THUẬT TOÁN TIẾN CHO HMM KHÔNG THUẦN NHẤT

Xác nhận của bộ môn

Hà Nội - 06/2022

LỜI MỞ ĐẦU

Khoa học thống kê đóng vai trò không thể thiếu trong bất cứ công trình nghiên cứu khoa học nào, nhất là khoa học thực nghiệm như y khoa, nông nghiệp, hóa học, và ngay cả xã hội học. Để phát hiện ra những quy luật đằng sau những con số, người làm thống kê phải tiến hành công việc suy luận thống kê. Hiểu một cách đơn giản, suy luận thống kê là quá trình tìm ra các quy luật từ dữ liệu thực tế.

Báo cáo học thuật đề cập đến việc nghiên cứu lớp mô hình MTT mà trong đó các mục tiêu được quan tâm là lớp mục tiêu nào đó trong số các mục tiêu có thể có của mô hình MTT. Trường hợp lớp mục tiêu được quan tâm là lớp con thực sự của lớp tất cả các mục tiêu có thể có của mô hình, thì các thuật toán về MTT đã được công bố cho đến thời điểm hiện tại không thể áp dụng được và hiện chưa có phương pháp giải nào được công bố. Báo cáo đã đưa ra phương pháp tiếp cận đó là sử dụng mô hình Markov ẩn để giải quyết. Trong báo cáo sẽ trình bày về mô hình Markov ẩn không thuần nhất và cách xây dựng thuật toán tiến cho mô hình Markov ẩn không thuần nhất này.

Báo cáo được chia thành ba chương:

- **Chương 1. Một số vấn đề về quá trình ngẫu nhiên** Trong chương này trình bày các kiến thức về quá trình ngẫu nhiên như: quá trình Poisson, quá trình điểm Poisson tổng quát, quá trình Poisson phức hợp, quá trình Markov,

- **Chương 2. Bài toán MTT và biến tiến**

  Nội dung chương này bao gồm các kiến thức cơ bản liên quan tới báo cáo : Mục 1.1 là mục "Tổng quan về bài toán MTT"; mục 1.2 là mục phát biểu và xây dựng mô hình toán học cho bài toán MTT được nghiên cứu trong báo cáo này.

- **Chương 3. Xây dựng thuật toán tiến cho mô hình HMM không thuần nhất**

  Tiếp theo Chương 3 sẽ trình bày các kiến thức liên quan đến mô hình Markov ẩn HMM không thuần nhất; các kết quả nghiên cứu mở rộng các kết quả trong HMM không thuần nhất cho biến tiến và xây dựng thuật toán tiến.

Hà Nội, ngày 25 tháng 06 năm 2022

Chương 1

Một số vấn đề về quá trình ngẫu nhiên

1.1. Quá trình Poisson

1.1.1. Phân phối Poisson

Định nghĩa 1.1.1. *Ta nói rằng biến ngẫu nhiên X có phân phối Poisson với tham số λ, $\lambda > 0$ nếu:*

$$P\left[X = k\right] = \frac{\lambda^k}{k!}\, e^{-\lambda}, \quad k = 0, 1, 2, \ldots$$

Tính chất 1.1..1.

- *Nếu X và Y là hai biến ngẫu nhiên có phân phối Poisson độc lập với tham số λ_1, λ_2 tương ứng ($\lambda_1 > 0$, $\lambda_2 > 0$) thì $X + Y$ có phân phối Poisson với tham số $\lambda = \lambda_1 + \lambda_2$.*

- *Giả sử N là biến ngẫu nhiên có phân phối Poisson với tham số λ, $\lambda > 0$, và X là biến ngẫu nhiên sao cho:*

  $$P\left[X = k | N = n\right] = C_n^k\, p^k (1 - p)^{n-k}, \quad k = 0, 1, 2, \ldots, n$$

  Khi đó, X có phân phối Poisson với tham số $\lambda.p$

1.1.2. Quá trình Poisson

Định nghĩa 1.1.2. *Giả sử A là biến cố nào đó, ký hiệu $N(t)$, $t \geq 0$, là số lần biến cố A xuất hiện trong khoảng thời gian từ 0 đến t (kể cả thời điểm t). Khi đó, $\{N(t), t \geq 0\}$ được gọi là quá trình đếm. Ký hiệu $N(s,t] = N(t) - N(s)$, $0 \leq s < t$ là số lần biến cố A xảy ra trong khoảng thời gian $(s,t]$. Khi đó, $\{N(s,t], 0 \leq s < t\}$ được gọi là quá trình điểm ứng với quá trình đếm $\{N(t), t \geq 0\}$.*

Để định nghĩa quá trình Poisson và chỉ ra mối quan hệ mật thiết của nó với quá trình đếm, chúng ta đưa ra các khái niệm và tính chất sau đây:

a. Tính chất có gia số độc lập: Quá trình ngẫu nhiên $\{Z(t), t \geq 0\}$ được gọi là có gia số độc lập nếu với mọi $m = 2, 3, \ldots$ và với mọi $0 = t_0 < t_1 < \cdots < t_m$, các gia số $Z(t_0, t_1]$, $Z(t_1, t_2]$,..., $Z(t_{m-1}, t_m]$ là các biến ngẫu nhiên độc lập, trong đó:
$$Z(t_k, t_{k+1}] = Z(t_{k+1}) - Z(t_k),\ k = 0, 1, \ldots, m-1.$$

b. Tính chất có gia số dừng: Quá trình $\{Z(t), t \geq 0\}$ được gọi là có gia số dừng nếu với mọi $s > 0, 0 \leq t_1 < t_2$, các gia số $Z(t_1 + s, t_2 + s]$, $Z(t_1, t_2]$ là các biến ngẫu nhiên có cùng phân phối xác suất.

c. Tính chất c: Ta nói quá trình $\{Z(t), t \geq 0\}$ có tính chất c nếu tồn tại một hằng số λ, $\lambda > 0$, sao cho với $h > 0$ đủ nhỏ thì:
$$P[Z(h) = 1] = \lambda h + o(h),$$
ở đây, $o(h)$ là ký hiệu vô cùng bé bậc cao hơn của h (khi $h \to 0$).

d. Tính chất d: Ta nói quá trình $\{Z(t), t \geq 0\}$ có tính chất d nếu với $h > 0$ đủ nhỏ thì:
$$P[Z(h) \geq 2] = o(h).$$

Định nghĩa 1.1.3. (Định nghĩa quá trình Poisson)
Ta nói rằng $\{X(t), t \geq 0\}$ là quá trình Poisson với cường độ λ (hoặc tham số λ), $\lambda > 0$, nếu:

i. $X(t)$ nhận các giá trị $0, 1, 2, \ldots, \forall t > 0$.

ii. $\{X(t), t \geq 0\}$ *là quá trình có gia số độc lập.*

iii. *Mỗi gia số* $X(t+s) - X(s)$, $s \geq 0, t > 0$, *là biến ngẫu nhiên có phân phối Poisson với tham số* λt.

iv. $X(0) = 0$.

Các tính chất trực tiếp (cần cho báo cáo):

Tính chất 1.1..2.

- *Nếu* $\{X(t), t \geq 0\}$ *là quá trình Poisson thì* $X(t)$ *là biến ngẫu nhiên có phân phối Poisson với tham số* λt.

- *Quá trình Poisson là quá trình đếm có các tính chất a, b, c và d; Ngược lại, quá trình đếm có các tính chất a, b, c, d là quá trình Poisson.*

Ghi chú:

Do trong báo cáo chỉ sử dụng đến quá trình Poisson thuần nhất theo thời gian, nghĩa là cường độ của quá trình là hằng số không phụ thuộc vào thời gian, nên ở đây chúng tôi chỉ định nghĩa và trích lọc một vài tính chất cần thiết của quá trình Poisson thuần nhất cần cho báo cáo.

