Thuật toán Cực đại hóa Kì vọng (EM)

Trần Quốc Long[1]

[1]Bộ môn Khoa học Máy tính
Khoa Công nghệ Thông tin
Trường Đại học Công nghệ

Thứ Tư, 30/03/2016

Nội dung

1 Giới thiệu

2 Thuật toán cực đại hóa kì vọng (EM)

3 Mô hình Markov ẩn

4 Tổng kết

Bài toán ước lượng mật độ (có tham số)

- Xét biến ngẫu nhiên $\mathbf{X}$ trên tập $\mathcal{X}$.
- Ta không biết phân bố thật sự $p(\mathbf{x})$ của $\mathbf{X}$ nhưng ta có dữ liệu là một mẫu lấy từ phân bố $p(\mathbf{x})$
$$\mathbf{x} \sim p(\mathbf{x}).$$
- Xét lớp hàm phân bố $p(\mathbf{x}; \theta), \mathbf{x} \in \mathcal{X}$ phụ thuộc vào tham số $\theta \in \Theta$.
- **Bài toán:** Cho mẫu $\mathbf{x} \sim p(\mathbf{x})$, tìm tham số θ để $p(\mathbf{x}; \theta)$ xấp xỉ $p(\mathbf{x})$.
- Lưu ý: Khi $\mathcal{X} = \mathcal{Z} \times \mathcal{Z} \times \ldots \times \mathcal{Z}$ và $p(\mathbf{x})$ có thể phân tích thành nhân tử
$$p(\mathbf{x}) = p(\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_n) = \prod_{i=1}^{n} p(\mathbf{z}_i), \mathbf{z}_i \in \mathcal{Z}$$
thì $\mathbf{x}$ tương đương với n mẫu học độc lập có cùng phân bố $\mathbf{z}_i, i = \overline{1, n}$.

Phương pháp ước lượng hợp lý cực đại (MLE)

Bài toán

- Bài toán: Cho mẫu $\mathbf{x} \sim p(\mathbf{x})$, tìm tham số θ để $p(\mathbf{x}; \theta)$ xấp xỉ $p(\mathbf{x})$.
- Sự hợp lý của tham số (likelihood function):

$$\mathcal{L}(\theta; x) = p(\mathbf{x}; \theta)$$

  là hàm của θ.
- Ước lượng hợp lý cực đại:

$$\begin{aligned} \boxed{\theta^{\text{MLE}} = \arg\max_{\theta} \mathcal{L}(\theta; x)} \\ = \arg\max_{\theta} \log \mathcal{L}(\theta; x) \end{aligned}$$

Phương pháp ước lượng hợp lý cực đại (MLE)

Ví dụ

- Giả sử n mẫu $z_1, z_2, \ldots, z_n$ độc lập và có cùng phân bố $p(z)$.
- Xét lớp hàm phân bố chuẩn $p(z; \mu, \sigma^2) = \mathcal{N}(z; \mu, \sigma^2)$.
- Sự hợp lý của tham số

$$\mathcal{L}(\underbrace{\mu, \sigma^2}_{\theta}; \underbrace{z_{1:n}}_{\mathbf{x}}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z_i-\mu)^2}{2\sigma^2}}$$

$$\log \mathcal{L}(\mu, \sigma^2; z_{1:n}) = n \log \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (z_i - \mu)^2$$

- Lấy đạo hàm và đặt bằng 0 được

$$\mu = \bar{z} = \frac{1}{n} \sum_{i=1}^{n} z_i; \quad \sigma^2 = \frac{\sum_{i=1}^{n} z_i^2}{n} - \bar{z}^2$$