Trong trường hợp không thuần nhất, nghĩa là trường hợp cường độ phụ thuộc vào thời gian, tức là $\lambda = \lambda(t)$, chúng ta có quá trình Poisson không thuần nhất (theo thời gian). Khi đó, gia số $X(t) - X(s)$ là biến ngẫu nhiên có phân phối Poisson với tham số:

$$\int_s^t \lambda(u)\, du \quad , \quad 0 \leq s < t$$

1.1.3. Quá trình điểm Poisson tổng quát

Giả sử S là một tập khác rỗng trong không gian tổng quát nào đó (chẳng hạn trong không gian $\mathbb{R}^d$); ký hiệu $\mathcal{A}$ là họ các tập con nào đó của S; ký hiệu $\mu(.)$ là độ đo xác định trên $\mathcal{A}$; khi đó, $(S, \mathcal{A}, \mu)$ là một không gian có độ đo tổng quát. Xét $(S, \mathcal{A}, \mu)$ là không gian có độ đo nào đó.

Định nghĩa 1.1.4. *Ta gọi quá trình điểm Poisson thuần nhất trong* S *là họ các biến ngẫu nhiên* $\{\mathcal{N}(A), A \in \mathcal{A}\}$ *sao cho:*

i. Với mỗi $A \in \mathcal{A}$, $\mu(A) < \infty$ thì $\mathcal{N}(A)$ là biến ngẫu nhiên Poisson với tham số $\lambda \cdot \mu(A)$; $\lambda = const$ nào đó.

ii. Nếu A_1, A_2,..., A_n là các tập con rời nhau thuộc $\mathcal{A}$ thì $\mathcal{N}(A_1)$, $\mathcal{N}(A_2)$,..., $\mathcal{N}(A_n)$ là các biến ngẫu nhiên độc lập và

$$\mathcal{N}\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} \mathcal{N}(A_k) \qquad h.c.c$$

Khi S là tập con trong không gian d-chiều ($\mathbb{R}^d$) với độ đo (Lebesgue) thông thường, thì $\{\mathcal{N}(A), A \in \mathcal{A}\}$ được gọi là quá trình (Poisson) không gian d-chiều. Ký hiệu $]|A|[$ là diện tích ($d = 2$), thể tích ($d = 3$),..., của A; chúng ta dùng một từ chung gọi $]|A|[$ là cỡ của A.

Do mục tiêu nghiên cứu và cách tiếp cận nên có một số định nghĩa quá trình điểm Poisson không giống nhau, song các định nghĩa đó là tương đương. Với mục đích phục vụ cho báo cáo nên ở đây chúng tôi dùng định nghĩa như vừa phát biểu ở trên.

Ta có tính chất quan trọng sau đây:

Định lý 1.1.1. *Giả sử $\{\mathcal{N}(A), A \in \mathcal{A}\}$ là một quá trình điểm Poisson tổng quát nào đó, $]|A|[> 0$ và A_1, A_2,..., A_m là một phân hoạch của A. Khi đó,*

$$\begin{aligned} & P\left[\mathcal{N}(A_1) = k_1, \mathcal{N}(A_2) = k_2, \ldots, \mathcal{N}(A_m) = k_m | \mathcal{N}(A) = n\right] \\ = \quad & \frac{n!}{k_1!\, k_2!\, \ldots k_m!} \prod_{i=1}^{m} \left(\frac{]|A_i|[}{]|A|[}\right)^{k_i}, \end{aligned}$$

trong đó, $n \geq 1$ và các k_1, k_2,...,k_m là các số nguyên dương sao cho $\sum_{i=1}^{m} k_i = n$.

Từ định lý này, chúng ta thấy nếu A chứa n điểm thì n điểm này độc lập và phân phối đều trong A.

1.1.4. Quá trình Poisson phức hợp

Định nghĩa 1.1.5. *Giả sử $\{X(t), t \geq 0\}$ là quá trình Poisson với cường độ λ, $\lambda > 0$; $\{Y_k, k = 1, 2, \ldots\}$ là dãy các biến ngẫu nhiên độc lập, cùng phân phối và dãy này độc lập với $\{X(t), t \geq 0\}$. Khi đó, chúng ta gọi:*

$$Z(t) = \sum_{k=1}^{X(t)} Y_k \quad , \quad t \geq 0$$

là quá trình Poisson phức hợp.

1.2. Quá trình Markov

1.2.1. Các định nghĩa và các khái niệm cơ bản

Tính Markov của quá trình ngẫu nhiên và các định nghĩa liên quan:

Giả sử $\{X(t), t \geq 0\}$ là một quá trình ngẫu nhiên. Ký hiệu $\mathcal{F}_{\leq s} = \sigma\left(\{X(l), l \leq s\}\right)$ là σ-đại số cảm sinh bởi họ các biến ngẫu nhiên $\{X(l), l \leq s\}$; $\mathcal{F}_s = \sigma\left(X(s)\right)$ là σ-đại số cảm sinh bởi biến ngẫu nhiên $X(s)$.

Định nghĩa 1.2.1. *Ta nói quá trình ngẫu nhiên $\{X(t), t \geq 0\}$ có tính Markov nếu:*

$$E\left\{X(t)|\mathcal{F}_{\leq s}\right\} = E\left\{X(t)|\mathcal{F}_s\right\}, \quad \forall\, t > s \geq 0.$$

Định nghĩa 1.2.2. *Ta nói quá trình ngẫu nhiên $\{X(t), t \geq 0\}$ có tính Markov nếu:*

$$P\left[X(t_{n+1})|X(t_0), X(t_1), \ldots, X(t_n)\right] = P\left[X(t_{n+1})|X(t_n)\right]$$

với bất kỳ $0 \leq t_0 < t_1 < \cdots < t_n < t_{n+1}$.

Hai định nghĩa trên là tương đương. Việc phát biểu cả hai dạng định nghĩa ở đây nhằm mục đích để tiện sử dụng trong các nội dung sau này.

Ký hiệu E là không gian giá trị của quá trình ngẫu nhiên $\{X(t), t \geq 0\}$ (E sẽ được gọi là không gian trạng thái của quá trình $\{X(t), t \geq 0\}$). Chúng ta có các khái niệm sau đây:

Định nghĩa 1.2.3. *Quá trình ngẫu nhiên $\{X(t), t \geq 0\}$ có tính Markov thì $\{X(t), t \geq 0\}$ được gọi là quá trình Markov.*

Nếu $\{X(t), t \geq 0\}$ là quá trình Markov thì E được gọi là không gian trạng thái của quá trình Markov $\{X(t), t \geq 0\}$.

Nếu quá trình Markov $\{X(t), t \geq 0\}$ có không gian trạng thái E có lực lượng không quá đếm được thì $\{X(t), t \geq 0\}$ được gọi là xích Markov.

Nếu xích Markov $\{X(t), t \geq 0\}$ có không gian trạng thái có lực lượng hữu hạn ($Card(E) < +\infty$) thì $\{X(t), t \geq 0\}$ được gọi là xích Markov hữu hạn trạng thái.

Nếu xích Markov $\{X(t), t \geq 0\}$ với t chỉ nhận các giá trị rời rạc (tương đương với $t = 0, 1, 2, \dots$) thì $\{X(t), t \geq 0\}$ được gọi là xích Markov với thời gian rời rạc hay còn được gọi là dãy Markov, và khi đó người ta thường dùng ký hiệu đơn giản là $\{X_n, n = 0, 1, 2, \dots\}$.

Đối với xích Markov với thời gian rời rạc $\{X_n, n = 0, 1, 2, \dots\}$, người ta thường quan tâm tới:

$$p_{ij} = P\left[X_{n+1} = j | X_n = i\right], \quad i, j \in E.$$

Khi đó, p_{ij} được gọi là xác suất chuyển sau một bước.

Nếu p_{ij} không phụ thuộc vào n thì xích Markov $\{X_n, n = 0, 1, 2, \dots\}$ được gọi là xích Markov thuần nhất; trong trường hợp ngược lại gọi là xích Markov không thuần nhất.