Nội dung

1 Giới thiệu

2 Thuật toán cực đại hóa kì vọng (EM)

3 Mô hình Markov ẩn

4 Tổng kết

Biến ẩn

- Trong đa số trường hợp, ta không có đầy đủ dữ liệu, một số thông tin đã bị ẩn đi.
- Ta chỉ quan sát được biến $\mathbf{x}$ mà không quan sát được biến $\mathbf{y}$ (ẩn).
- Bài toán: Cho mẫu $\mathbf{x}$, ước lượng mật độ $p(\mathbf{x}, \mathbf{y})$ với lớp hàm phân bố $p(\mathbf{x}, \mathbf{y}; \theta)$.
- Ước lượng hợp lý cực đại (MLE)

$$\boxed{\theta^{\text{MLE}} = \arg\max_\theta \log p(\mathbf{x}; \theta)}$$
$$= \arg\max_\theta \log \left[\int_{\mathbf{y}} p(\mathbf{x}, \mathbf{y}; \theta)\, d\mathbf{y} \right]$$

Bất đẳng thức biến phân cho EM (variational inequality)

Xét một phân bố bất kì $q(\mathbf{y})$, ta có

$$\begin{aligned}\log p(\mathbf{x};\theta) &= \int_{\mathbf{y}} \log p(\mathbf{x};\theta)\, q(\mathbf{y})\, d\mathbf{y} \\ &= \int_{\mathbf{y}} \log p(\mathbf{x},\mathbf{y};\theta)\, q(\mathbf{y})\, d\mathbf{y} - \int_{\mathbf{y}} \log p(\mathbf{y}|\mathbf{x};\theta)\, q(\mathbf{y})\, d\mathbf{y}\end{aligned}$$

$$\int_{\mathbf{y}} \log p(\mathbf{x},\mathbf{y};\theta)\, q(\mathbf{y})\, d\mathbf{y} = E_q[\log p(\mathbf{x},\mathbf{y};\theta)]$$

$$-\int_{\mathbf{y}} \log p(\mathbf{y}|\mathbf{x};\theta)\, q(\mathbf{y})\, d\mathbf{y} = \underbrace{\int_{\mathbf{y}} \log \frac{q(\mathbf{y})}{p(\mathbf{y}|\mathbf{x};\theta)}\, q(\mathbf{y})\, d\mathbf{y}}_{D_{\mathrm{KL}}[q\|p(\mathbf{y}|\mathbf{x};\theta)]} \underbrace{- \int_{\mathbf{y}} \log q(\mathbf{y})\, q(\mathbf{y})\, d\mathbf{y}}_{\mathcal{E}[q]}$$

- $E_q[\log p(\mathbf{x},\mathbf{y};\theta)]$: kì vọng theo phân bố $q(\mathbf{y})$.
- $D_{\mathrm{KL}}[q\|p(\mathbf{y}|\mathbf{x};\theta)]$: khoảng cách Kullback-Leibler.
- $\mathcal{E}[q]$: entropy của phân bố $q(\mathbf{y})$.

Bất đẳng thức biến phân cho EM (variational inequality)

Xét một phân bố bất kì $q(\mathbf{y})$, ta có

$$\boxed{\underbrace{\log p(\mathbf{x};\theta)}_{\text{sự hợp lý}} = \underbrace{E_q[\log p(\mathbf{x},\mathbf{y};\theta)]}_{\text{kì vọng}} + \underbrace{D_{\mathrm{KL}}[q\|p(\mathbf{y}|\mathbf{x};\theta)]}_{\text{khoảng cách KL}\geq 0} + \underbrace{\mathcal{E}[q]}_{\text{entropy}}}$$

- Sự hợp lý bị chặn dưới bởi kì vọng + entropy.
- Cận dưới này chặt nhất (dấu bằng xảy ra) khi

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

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

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

Bản chất thuật toán EM là cực đại hóa cận dưới của sự hợp lý.

Thuật toán cực đại hóa kì vọng (EM)

Thuật toán 1 Cực đại hóa kì vọng (EM)

1: **Input**: mẫu $\mathbf{x}$ (thông tin về $\mathbf{y}$ bị ẩn).
2: **Khởi tạo**: chọn tham số $\theta^{(0)}$ của phân bố $p(\mathbf{x}, \mathbf{y}; \theta)$ và $k = 0$.
3: **while** chưa hội tụ **do**
4: Bước E: tính phân bố hậu nghiệm $p(\mathbf{y}|\mathbf{x}; \theta^{(k)})$ và biểu thức kì vọng

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

5: Bước M: cực đại hóa kì vọng

$$\theta^{(k+1)} = \arg\max_{\theta} Q(\theta|\theta^{(k)})$$

6: $k \leftarrow k + 1$.
7: **end while**

- Đặt tên bởi Arthur Dempster, Nan Laird, và Donald Rubin (1977).

Thuật toán cực đại hóa kì vọng (EM)

Đặc biệt khi $p(\mathbf{x}, \mathbf{y}; \theta)$ có phân bố dạng mũ (exponential family)

$$\boxed{p(\mathbf{x}, \mathbf{y}; \theta) = h(\mathbf{x}, \mathbf{y}) g(\theta) \exp\{\langle \overrightarrow{\eta(\theta)}, \overrightarrow{T(\mathbf{x}, \mathbf{y})} \rangle\}}$$

- Véc-tơ $\overrightarrow{T(\mathbf{x}, \mathbf{y})}$ gọi là thống kê đủ (sufficient statistics) của phân bố dạng mũ.
- Bài toán MLE khi biết cả $\mathbf{x}$ và $\mathbf{y}$

$$\arg\max_{\theta} \log p(\mathbf{x}, \mathbf{y}; \theta) = \arg\max_{\theta} [\log g(\theta) + \langle \overrightarrow{\eta(\theta)}, \overrightarrow{T(\mathbf{x}, \mathbf{y})} \rangle]$$

- Thuật toán EM khi $\mathbf{y}$ bị ẩn đi

$$\arg\max_{\theta} E_q[\log p(\mathbf{x}, \mathbf{y}; \theta)] = \arg\max_{\theta} [\log g(\theta) + \langle \overrightarrow{\eta(\theta)}, E_q[\overrightarrow{T(\mathbf{x}, \mathbf{y})}] \rangle]$$

Thuật toán cực đại hóa kì vọng (EM)

$$\text{(MLE)} : \arg\max_{\theta}[\log g(\theta) + \langle \overrightarrow{\eta(\theta)}, \overrightarrow{T(\mathbf{x},\mathbf{y})} \rangle]$$

$$\text{(EM)} : \arg\max_{\theta}[\log g(\theta) + \langle \overrightarrow{\eta(\theta)}, E_q[\overrightarrow{T(\mathbf{x},\mathbf{y})}] \rangle]$$

- Rolf Sundberg (1974): Bước M của thuật toán EM chính là phương pháp MLE sau khi thay thống kê đủ $T(\mathbf{x},\mathbf{y})$ bằng kì vọng của nó theo phân bố $q(\mathbf{y})$.

Ví dụ

Trộn phân bố chuẩn (mixture of gaussians)

- Giả sử n mẫu $z_1, z_2, \ldots, z_n$ độc lập và có cùng phân bố $p(z)$.
- Xét lớp hàm phân bố là trộn của K phân bố chuẩn

$$p(z\,;\ \mu_{1:K}, \sigma^2_{1:K}) = \sum_{k=1}^{K} p_k \mathcal{N}(z\,;\ \mu_k, \sigma^2_k)$$

- Biến ẩn: gọi $c_{ik} \in \{0, 1\}$ là biến nhị phân chỉ ra mẫu z_i có thuộc phân bố chuẩn thứ k hay không.

$$p(z_{1:n}, \boldsymbol{c}\,;\ \mu_{1:K}, \sigma^2_{1:K}) = \underbrace{\prod_{i=1}^{n}\prod_{k=1}^{K}\left(p_k \mathcal{N}(z_i; \mu_k, \sigma^2_k)\right)^{c_{ik}}}_{\text{Phân bố dạng mũ}}$$