1.2.2. Một số kết quả, tính chất đối với xích Markov rời rạc và thuần nhất

Giả sử $\{X_n, n = 0, 1, 2, \dots\}$ là xích Markov rời rạc và thuần nhất, như trước đây chúng ta ký hiệu:

$$p_{ij} = P\left[X_{n+1} = j | X_n = i\right], \quad i, j \in E$$

là xác suất chuyển sau một bước (không phụ thuộc vào n);

Ký hiệu:

$$p_{ij}^{(k)} = P\left[X_{k+m} = j | X_m = i\right]; \quad i, j \in E; k \geq 1; m \geq 0$$

và gọi $p_{ij}^{(k)}$ là xác suất chuyển sau k bước:

Do tính thuần nhất, chúng ta cũng có:

$$p_{ij}^{(k)} = P\left[X_{k+m} = j | X_m = i\right] = P\left[X_k = j | X_0 = i\right]$$

- Ma trận $\mathbb{P} = [p_{ij}]_{i,j \in E}$ được gọi là ma trận xác suất chuyển sau một bước.

  Ma trận $\mathbb{P}$ có tính chất: $\begin{cases} 0 \leq p_{ij} \leq 1 \\ \sum\limits_{j \in E} p_{ij} = 1 \end{cases}$.

- Ma trận $\mathbb{P}^{(k)} = [p_{ij}^{(k)}]_{i,j\in E}$ được gọi là ma trận xác suất chuyển sau k bước.

- Phương trình

$$p_{ij}^{(n+m)} = \sum_{l\in E} p_{il}^{(n)} \cdot p_{lj}^{(m)}; \quad i,j \in E; n,m \geq 1$$

  được gọi là phương trình Chapman-Kolmogorov.

- Phân phối hữu hạn chiều được tính bằng công thức:

$$P\left[X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}, X_n = i_n\right] = p_{i_0} \cdot p_{i_0 i_1} \cdot \cdots \cdot p_{i_{n-1} i_n},$$

  trong đó, $p_{i_0} = P\left[X_0 = i_0\right]$

- Phân phối của xích Markov tại thời điểm n được tính theo công thức:

$$p_j^{(n)} = P[X_n = j], n = 0, 1, 2, \ldots; j \in E.$$

  Ký hiệu véc tơ hàng:

$$\Pi^{(n)} = \left(p_j^{(n)}, j \in E\right)$$

  và gọi $\Pi = \Pi^{(0)}$ là phân phối ban đầu của xích Markov.

  Ta có các kết quả (tính chất) sau:

$$\mathbb{P}^{(n+1)} = \mathbb{P} \cdot \mathbb{P}^{(n)} = \mathbb{P}^{(n)} \cdot \mathbb{P}$$

$$\mathbb{P}^{(n+m)} = \mathbb{P}^{(n)} \cdot \mathbb{P}^{(m)} = \mathbb{P}^{(m)} \cdot \mathbb{P}^{(n)}$$

  (từ đó dễ thấy $\mathbb{P}^{(k)} = \mathbb{P}^k$)

  và ta cũng có:

$$\Pi^{(n)} = \Pi \cdot \mathbb{P}^{(n)}$$

$$\Pi^{(n+1)} = \Pi^{(1)} \cdot \mathbb{P}^{(n)} = \Pi^{(n)} \cdot \mathbb{P}$$

$$\Pi^{(n+m)} = \Pi^{(n)} \cdot \mathbb{P}^{(m)}$$

- Phân phối ban đầu được gọi là dừng nếu $\Pi^{(n)}$ không phụ thuộc vào n, nghĩa là $\Pi = \Pi^{(n)}$, hay tương đương $\Pi = \Pi \cdot \mathbb{P}$.

Các kết quả nghiên cứu sâu hơn của xích Markov rời rạc thuần nhất như nghiên cứu về phân phối dừng, phân phối giới hạn, tính ergodic, phân lớp trạng thái,..., chúng ta không đưa ra ở đây vì không thuộc phạm vi nội dung báo cáo.

Chương 2

Bài toán MTT và biến tiến

2.1. Tổng quan về bài toán MTT

Trong thực tiễn có nhiều lớp mô hình MTT. Ở chương này báo cáo dành để nghiên cứu một lớp MTT có nhiều ứng dụng và thường gặp trong thực tiễn. Điều đặc biệt ở đây là các phương pháp truyền thống và thông dụng đã được công bố để xử lý các mô hình MTT không áp dụng được cho lớp mô hình MTT này. Chúng ta nêu một số ví dụ trong thực tiễn:

Ví dụ 2.1.1. Trong chiến dịch "Điện Biên Phủ trên không" tại Hà Nội tháng 12 năm 1972, do lượng tên lửa SAM-2 hạn chế không đủ đáp ứng yêu cầu chiến đấu. Bộ Tổng tham mưu và và Quân chủng Phòng không - Không quân quyết định tập trung tên lửa SAM-2 và không quân để đánh B-52, các mục tiêu khác (các loại máy bay khác của Mỹ) chỉ đối phó bằng hỏa lực phòng không thông thường. Hệ thống ra đa trong trường hợp này có một hệ thống chức năng phải xử lý mô hình MTT mà trong đó đối tượng quan tâm chỉ là các mục tiêu máy bay B-52; Lớp mục tiêu cần quan tâm này là một lớp con trong lớp tất cả các mục tiêu có thể có đối với hệ thống ra đa phòng không.

Ví dụ 2.1.2. Với hệ thống phòng thủ trong chiến tranh hạt nhân có một hệ thống chức năng trong đó mô hình MTT có nhiệm vụ phải quan tâm xác định các mục tiêu có thể mang đầu đạn hạt nhân (tên lửa mang đầu đạn hạt nhân, máy bay chiến lược mang bom hạt nhân, ...) trong số các mục tiêu có thể có trong không gian phòng thủ.

Như vậy chúng ta gặp một lớp mô hình MTT mà trong đó đối tượng quan tâm theo dõi và xác định không phải là tất cả các mục tiêu mà chỉ là một lớp con trong số đó. Trong trường hợp lớp mục tiêu được quan tâm là lớp con thực sự của lớp tất cả các mục tiêu có thể có, thì các phương pháp, các thuật toán xử lý với MTT đã được công bố không còn phù hợp nữa. Báo cáo đã đề xuất phương pháp tiếp cận bằng cách sử dụng mô hình Markov ẩn để giải quyết lớp mô hình MTT này. Các kết quả theo hướng nghiên cứu đó được trình bày trong báo cáo này.

2.2. Mô hình toán học bài toán MTT

2.2.1. Mô hình toán học bài toán MTT

Giả sử ta quan sát các đối tượng (hay còn gọi là mục tiêu) di động nào đó trong một miền không gian và trong một khoảng thời gian nào đó. Ký hiệu $\mathcal{R}$ là miền không gian mà ta cần quan tâm, ở đây $\mathcal{R} \subset \mathbb{R}^{n_x}$, với $\mathbb{R}^{n_x}$ là không gian trạng thái của mục tiêu, n_x là số chiều của véc tơ trạng thái của mục tiêu. $\mathcal{R}$ được gọi là miền quan sát.

Ký hiệu $[1, T]$, $T > 1$, $T \in \mathbb{R}^+$, là khoảng thời gian mà ta cần quan tâm. $[1, T]$ được gọi là khoảng thời gian của quá trình quan sát. Do các thời điểm quan sát: t_1, t_2, ..., t_n; $1 = t_1 < t_2 < \ldots < t_n = T$, là rời rạc, nên không mất tính tổng quát, khi nói đến thời điểm thứ i (t_i), chúng ta có thể quy ước: $T \in \mathbb{Z}^+$, $t_i \in \mathbb{Z}^+$ và đồng nhất $t_i = i$, $i = 1, 2, \ldots, T$; trong đó, $t_1 = 1$ là lần quan sát đầu tiên và $t_n = T$ là lần quan sát cuối cùng của quá trình quan sát.

Các mục tiêu xuất hiện, biến mất một cách ngẫu nhiên. Các mục tiêu xuất hiện ở vị trí ngẫu nhiên có phân bố đều trong $\mathcal{R}$. Các mục tiêu xuất hiện chuyển động và biến mất một cách độc lập với nhau. Các mục tiêu giả FA được xem như là một loại mục tiêu và cũng tuân theo các giả thiết (về mục tiêu) như các mục tiêu khác.

Ký hiệu X_t^k, $t \in \left[t_i^k, t_f^k\right]$, $k = 1, 2, \ldots$ là trạng của mục tiêu thứ k tại thời điểm t (t_i^k và t_f^k là thời điểm xuất hiện và biến mất tương ứng).

Ký hiệu $\mathcal{M}$ là lớp mục tiêu mà mô hình MTT quan tâm.

Số mục tiêu cần quan tâm có trong $\mathcal{R}$ tại thời điểm t ký hiệu là M_t, khi đó ta có:

$$M_t = M_t(\omega) = Card\{X_t^k | k \in \mathcal{M}\}.$$

Số mục tiêu không thuộc lớp mục tiêu quan tâm có trong $\mathcal{R}$ tại thời điểm t ký hiệu là G_t, khi đó:

$$G_t = G_t(\omega) = Card\{X_t^s | s \notin \mathcal{M}\}.$$

Một cách hợp lý và tự nhiên, chúng ta giả thiết:

+ $M_t = M_t(\omega)$ là biến ngẫu nhiên Poisson với tham số λ_m, $\lambda_m > 0$.

+ Các X_t^k, $k \in \mathcal{M}$, xuất hiện với xác suất p_m, $0 < p_m < 1$.

+ $G_t = G_t(\omega)$, là biến ngẫu nhiên Poisson với tham số λ_g, $\lambda_g > 0$.

+ Các X_t^s, $s \notin \mathcal{M}$, xuất hiện với xác suất p_g, $0 < p_g < 1$ (trong các mô hình thực tế thường $p_g \neq p_m$ và thậm chí $p_g \gg p_m$).

Lưu ý: Giá trị p_m và λ_m (tương ứng p_g và λ_g) có mối liên hệ toán học do định lý giới hạn địa phương (phân bố $B\,[n, p]$ dẫn tới phân bố $P(\lambda)$ khi $n \to \infty$).

Ký hiệu:

$$Y(t) = \left\{Y_t^j \mid j = 1, 2, \ldots, n_t\right\}$$

là tập các giá trị quan sát được tại thời điểm t, $t = t_1, t_2, \ldots, t_n$; n_t là số lượng các kết quả quan sát được tại thời điểm t, chính xác về mặt toán học:

$$n_t = Card\Big(Y(t)\Big)$$

là một biến ngẫu nhiên và

$$n_t = n_t(\omega) = M_t(\omega) + G_t(\omega)$$

Vì $M_t(\omega)$ và $G_t(\omega)$ độc lập, $M_t(\omega) \simeq P(\lambda_m)$, $G_t(\omega) \simeq P(\lambda_g)$, nên

$$n_t = n_t(\omega) \simeq P(\lambda_m + \lambda_g).$$

Mỗi giá trị quan sát có thể là giá trị quan sát thu được từ mục tiêu nào đó hoặc có thể là giá trị quan sát do mục tiêu giả gây ra.

Yêu cầu của bài toán MTT là: Hãy xác định số lượng mục tiêu có trong lớp $\mathcal{M}$ tại mỗi thời điểm t trong miền thời gian quan sát trong $\mathcal{R}$, nghĩa là xác định $M_t(\omega)$.

2.2.2. Mô hình xấp xỉ

Trong mô hình toán học của bài toán MTT, về mặt toán học biến ngẫu nhiên n_t có phân phối Poisson với tham số $(\lambda_m + \lambda_g)$. Do đó, tập giá trị của n_t là $N = \{0, 1, 2, \dots\}$.

Khi đó ta có:

$$\lim_{k \to +\infty} P\left[n_t = k\right] = \lim_{k \to +\infty} \frac{(\lambda_m + \lambda_g)^k}{k!} \cdot e^{-(\lambda_m + \lambda_g)} = 0.$$

Do vậy, với $\varepsilon > 0$ tùy ý bé, tồn tại $N^* = N(\varepsilon) \in N^+$ sao cho

$$P\left[n_t \leq N(\varepsilon)\right] \geq 1 - \varepsilon, \ \forall t \in [1, T]$$

Hoàn toàn tương tự, vì M_t có phân phối Poisson với tham số λ_m, $\lambda_m > 0$ nên:

$$\lim_{k \to +\infty} P\left[M_t = k\right] = \lim_{k \to +\infty} \frac{(\lambda_m)^k}{k!} \cdot e^{-\lambda_m} = 0$$

Nên với $\varepsilon > 0$ tùy ý bé, tồn tại $M^* = M(\varepsilon) \in N^+$ sao cho:

$$P\left[M_t \leq M(\varepsilon)\right] \geq 1 - \varepsilon, \ \forall t \in [1, T]$$

Ta đưa ra giả thiết sau đây:

Giả thiết 2.2..1. $\exists\ M^*, N^* \in \mathbb{N}^+$ *sao cho:*

- $M_t(\omega) \leq M^* \pmod P, \ \forall t \in [1, T]$
- $n_t(\omega) \leq N^* \pmod P, \ \forall t \in [1, T]$

Lưu ý: Tùy theo độ chính xác yêu cầu mà M^* và N^* được chọn tương ứng. Song M^* và N^* càng chọn với giá trị lớn thì độ phức tạp tính toán trong lời giải càng cao.

Bài toán MTT được phát biểu trong mục 2.2.1 với điều kiện tuân theo Giả thiết 2.2.1 được gọi là mô hình xấp xỉ hay mô hình gần đúng. Mô hình này sẽ là đối tượng quan tâm nghiên cứu trong báo cáo này.

Chương 3

Xây dựng thuật toán tiến cho mô hình HMM không thuần nhất

3.1. Mô hình Markov ẩn HMM (Hidden Markov Model)

Xét một quy luật ngẫu nhiên nào đó thay đổi trạng thái theo thời gian. Số trạng thái của quy luật ngẫu nhiên này là hữu hạn và không quan sát được, nhưng biết rằng mỗi trạng thái đó gắn liền với một hiện tượng ngẫu nhiên khác quan sát được. Về mặt toán học, chúng ta có thể toán học hóa công việc đó như sau:

- Xét một quá trình ngẫu nhiên S-giá trị X_t; S ở đây được giả thiết là $Card(S) = M < +\infty$. Không mất tính tổng quát, chúng ta ký hiệu:

  $$S = \{S_1, S_2, ..., S_M\}$$

  S sẽ được gọi là không gian trạng thái của quá trình X_t (X_t là quá trình nhận giá trị trên S; và sau này liên quan đến giả thiết Markov thì chúng ra sẽ xét X_t là quá trình Markov).

- Ký hiệu q_t là trạng thái của quá trình X_t tại thời điểm t, $t \in [1, T]$. Khi đó, q_t nhận giá trị trong S. Các trạng thái $q_t = S_j$, $1 \leq j \leq M$, $S_j \in S$, là không quan sát được nhưng nó có liên kết mật thiết (phép tương ứng "1-1" hay có thể gọi là phép nhúng) với một biến ngẫu nhiên quan sát được V-giá trị với $Card(V) = N < +\infty$. Không mất tính tổng quát, chúng ta giả sử rằng:

$$V = \{v_1, v_2, ..., v_N\}$$

Các phần tử v_k, $v_k \in V$, $1 \leq k \leq N$, là quan sát được và V được gọi là không gian các giá trị quan sát. Ký hiệu O_t là giá trị quan sát được của hệ thống ở trạng thái S_j tại thời điểm t, $t \in [1, T]$, thì phân phối của biến ngẫu nhiên quan sát được V-giá trị tương ứng với trạng thái $q_t = S_j$ cho bởi phân phối

$$B = \{b_j(k)\}$$

với $$b_j(k) = P\left[O_t = v_k | q_t = S_j\right],\ 1 \leq k \leq N,\ 1 \leq j \leq M,$$

trong đó,

$$b_j(k) \geq 0; \quad 1 \leq k \leq N, \quad 1 \leq j \leq M;$$
$$\sum_{k=1}^{N} b_j(k) = 1, \quad 1 \leq j \leq M.$$