Ví dụ
Trộn phân bố chuẩn: MLE khi biết c_{ik}

- Thống kê đủ

$$n_k = \sum_{i=1}^{n} c_{ik}, \quad t_{1,k} = \sum_{i=1}^{n} c_{ik} z_i, \quad t_{2,k} = \sum_{i=1}^{n} c_{ik} z_i^2$$

- MLE

$$p_k = \frac{n_k}{n}$$
$$\mu_k = \frac{t_{1,k}}{n_k}$$
$$\sigma_k^2 = \frac{t_{2,k}}{n_k} - \left(\frac{t_{1,k}}{n_k}\right)^2$$

Ví dụ

Trộn phân bố chuẩn: EM khi không biết c_{ik}

- Bước E: tính xác suất hậu nghiệm và kì vọng của thống kê đủ

$$r_{ik} = E[c_{ik}|z_i; \boldsymbol{\mu}, \boldsymbol{\sigma}^2] = p(c_{ik} = 1|z_i; \boldsymbol{\mu}, \boldsymbol{\sigma}^2) = \frac{p_k \mathcal{N}(z_i|\mu_k, \sigma_k^2)}{\sum_{k'} p_{k'} \mathcal{N}(z_i|\mu_{k'}, \sigma_{k'}^2)}$$

$$\bar{n}_k = \sum_{i=1}^{n} r_{ik}, \quad \bar{t}_{1,k} = \sum_{i=1}^{n} r_{ik} z_i, \quad \bar{t}_{2,k} = \sum_{i=1}^{n} r_{ik} z_i^2$$

- Bước M: thế vào công thức MLE

$$p_k = \frac{\bar{n}_k}{n}$$

$$\mu_k = \frac{\bar{t}_{1,k}}{\bar{n}_k}$$

$$\sigma_k^2 = \frac{\bar{t}_{2,k}}{\bar{n}_k} - \left(\frac{\bar{t}_{1,k}}{\bar{n}_k}\right)^2$$

Nội dung

1 Giới thiệu

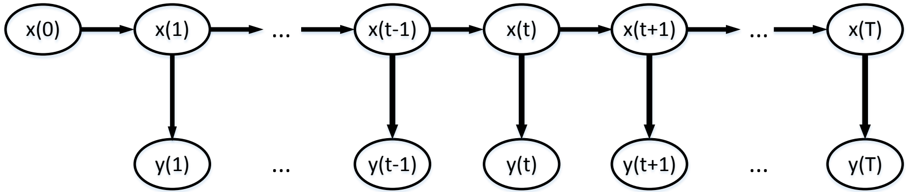
2 Thuật toán cực đại hóa kì vọng (EM)

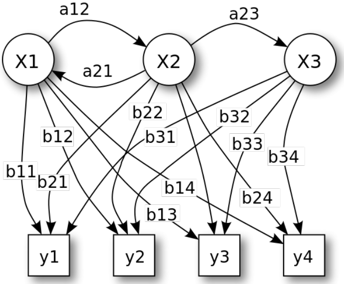
3 Mô hình Markov ẩn

4 Tổng kết

Mô hình Markov ẩn

Giới thiệu





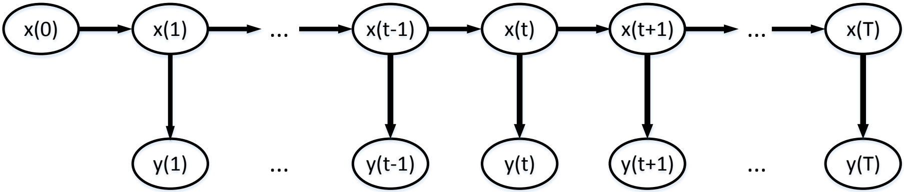
- Biến trạng thái $\mathbf{x} \in \{x_1, \ldots, x_N\}$ (ẩn).
- Biến nhãn $\mathbf{y} \in \{y_1, \ldots, y_K\}$ (hiện).
- Tính chất Markov:

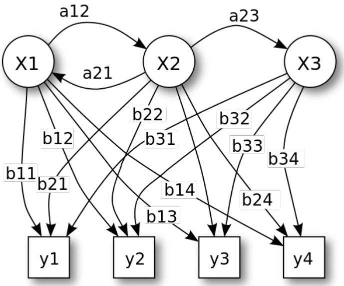
$$x(u) \perp x(v) \mid x(t), u > t, v < t$$

$$x(u) \perp y(v) \mid x(t), u > t, v \leq t.$$

Mô hình Markov ẩn

Tham số của mô hình





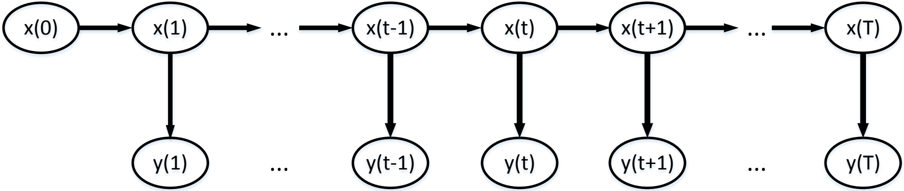
- Xác suất chuyển trạng thái

$$P(x(t+1) = x_j | x(t) = x_i) = a_{ij}$$

- Xác suất sinh ra $\mathbf{y}$

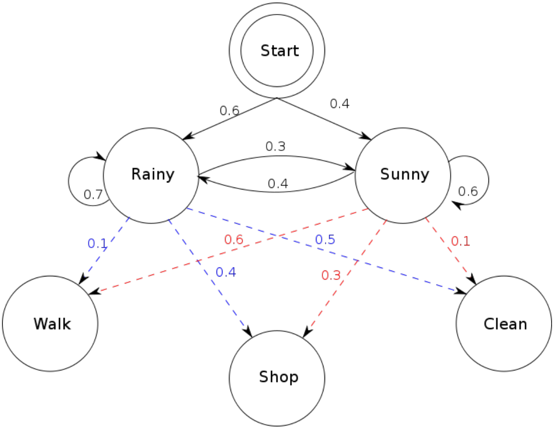
$$P(y(t) = y_k | x(t) = x_i) = b_{ik}$$

Mô hình Markov ẩn
Ví dụ



- Biến ẩn $\mathbf{x}$ có thể là chức năng đoạn gen, loại của từ trong câu, v.v...
- Biến quan sát được $\mathbf{y}$ là kí hiệu nucleotide ATCG, các từ cụ thể, v.v...
- Xác suất chuyển trạng thái là xác suất chuyển giữa các đoạn gen có chức năng khác nhau, v.v...
- Xác suất sinh dữ liệu là xác suất để mỗi đoạn gen sinh ra các nucleotide khác nhau, v.v...
- Bài toán: Quan sát được (biến hiện) $\mathbf{y}$, tính tham số của mô hình Markov ẩn $\mathbf{A} = [a_{ij}], \mathbf{B} = [b_{ik}]$.

Mô hình Markov ẩn
Ví dụ: Wikipedia Alice - Bob



Xác suất đầy đủ của HMM

Giả sử ta biết cả biến trạng thái $\mathbf{x} = (x(0) = x_1, x(1), x(2), \ldots, x(T))$ và biến quan sát được $\mathbf{y} = (y(1), y(2), \ldots, y(T))$.