- Các nghiên cứu về Mô hình Markov ẩn (HMM), cho đến thời điểm hiện tại, được công bố đều giả thiết rằng HMM phải thỏa mãn ba giả thuyết sau đây:

  a/ Giả thuyết về tính Markov

  Quá trình ẩn X_t có tính Markov, có nghĩa là:

  $$P\left[q_{t+1} = S_j | q_1 = S_{i_1}, q_2 = S_{i_2}, \ldots, q_{t-1} = S_{i_{t-1}}, q_t = S_i\right] =$$
  $$= P\left[q_{t+1} = S_j | q_t = S_i\right] = a_{ij}\left(t\right)$$

  và $a_{ij}\left(t\right)$ được gọi là xác suất chuyển từ trạng thái S_i sang trạng thái S_j ở thời điểm (bước thứ) t.

  b/ Giả thuyết về tính dừng (thuần nhất)

  $$P\left[q_{t+1} = S_j | q_t = S_i\right] = P\left[q_{l+1} = S_j | q_l = S_i\right] = a_{ij}, \quad \forall t, l.$$

  Hay nói một cách khác $$a_{ij}\left(t\right) \equiv a_{ij}, \quad \forall t.$$

  a_{ij} được gọi là xác suất chuyển trạng thái sau một bước, $A := [a_{ij}]_{1 \leq i,j \leq M}$ được gọi là ma trận chuyển trạng thái sau một bước. Lẽ đương nhiên, trong đó:

$$a_{ij} \geq 0, \quad 1 \leq i, j \leq M$$
$$\sum_{j=1}^{M} a_{ij} = 1, \quad 1 \leq i \leq M$$

c/ Giả thuyết về tính độc lập quan sát

Giả thiết này yêu cầu về mặt thống kê quan sát hiện tại độc lập với các quan sát trước đó. Về mặt toán học điều đó có nghĩa là: Giả sử với HMM có dãy quan sát $O = O_1O_2...O_T$ (ứng với dãy trạng thái $Q = q_1q_2...q_T$) thì:

$$P[O|Q, HMM] = \prod_{t=1}^{T} P[O_t|q_t, HMM]$$

- Với HMM người ta cần phải xác định phân bố trạng thái ban đầu của mô hình. Ký hiệu: $\Pi = \{\pi_i, 1 \leq i \leq M\}$ với $\pi_i = P[q_1 = S_i]$ và gọi Π là phân phối trạng thái ban đầu.

  Trong đó:

$$\pi_i > 0, \quad 1 \leq i \leq M$$
$$\sum_{i=1}^{M} \pi_i = 1$$

Với những điều bổ trợ đã đưa vào như trên, bây giờ chúng ta mô tả thành phần cấu trúc, quá trình hoạt động và những vấn đề (bài toán) cơ bản của HMM.

Các thành phần của HMM

1/ Tham số chỉ số lượng trạng thái của mô hình

Xét xích Markov ẩn X_t hữu hạn trạng thái như đã mô tả ở trên có M trạng thái. Khi đó, M được gọi là tham số chỉ số lượng trạng thái của mô hình Markov ẩn. Mặc dù các trạng thái là ẩn song một cách tổng quát giữa các các trạng thái trong quá trình mô hình hoạt động theo thời gian có tồn tại một quy luật mà mô hình chuyển từ trạng thái này sang trạng thái khác. Về mặt toán học, quy luật đó là quy luật thay đổi theo thời gian của quá trình ẩn X_t. Đặc biệt nếu yêu cầu một trạng thái bất kỳ có thể đạt được từ bất kỳ trạng thái khác thì đó là mô hình có tính ergodic. Cũng như trên, chúng ta ký hiệu tập các trạng thái riêng biệt là

$$S = \{S_1, S_2, ..., S_M\}$$

và q_t là trạng thái của mô hình ở thời điểm t, $t \in [1, T]$. Khi đó, q_t nhận giá trị trong S, và S được gọi là không gian trạng thái.

2/ Tham số chỉ số lượng các giá trị quan sát

Ký hiệu N là số các biểu tượng (giá trị) quan sát có thể có liên quan đến các trạng thái của mô hình. Chúng ta ký hiệu tập các giá trị quan sát riêng biệt là:

$$V = \{v_1, v_2, ..., v_N\}.$$

Quan sát thu được tại thời điểm t ký hiệu là O_t. Như vậy, O_t nhận giá trị trong V và tập V được gọi là không gian các giá trị quan sát.

3/ Phân phối xác suất chuyển trạng thái

$$A := [a_{ij}]_{1 \le i,j \le M}$$

trong đó,
$$a_{ij} = P\left[q_{t+1} = S_j | q_t = S_i\right],\ 1 \le i, j \le M.$$

Với HMM mà một trạng thái bất kỳ có thể đạt được từ một trạng thái bất kỳ khác chỉ sau một bước thì $a_{ij} > 0$, $\forall i, j$. Trong trường hợp ngược lại chúng ta có $a_{ij} = 0$ với một hay một số cặp (i, j). Lẽ đương nhiên yêu cầu A phải là ma trận ngẫu nhiên, nghĩa là:

$$\sum_{j=1}^{M} a_{ij} = 1$$

Trong báo cáo, để tường minh hơn, liên quan đến trạng thái, cùng với ký hiệu

$$a_{ij} = P\left[q_{t+1} = S_j | q_t = S_i\right]$$

đôi khi còn ký hiệu
$$a_{q_t q_{t+1}} = P\left[q_{t+1} = S_j | q_t = S_i\right]$$

có nghĩa là cùng với a_{ij} ta cũng còn dùng $a_{q_t q_{t+1}}$ và chúng có cùng ý nghĩa như biểu thức nêu trên.

4/ Phân phối xác suất của các quan sát khi hệ thống ở trạng thái: S_j, $1 \le j \le M$ là $B := \{b_j(k), 1 \le k \le N\}$, trong đó: $b_j(k) = P\left[O_t = v_k | q_t = S_j\right]$, $1 \le k \le N$, $1 \le j \le M$.

5/ *Phân phối của trạng thái ban đầu* $\Pi = \{\pi_i,\ 1 \le i \le M\}$ trong đó, $\pi_i = P\left[q_1 = S_i\right],\ 1 \le i \le M$.

Để tiện trong trình bày, ký hiệu bộ ba phân phối A, B, Π là Λ:

$$\Lambda = (A, B, \Pi)$$

Như vậy, khi xác định được M, N và Λ có nghĩa là ta đã xác định được một mô hình Markov ẩn và HMM này cũng dùng ký hiệu là Λ.

Vận hành của HMM theo thời gian

Giả sử quan tâm trong miền thời gian $[1, T]$, $T > 1$, $T \in N^+$.

Giả sử đã cho một mô hình Markov ẩn (HMM) Λ. Khi đó,

1/ Λ xác định trạng thái ban đầu $q_1 = S_i$ theo phân phối Π.

2/ Đặt $t = 1$.

3/ Λ xác định quan sát $O_t = v_k$ theo phân phối B (cụ thể là dòng thứ i: $\{b_i(k), 1 \le k \le N\}$)

4/ Chuyển trạng thái từ q_t sang $q_{t+1} = S_j$ theo phân phối xác suất A.

5/ Đặt $t = t + 1$ và quay lại Bước 3

Quá trình sẽ dừng lại khi $t = T$.

Như vậy, HMM cho ra một dãy quan sát: $O = O_1 O_2 \cdots O_T$ với T là số quan sát.

Các vấn đề cơ bản khi nghiên cứu HMM

Mô hình HMM mà báo cáo vừa trình bày là mô hình HMM với thời gian rời rạc, hữu hạn trạng thái và không gian các giá trị quan sát hữu hạn và có tên gọi là Mô hình Markov ẩn rời rạc.

Các công trình nghiên cứu về HMM được công bố cho đến thời điểm hiện tại có thể chia thành 3 hướng sau:

+ Hướng mở rộng sang những lớp mô hình Markov ẩn mới như: HMM ergodic, HMM dạng Bakis; HMM trái-phải; HMM có mật độ quan sát liên tục, HMM tự hồi quy,....

+ Hướng nghiên cứu một vài dạng biến thể trên những cấu trúc HMM. Hướng này chủ yếu với mục tiêu áp dụng cho những mô hình toán học cần giải quyết cụ thể.

+ Hướng nghiên cứu giải quyết ba bài toán cơ bản đối với HMM.

Báo cáo xin trình bày cụ thể hơn về hướng thứ ba này.

Khi cho một mô hình HMM là Λ và cho một dãy quan sát sinh ra bởi HMM đó là: $O = O_1 O_2 \cdots O_T$.

Có ba bài toán cơ bản được đặt ra là:

+ *Bài toán 1:* Cho dãy quan sát $O = O_1 O_2 \cdots O_T$ và mô hình Markov ẩn Λ. Hãy tính $P(O|\Lambda)$.

+ *Bài toán 2:* Cho dãy quan sát $O = O_1 O_2 \cdots O_T$ và mô hình Markov ẩn Λ. Hãy xác định dãy trạng thái tương ứng $Q = q_1 q_2 ... q_T$ tối ưu nhất.

+ *Bài toán 3:* Cho dãy quan sát $O = O_1 O_2 \cdots O_T$. Làm thế nào để điều chỉnh các tham số của mô hình $\Lambda = (A, B, \Pi)$ sao cho với mô hình được điều chỉnh thì $P(O|\Lambda)$ đạt cực đại.

Trong phạm vi nghiên cứu của báo cáo, chúng ta quan tâm đến Bài toán 1 và Bài toán 2.

Bài toán 1 là bài toán đánh giá, tính xác suất. Với các công trình đã được công bố để giải bài toán này thì người ta thường dùng “thuật toán tiến-lùi” (Forward-Backward Algorithm).

Bài toán 2 là bài toán mà trong đó ta tìm ra phần ẩn của mô hình. Với các công trình đã được công bố để giải bài toán này thì người ta sử dụng thuật toán Viterbi (Viterbi Algorithm).

Với mục đích là xây dựng công cụ để giải bài toán MTT được phát biểu ở mục 3.2.2 thì các thuật toán tiến-lùi (Forward-Backward Algorithm) và thuật toán Viterbi (Viterbi Algorithm) đã được công bố là không thể áp dụng được bởi lẽ: đối với bài toán MTT thì dãy số liệu quan sát tiến dần theo thời gian và đến đâu thì phải xử lý ngay tới đó. Tại thời điểm xử lý hiện tại chỉ có dãy số liệu qúa khứ tiến dần theo thời gian cho đến thời điểm hiện tại chứ không thể có số liệu quan sát trước của tương lai. Do đó biến lùi là không tồn tại và vô nghĩa.

Mặt khác đối với HMM các giả thiết a/ (giả thiết về tính Markov) và c/ (giả thiết về tính độc lập thống kê của các quan sát) là bắt buộc không thể bỏ qua. Nhưng giả thiết b/ (giả thiết về tính thuần nhất) có thể mở rộng ra cho trường hợp không thuần nhất. Sự mở rộng này chỉ ảnh hưởng đến việc giải bài toán cơ bản số 3 (Bài toán 3). Sự ảnh hưởng này liên quan trực tiếp đến cơ chế điều chỉnh HMM và thủ tục học máy (machine learning) trong ứng dụng HMM vào bài toán nhận dạng. Đối với mục đích giải bài toán MTT chúng ta chỉ cần dùng đến Bài toán 1 và Bài toán 2 mà thôi. Bởi những lý do đó báo cáo sẽ trình bày một số kết quả đóng góp mới trong lý thuyết HMM.

3.2. Xây dựng thuật toán tiến

Trong mục này báo cáo xây dựng thuật toán tiến (Forward Algorithm) và thuật toán Viterbi cải tiến (Modified Viterbi Algorithm). Thuật toán Viterbi cải tiến thực chất là thuật toán Viterbi nhưng chỉ sử dụng biến tiến và thuật toán tiến. Cả hai thuật toán này được xây dựng trên HMM không thuần nhất. Lẽ đương nhiên HMM thuần nhất là trường hợp riêng của HMM không thuần nhất.

Xét HMM với các thành phần cấu trúc như đã nêu trong mục 3.3, nhưng phân phối xác suất chuyển trạng thái phụ thuộc vào thời gian (không thuần nhất theo thời gian). Nghĩa là,

$$P\left[q_{t+1} = S_j | q_t = S_i\right] = a_{ij}(t)$$

Do HMM rời rạc nên các mốc chuyển trạng thái là các mốc thời gian rời rạc nghĩa là: $t_1 = 1 < t_2 < t_3 < ... < t_n = T$, bởi vậy để tiện lợi trong trình bày cùng với $a_{ij}(t_k)$ chúng ta có thể dùng gọn hơn là $a_{ij}(k)$.

Trong mục tiếp sau (mục 3.5) giá trị $a_{ij}(k)$ sẽ được chỉ ra công thức giải tích cụ thể đối với HMM dùng cho bài toán MTT.

Ký hiệu:

$$A(k) = \left[a_{ij}(k)\right]_{1 \le i,j \le M}, \; k = \overline{1, n-1}$$

$$\mathcal{A} = \{A(1), A(2), \cdots, A(n-1)\}$$

Mô hình Markov ẩn $\Lambda = (\mathcal{A}, B, \Pi)$ được gọi là mô hình Markov ẩn không thuần nhất.

Với HMM không thuần nhất chúng ta có các kết quả sau đây:

3.2.1. Bài toán cơ bản thứ nhất và thuật toán tiến

Xét mô hình Markov ẩn không thuần nhất $\Lambda = (\mathcal{A}, B, \Pi)$.

Vì Λ là mô hình HMM rời rạc, nên các thời điểm t được xét trong $[1, T]$ là các thời điểm rời rạc $1 = t_1 < t_2 < \cdots < t_k < \cdots < t_n = T$. Bởi vậy, với $t = t_k$ thì cùng với quan sát O_{t_k} đôi khi ra ký hiệu gọn là O_k; cùng với trạng thái q_{t_k} đôi khi ta ký hiệu gọn là q_k.

Tại thời điểm $t = t_k$ bất kỳ, $t_k \in [1, T]$, với Λ ta có dãy quan sát

$$O = O_1 O_2 \cdots O_t \tag{3.1}$$

Bài toán cơ bản thứ nhất là đi tính xác suất của dãy quan sát O khi HMM không thuần nhất Λ đã cho. Nghĩa là tính $P(O|\Lambda)$.

Công thức giải và thuật toán giải bài toán 1 được cho thông qua bổ đề và thuật toán sau đây:

Bổ đề 3.2..1.

$$P(O|\Lambda) = \sum_{\forall (q_1 q_2 \ldots q_k)} \left\{ \prod_{s=1}^{k} a_{q_{s-1} q_s}(s-1)\, b_{q_s}(O_s) \right\} \tag{3.2}$$

trong đó ký hiệu hình thức:

$$a_{q_0 q_1}(0) = \pi_{q_1}.$$

Chứng minh. Xét một dãy trạng thái bất kỳ cố định nào đó:

$$Q = q_1 q_2 \ldots q_k \tag{3.3}$$

ở đây, q_1 là trạng thái ban đầu.