- Đặt $\delta^s_{t,i} \in \{0, 1\}$ là biến nhị phân chỉ ra $x(t) = x_i$ (biến ẩn).
- Đặt $\delta^o_{t,k} \in \{0, 1\}$ là biến nhị phân chỉ ra $y(t) = y_k$ (biến hiện).
- Xác suất chuyển trạng thái $x(t-1) \to x(t)$ là

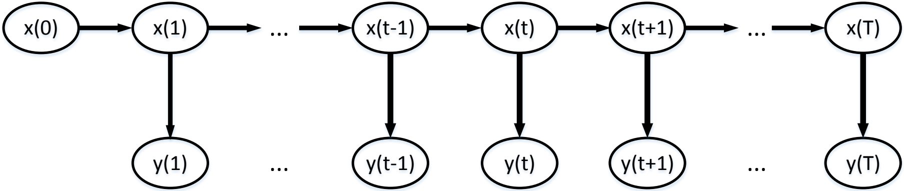
$$\prod_{i=1}^{N} \prod_{j=1}^{N} a_{ij}^{\delta^s_{t-1,i} \ \delta^s_{t,j}}$$

  (chỉ có một số khác 1)
- Xác suất sinh dữ liệu $x(t) \to y(t)$ là

$$\prod_{i=1}^{N} \prod_{k=1}^{K} b_{ik}^{\delta^s_{t,i} \ \delta^o_{t,k}}$$

Xác suất đầy đủ của HMM

Giả sử ta biết cả biến trạng thái $\mathbf{x} = (x(0) = x_1, x(1), x(2), \ldots, x(T))$ và biến quan sát được $\mathbf{y} = (y(1), y(2), \ldots, y(T))$.



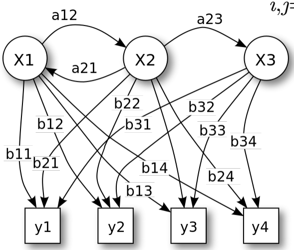
- Xác suất đầy đủ (phân phối dạng mũ)

$$P(\underbrace{\boldsymbol{\delta}^s}_{\text{biến ẩn}}, \underbrace{\boldsymbol{\delta}^o}_{\text{biến hiện}}; \underbrace{\mathbf{A}, \mathbf{B}}_{\text{tham số } \theta}) = \prod_{t=1}^{T} \left[\prod_{i=1}^{N} \prod_{j=1}^{N} a_{ij}^{\delta^s_{t-1,i}\delta^s_{t,j}} \prod_{i=1}^{N} \prod_{k=1}^{K} b_{ik}^{\delta^s_{t,i}\delta^o_{t,k}} \right]$$

Thống kê đủ

$$\log P(\boldsymbol{\delta}^s, \boldsymbol{\delta}^o; \mathbf{A}, \mathbf{B}) = \sum_{i,j=1}^{N,N} (\log a_{ij}) \left[\sum_{t=1}^{T} \delta^s_{t-1,i} \delta^s_{t,j} \right] + \sum_{i,k=1}^{N,K} (\log b_{ik}) \left[\sum_{t=1}^{T} \delta^s_{t,i} \delta^o_{t,k} \right]$$

$$= \sum_{i,j=1}^{N,N} \log a_{ij} \cdot U_{ij} + \sum_{i,k=1}^{N,K} \log b_{ik} \cdot V_{ik}$$



- Thống kê đủ (số đếm trên cạnh đồ thị)

$$\underbrace{U_{ij} = \sum_{t=1}^{T} \delta^s_{t-1,i}\ \delta^s_{t,j}}_{\text{số lần } x_i \to x_j} \quad , \quad \underbrace{V_{ik} = \sum_{t=1}^{T} \delta^s_{t,i}\ \delta^o_{t,k}}_{\text{số lần } x_i \to y_k}$$

MLE khi biết δ^s

Tính a_{ij}, b_{ik}

- Bài toán tối ưu có ràng buộc

$$\max_{a_{ij}, j=1,\dots,N} \sum_{j=1}^{N} \log a_{ij} \cdot U_{ij} \text{ , sao cho } \sum_{j=1}^{N} a_{ij} = 1$$

$$\max_{b_{ik}, k=1,\dots,K} \sum_{k=1}^{K} \log b_{ik} \cdot V_{ik} \text{ , sao cho } \sum_{k=1}^{K} b_{ik} = 1$$