Khi đó, xác suất của dãy quan sát (3.1) khi cho dãy trạng thái (3.3) sẽ là:

$$P(O|Q, \Lambda) = \prod_{s=1}^{k} P(O_s|q_s, \Lambda). \tag{3.4}$$

Trong (3.4) đã sử dụng giả thiết độc lập thống kê của dãy quan sát (3.1), từ đó ta có:

$$P(O|Q,\Lambda) = \prod_{s=1}^{k} P(O_s|q_s,\Lambda) = \prod_{s=1}^{k} b_{q_s}(O_s). \tag{3.5}$$

Mặt khác: $P(Q|\Lambda) = \pi_{q_1}\, a_{q_1q_2}(1)\, a_{q_2q_3}(2)...a_{q_{k-1}q_k}(k-1)$, với ký hiệu: $a_{q_0q_1}(0) := \pi_{q_1}$, ta có:

$$P(Q|\Lambda) = \prod_{s=1}^{k} a_{q_{s-1}q_s}(s-1) \tag{3.6}$$

Từ công thức (3.5) và (3.6) ta có:

$$\begin{aligned} P(O|\Lambda) &= \sum_{\forall Q} P(O|Q,\Lambda)\cdot P(Q|\Lambda) \\ &= \sum_{\forall Q}\left\{\prod_{s=1}^{k} a_{q_{s-1}q_s}(s-1)\cdot b_{q_s}(o_s)\right\} \\ &= \sum_{\forall(q_1q_2...q_k)}\left\{\prod_{s=1}^{k} a_{q_{s-1}q_s}(s-1)\cdot b_{q_s}(o_s)\right\} \end{aligned}$$

Bổ đề được chứng minh. □

Công thức (3.2) có thể diễn giải tường minh như sau: Ở thời điểm ban đầu $t_1(s=1)$, mô hình ở trạng thái $q_1 = q_{t_1}$ với xác suất $\Pi_{q_1}(=: a_{q_0q_1}(0))$ và tại thời điểm này với trạng thái q_1 đó chúng ta quan sát được giá trị quan sát $O_1 = O_{t_1}$ với xác suất $b_{q_1}(O_1)$. Sự thay đổi từ t_s tới t_{s+1} (vắn tắt từ s sang $s+1$: $(s=2)$) mô hình chuyển từ trạng thái $q_1 = q_{t_1}$ sang trạng thái $q_2 = q_{t_2}$ với xác suất $a_{q_1q_2}(1)$ và ở trạng thái q_2 này chúng ta quan sát được giá trị quan sát $O_2 = O_{t_2}$ với xác suất $b_{q_2}(O_2)$. Quá trình cứ tiếp tục cho tới cuối khi từ t_{k-1} chuyển sang thời điểm $t_k = t$ $(s = k-1)$ mô hình chuyển từ trạng thái $q_{k-1} = q_{t_{k-1}}$ sang trạng thái $q_k = q_{t_k}$ với xác suất $a_{q_{k-1}q_k}(k-1)$ và ở trạng thái q_k này chúng ta quan sát được giá trị quan sát $O_k = O_{t_k}$ với xác suất $b_{q_k}(O_k)$.

Để tính công thức (3.2) chúng ta thấy rằng nếu tính trực tiếp thì độ phức tạp của thuật toán có cấp của $2tM^t$ phép toán. Trong HMM thuần nhất người ta đưa ra thuật toán tiến-lùi (Forward-Backward Algorithm) để tính công thức (3.2) với $t = T$. Đối với HMM với mục tiêu áp dụng để giải bài toán MTT thì thuật toán đó không dùng được. Ở đây báo cáo đưa ra thuật toán gọi là thuật toán tiến (Forward Algorithm) để tính công thức (3.2) với $\forall t \in [1, T]$ như sau.

- **Thuật toán tiến**

Ký hiệu $\alpha_\tau(i) = P(O_1 O_2 \cdots O_\tau; q_\tau = S_i|\Lambda)$; nghĩa là $\alpha_\tau(i)$ là xác suất của phần đầu của dãy quan sát cho đến thời điểm t_τ và tại thời điểm t_τ đó, trạng thái $q_\tau = q_{t_\tau} = S_i$, $S_i \in S$. Khi đó $\alpha_\tau(i)$ được gọi là biến tiến.

Xác suất $P(O|\Lambda)$ được tính theo biến tiến $\alpha_\tau(i)$ theo thủ tục quy nạp như sau:

1/ Bước khởi đầu:

$$\alpha_1(i) = \pi_i \, b_i\,(O_1), 1 \leq i \leq M.$$

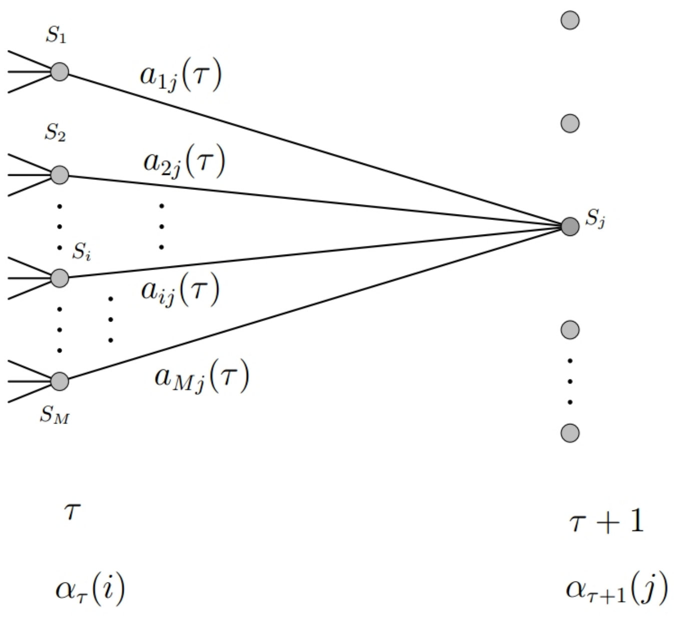
2/ Bước quy nạp:

$$\alpha_{\tau+1}(j) = \left[\sum_{i=1}^{M} \alpha_\tau(i)\, a_{ij}(\tau)\right] \cdot b_j\,(O_{\tau+1});$$
$$1 \leq \tau \leq t-1 = t_k - 1,\ 1 \leq j \leq M$$

3/ Kết thúc:

$$P(O|\Lambda) = \sum_{i=1}^{M} \alpha_t(i) = \sum_{i=1}^{M} \alpha_{t_k}(i)$$

Cốt lõi và trung tâm của thuật toán là bước quy nạp. Sự tính toán ở bước quy nạp được tính toán mô tả qua Hình vẽ 2.1:



Hình 3.1: Lưới tính biến tiến

Hình 2.1 chỉ ra cách mà trạng thái S_j đạt được tại thời điểm $\tau+1$ từ M trạng thái có thể S_i, $1 \leq i \leq M$, ở thời điểm τ. Vì $\alpha_\tau(i)$ là xác suất của biến cố mà dãy quan sát đầu là $O_1O_2\cdots O_\tau$ và trạng thái của HMM ở thời điểm τ là S_i. Tích số $\alpha_\tau(i)\cdot a_{ij}(\tau)$ là xác suất của biến cố mà $O_1O_2\cdots O_\tau$ là dãy đầu của các quan sát và trạng thái S_j đạt được tại thời điểm $(\tau+1)$ theo con đường đi qua trạng thái S_i ở thời điểm τ. Tổng của tất cả các tích này theo tất cả M trạng thái S_i, $1 \leq i \leq M$, có thể có ở thời điểm τ chúng ta thu được xác suất của S_j ở thời điểm $\tau+1$ với tất cả các quan sát bộ phận đi kèm trước đó. Từ đó ta tính được xác suất $\alpha_{\tau+1}(j)$ với hệ ở trạng thái S_j và giá trị quan sát là $O_{\tau+1}$ ở thời điểm $\tau+1$.

Từ sự phân tích và minh họa trên, chúng ta đã chỉ ra tính đúng đắn của thuật toán.

3.2.2. Bài toán cơ bản thứ hai và thuật toán Viterbi cải tiến

Bài toán cơ bản thứ 2 đối với HMM mục đích là phát hiện ra phần ẩn của mô hình nghĩa là đi tìm dãy trạng thái hợp lý nhất, dãy trạng thái tối ưu tương ứng với dãy quan sát đã cho.

Vấn đề quan trọng đầu tiên là tiêu chuẩn thế nào là hợp lý nhất? Thế nào là tối ưu?