- Dùng phương pháp nhân tử Lagrange

$$a_{ij} = \frac{U_{ij}}{\sum_{j'} U_{ij'}} \propto U_{ij}, j = 1, \dots, N$$

$$b_{ik} = \frac{V_{ik}}{\sum_{k'} V_{ik'}} \propto V_{ik}, k = 1, \dots, K$$

EM khi không biết $\boldsymbol{\delta}^s$

Thế vào công thức MLE

- Tính kì vọng hậu nghiệm

$$\overline{U}_{ij} = E[U_{ij}|\boldsymbol{\delta}^o; \mathbf{A}, \mathbf{B}], \quad \overline{V}_{ik} = E[V_{ik}|\boldsymbol{\delta}^o; \mathbf{A}, \mathbf{B}]$$

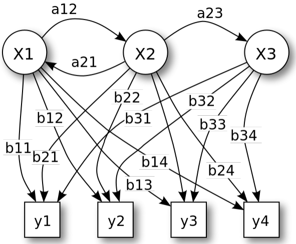
- Thế vào công thức MLE

$$a_{ij} = \frac{\overline{U}_{ij}}{\sum_{j'} \overline{U}_{ij'}}, \quad b_{ik} = \frac{\overline{V}_{ik}}{\sum_{k'} \overline{V}_{ik'}},$$

với $i, j = 1, \ldots, N, k = 1, \ldots, K.$

EM khi không biết $\boldsymbol{\delta}^s$

Kì vọng hậu nghiệm



$$\overline{U}_{ij} = E[U_{ij} \mid \boldsymbol{\delta}^o] = \sum_{t=1}^{T} E[\delta^s_{t-1,i}\ \delta^s_{t,j} \mid \boldsymbol{\delta}^o]$$

$$E[\delta^s_{t,i}\ \delta^s_{t+1,j} \mid \boldsymbol{\delta}^o] = P(x(t) = x_i, x(t+1) = x_j \mid \boldsymbol{\delta}^o)$$

$$\overline{V}_{ik} = E[V_{ik} \mid \boldsymbol{\delta}^o] = \sum_{t=1}^{T} E[\delta^s_{t,i}\ \delta^o_{t,k} \mid \boldsymbol{\delta}^o]$$

$$E[\delta^s_{t,i}\ \delta^o_{t,k} \mid \boldsymbol{\delta}^o] = P(x(t) = x_i, y(t) = y_k \mid \boldsymbol{\delta}^o)$$

Như vậy, thuật toán EM được quy về việc tính các xác suất hậu nghiệm (xác suất cạnh của đồ thị)

$$\xi_{ij}(t) = P(x(t) = x_i, x(t+1) = x_j \mid \boldsymbol{\delta}^o)$$

$$\gamma_{ik}(t) = P(x(t) = x_i, y(t) = y_k \mid \boldsymbol{\delta}^o)$$

EM khi không biết $\boldsymbol{\delta}^s$

Tính các xác suất hậu nghiệm

- Do biến hiện $\boldsymbol{\delta}^o$ là hằng số, nhân 2 vế với $P(\boldsymbol{\delta}^o) = \text{const}$

$$\xi_{ij}(t) \propto P(x(t) = x_i, x(t+1) = x_j, \boldsymbol{\delta}^o)$$
$$\gamma_{ik}(t) \propto P(x(t) = x_i, y(t) = y_k, \boldsymbol{\delta}^o)$$

- Quy hoạch động

$$\alpha_i(t) = P(x(t) = i, \delta_1^o, \ldots, \delta_t^o)$$
$$\beta_i(t) = P(\delta_{t+1}^o, \ldots, \delta_T^o | x(t) = i)$$

- Dùng công thức Bayes

$$P(x(t) = x_i, x(t+1) = x_j, \boldsymbol{\delta}^o) = \alpha_i(t) a_{ij} \beta_j(t+1) \prod_{k=1}^{K} b_{jk}^{\delta_{t,k}^o}$$
$$P(x(t) = x_i, y(t) = y_k, \boldsymbol{\delta}^o) = \delta_{t,k}^o \times \alpha_i(t)\beta_i(t)$$

EM khi không biết δ^s

Quy hoạch động tính $\alpha_i(t), \beta_i(t)$

Công thức đệ quy (lại sử dụng công thức Bayes)

- Thuật toán xuôi

$$\alpha_i(t) = \begin{cases} 1 & , \quad t = 0, i = 1 \\ 0 & , \quad t = 0, i \neq 1 \\ \left[\sum_{u=1}^{N} \alpha_u(t-1) a_{ui}\right] \prod_{k=1}^{K} b_{ik}^{\delta^o_{t,k}} & , \quad t > 0 \end{cases}$$

- Thuật toán ngược

$$\beta_i(t) = \begin{cases} 1 & , \quad t = T \\ \sum_{j=1}^{N} \beta_j(t+1) a_{ij} \prod_{k=1}^{K} b_{jk}^{\delta^o_{t+1,k}} & , \quad t < T \end{cases}$$

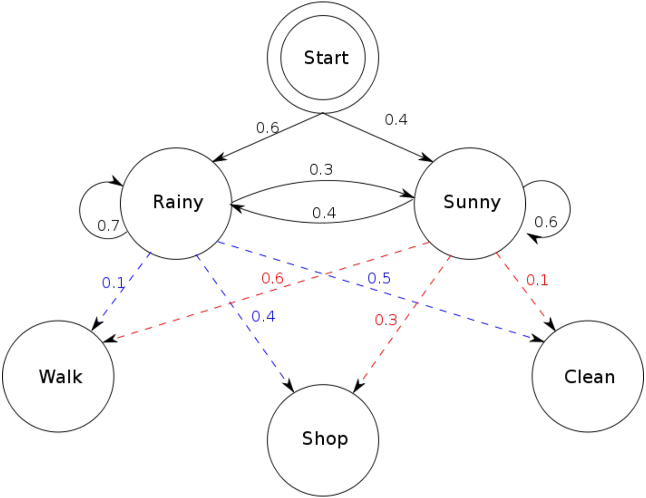
Thuật toán EM cho mô hình Markov ẩn

Thuật toán 2 EM cho HMM

Input: chuỗi $\mathbf{y}$ (tức là các biến nhị phân $\boldsymbol{\delta}^o$).
Khởi tạo: chọn tham số $\mathbf{A}^{(0)}, \mathbf{B}^{(0)}$ và $r = 0$.
while chưa hội tụ **do**
  Bước E:
  - Tính $\alpha_i(t), \beta_i(t)$ bằng quy hoạch động.
  - Tính xác suất hậu nghiệm $\xi_{ij}(t), \gamma_{ik}(t)$.
  - Tính kì vọng của thống kê đủ $\overline{U}_{ij}, \overline{V}_{ik}$.

  Bước M: cập nhật $a_{ij}^{(r+1)}, b_{ik}^{(r+1)}$ (thế $\overline{U}_{ij}, \overline{V}_{ik}$ vào công thức MLE).
  $r \leftarrow r + 1$.
end while

Ví dụ: Wikipedia Alice - Bob



Nội dung

1 Giới thiệu

2 Thuật toán cực đại hóa kì vọng (EM)

3 Mô hình Markov ẩn

4 Tổng kết

Tổng kết

- EM là thuật toán ước lượng mật độ (có tham số) cho dữ liệu không đầy đủ (có thông tin bị ẩn đi).
- Bản chất EM là cực đại hóa cận dưới sự hợp lý của tham số.

> Với phân bố dạng mũ, bước M của EM là ước lượng hợp lý cực đại (MLE) với thống kê đủ được thay bằng kì vọng hậu nghiệm.

- Lớp phân bố dạng mũ rất rộng, bao gồm cả mô hình trộn phân bố chuẩn, mô hình Markov ẩn.