Có hai dạng yêu cầu như sau:

- *Dạng 1 :*

  Cho dãy quan sát: $O = O_1O_2\cdots O_t$ sinh ra bởi HMM không thuần nhất Λ. Hãy tìm trạng thái $q_t = q_t^*$ tối ưu theo nghĩa cực đại xác suất

- *Dạng 2:*

  Cho dãy quan sát: $O = O_1O_2\cdots O_t$ sinh ra bởi HMM không thuần nhất Λ. Hãy tìm dãy trạng thái $Q^* = q_1^*q_2^*\cdots q_t^*$ của Λ tối ưu theo nghĩa cực đại xác suất

a/ Phương pháp tìm lời giải cho dạng 1.

Đây là bài toán tìm trạng thái tối ưu riêng biệt $q_t = q_t^*$ tại thời điểm hiện tại t.

Chúng ta xây dựng biến:

$$\gamma_t(i) = P(q_t = S_i | O, \Lambda). \tag{3.7}$$

Dễ dàng biểu diễn $\gamma_t(i)$ qua biến tiến $\alpha_t(i)$ theo công thức sau:

$$\gamma_t(i) = \frac{\alpha_t(i)}{P(O|\Lambda)} \tag{3.8}$$

và $\gamma_t(i)$ là một phân phối xác suất

$$\gamma_t(i) \geq 0; \quad \sum_{i=1}^{M} \gamma_t(i) = 1.$$

Từ công thức (3.7) chúng ta thu được lời giải:

$$q_t^* = arg \max_{1\leq i\leq M} \gamma_t(i). \tag{3.9}$$

- **Thuật toán Viterbi cải tiến 1**

Theo thuật toán tiến của mục 3.4.1. chúng ta dễ dàng dùng thuật toán tiến để thu được $\gamma_t(i)$ thông qua công thức (3.8) và từ đó thu được q_t^* qua công thức (3.9).

b/ Phương pháp tìm lời giải cho dạng 2.

Để tìm ra dãy trạng thái tốt nhất $Q^* = q_1^* q_2^* \cdots q_t^*$ khi cho trước dãy quan sát $O = O_1O_2...O_t$ của Λ, báo cáo đề xuất thuật toán sau đây và gọi là thuật toán Viterbi cải tiến 2 đối với HMM không thuần nhất. Sở dĩ gọi là "thuật toán Viterbi cải tiến" vì về mặt kỹ thuật khá tương đồng với thuật toán Viterbi đã được công bố đối với HMM thuần nhất, song nó chỉ sử dụng thuật toán tiến và biến tiến.

Định nghĩa đại lượng :

$$\delta_\tau(i) = \max_{q_1q_2\cdots q_{\tau-1}} P(q_1q_2...q_{\tau-1}q_\tau, q_\tau = S_i; O_1O_2\cdots O_\tau|\Lambda) \tag{3.10}$$

nghĩa là $\delta_\tau(i)$ là xác suất lớn nhất dọc theo dãy trạng thái đơn cho đến thời điểm τ và kết thúc ở τ tại trạng thái S_i. Từ (3.13) chúng ta có công thức quy nạp cho $\delta_\tau(j)$ theo công thức sau:

$$\delta_\tau(j) = \left\{ \max_{1\leq i\leq M} \delta_{\tau-1}(i) \cdot a_{ij}(\tau - 1) \right\} \cdot b_j(O_\tau). \tag{3.11}$$

Để tính ra được dãy trạng thái cần tìm trong quá trình quy nạp theo công thức (3.11) ta phải giữ lại đối số (trạng thái) đạt cực đại trong (3.11) đối với mỗi τ và j. Bởi vậy cùng với $\delta_\tau(i)$, thực hiện quy nạp cùng với đại lượng $\psi_t(j)$ như sau:

- **Thuật toán Viterbi cải tiến 2**

1/ Bước khởi tạo:

$$\delta_1(i) = \pi_i \, b_i \, (O_1), \quad 1 \le i \le M.$$
$$\psi_1(i) = 0.$$

2/ Bước quy nạp:

$$\delta_\tau(j) = \left\{ \max_{1 \le i \le M} \delta_{\tau-1}(i) \cdot a_{ij}(\tau - 1) \right\} \cdot b_j(O_\tau), \; 2 \le \tau \le t, \, 1 \le j \le M.$$
$$\psi_\tau(j) = arg \max_{1 \le i \le M} \left\{ \delta_{\tau-1}(i) \cdot a_{ij}(\tau - 1) \right\}, \; 2 \le \tau \le t, \, 1 \le j \le M.$$

3/ Kết thúc:

$$P^* = \max_{1 \le i \le M} \left\{ \delta_t(i) \right\}.$$
$$q_t^* = arg \max_{1 \le i \le M} \left\{ \delta_t(i) \right\}.$$

4/ Truy ngược

$$q_\tau^* = \psi_{\tau+1}(q_{\tau+1}^*), \; \tau = t-1, t-2, \cdots, 1.$$

Kết thúc thuật toán chúng ta xác định được dãy trạng thái tối ưu

$$Q^* = q_1^* q_2^* \cdots q_t^*$$

Kết luận

- Bài toán MTT trong mục 2.2.2. chỉ ước lượng số lượng mục tiêu tại các thời điểm quan tâm. Vấn đề mở ở đây là nếu quan tâm tới quỹ đạo của các mục tiêu đó như trong bài toán MTT tổng quát được nghiên cứu trong chương 2, thì liệu có thể sử dụng công cụ HMM được không? Tác giả cũng đã tiến hành nghiên cứu vấn đề này, song muốn dùng HMM thì các trạng thái của HMM liên quan đến bó ánh xạ từ i điểm tại thời điểm t_{k-1} sang j điểm tại thời điểm t_k (như dạng ánh xạ đệ quy được nêu ra ở Chương 2). Với hướng đi đó việc tính xác suất chuyển trạng thái $a_{ij}(k)$ sẽ phức tạp tương đương với việc tính xác suất hậu nghiệm. Hiện tại chưa có kết quả nào được công bố theo hướng bài toán mở này.

- Một vấn đề nữa được đặt ra là: Lời giải tối ưu theo dạng 1 có nằm trong dãy của lời giải tối ưu theo dạng 2 hay không? Đây là vấn đề rất ý nghĩa song đáng tiếc là tất cả các công trình đã được công bố cho đến thời điểm hiện tại ngay cả đối với HMM thuần nhất cũng chưa có lời giải bởi lẽ lời giải tối ưu chưa chắc đã duy nhất. Đây cũng là hướng nghiên cứu tiếp theo.

Như vậy, Chương 3 của báo cáo đã tập trung nghiên cứu lớp bài toán MTT với yêu cầu ước lượng số lượng mục tiêu của lớp mục tiêu được quan tâm tại mỗi thời điểm. Với lớp bài toán này, báo cáo dùng phương pháp tiếp cận theo hướng sử dụng HMM không thuần nhất. Các kết quả thu được của chương này là:

1/ Đề xuất "thuật toán tiến" và "thuật toán Viterbi cải tiến" đối với HMM không thuần nhất.

2/ Xây dựng HMM tương thích với bài toán MTT được nghiên cứu trong chương 3.

3/ Sử dụng các thuật toán được đề xuất và HMM tương thích để giải bài toán MTT đã nêu.

Tài liệu tham khảo

Tài liệu Tiếng Việt

[1] Nguyễn Văn Hữu, Nguyễn Hữu Dư, *Phân tích thống kê và dự báo*, (2003)

Tài liệu Tiếng Anh

[2] Andrew Gelman, Jonh B. Carlin, Hal S. Stern, Donald B. Rubin, *Bayesian Data Analysis*, Chapman & Hall/CRC (2004).

[3] Cari G. Kaufman, Stephan R. Sain, *Bayesian Functional ANOVA Modeling Using Gaussian Process Prior Distributions* , (2010).

[4] Fabrizio Solari, Brunero Liseo, Dongchu Sun , *Some remarks on Bayesian inference for one-way ANOVA models* ,(2007)

[5] Jean-Michel Marin, Christian P. Robert, *Bayesian Core: A Practical Approach to Computational Bayesian Statistics* , (2007).

[6] Peter D.Hoff, *A First Course in Bayesian Statistical Methods*, (2009)

[7] William M. Bolstad, *Introduction to Bayesian Statistic*, (2007).