AMB

Зубков Максим, 777 группа

17 апреля 2019 г.

Теперь алгоритм foraward-backward. Допустим нам дана некоторая марковская модель λс матрицей A и матрицей B, пусть там также дана последовательность наблюдейний$obs = \{o_i\}_{i=1}^T$ и начальное распределение pi. Нам необходимо найти величину $P(obs|\lambda)$, то есть вероятность того насколько возможно получить данную последовательность наблюдений в условиях данной модели. Алгоритм backward-forward является частью алгоритма baum-welch, который изменяет модельλтак, чтобы вероятность $P(obs|\lambda)$была максимальной, то есть настраивает модель под данного пользователя Обозначим $\alpha_t(i) = P(o_1, o_2, ..., o_t, s_t = i|\lambda)$- вероятность на t-том шаге встретить последовательность $o_1, o_2, ..., o_t$и оказаться в состоянии i. Таким образом для любого $i : \alpha_1(i) = P(o_1, s_1 = i|\lambda) = P(o_1|s_1 = i, \lambda) * P(s_1 = i|\lambda) = \pi_i * B[i, o_1]$ Далее по индукции заполним $\alpha_t(i) = P(o_1, o_2, ..., o_t, s_t = i|\lambda) = \sum_{j=1}^N P(o_1, o_2, ..., o_t, s_{t-1} = j, s_t = i|\lambda) = \sum_{j=1}^N P(o_t|s_{t-1} = j, s_t = i, o_1, o_2, ..., o_{t-1}, \lambda) * P(s_t = i|s_{t-1} = j, o_1, o_2, ..., o_{t-1}, \lambda)*P(s_{t-1} = j, o_1, o_2, ..., o_{t-1}, \lambda) = P(o_t|s_t = i, \lambda) \sum_{j=1}^N P(s_t = i|o_1, o_2, ..., o_t, s_{t-1} = j, \lambda)P(o_1, o_2, ..., o_{t-1}, s_{t-1} = j, \lambda) = B[i, o_t]*\sum_{j=1}^N \alpha_{t-1}(i)A[i, j]$ Доказательство второго равенства будет в pdf $P(obs|\lambda) = \sum_{i=1}^N P(o_1, ..., o_T, s_T = i|\lambda) = \sum_{i=1}^N \alpha_T(i)\beta_t(i) = P(o_{t+1}, o_{t+2}, ..., o_T|s_t = i, \lambda)$ - вероятность наблюдаемой части последовательных наблюдений $o_{t+1}, o_{t+2}, ..., o_T$ при условии, что в момент времени t состояние было i и при модели $\lambda.\beta_T(i) = 1\forall i$ Далее по индукции заполним $\beta_t(j) = P(o_{t+1}, o_{t+2}, ..., o_T|s_t = j, \lambda) =$ Объяснение следует закинуть в tex

$$\sum_{i=1}^{N} P(o_{t+1}, ..., o_T, s_{t+1} = i|s_t = i, \lambda) =$$

$$\sum_{i=1}^{N} \frac{P(o_{t+1}, ..., o_T, s_{t+1} = i, s_t = j, \lambda)}{P(s_t = j, \lambda)} =$$

$$\sum_{i=1}^{N} P(o_{t+1}|o_{t+2}..., o_T, s_{t+1} = i, s_t = j, \lambda) \cdot P(o_{t+2}..., o_T|s_{t+1} = i, s_t = j, \lambda) \cdot \frac{P(s_{t+1} = i, s_t = j, \lambda))}{P(s_t = j, \lambda)}$$

Согласно свойству марковских цепей: $P(S_t|S_{t-1}, O_{t-1}, \ldots, S_1, O_1) = P(S_t|S_{t-1})$ и $P(O_t|Q_t, Q_{t-1}, O_{t-1}, \ldots, S_1, O_1) = P(O_t|S_t)$, тогда $P(o_{t+1}|o_{t+2}..., o_T, s_{t+1} = i, s_t = j, \lambda) = P(o_{t+1}|s_{t+1} = i) = B_i(o_{t+1})$, кроме того легко понять, что $P(o_{t+2}..., o_T|s_{t+1} = i, s_t = j, \lambda) = \beta_{t+1}(i)$ и по формуле Байеса $\frac{P(s_{t+1}=i, s_t=j, \lambda))}{P(s_t=j, \lambda)} = P(s_{t+1} = i|s_t = j, \lambda) = a_{ij}$, тогда

$$\beta_t(j) = \sum_{i=1}^{N} A[i, j]\beta_{t+1}(j)B[i, obs(t + 1)]$$

$P(obs|\lambda) = ☼_{i=1}^{N} P(o_1, ..., o_T, s_T = i|\lambda) = \sum_{i=1}^{N} \alpha_T(i)$ Зная $\alpha\beta$ можно посчитать величину $\gamma_t(i) = \frac{P(s_t=i, obs|\lambda)}{P(obs|\lambda)} = \frac{P(o_1, ...o_t, s_t=i|\lambda)}{P(obs|\lambda)}$. Легко доказать, что $P(o_1, ...o_t, s_t = i|\lambda) = P(o_1, ...o_t|s_t = i, \lambda)P(o_{t+1}, ...o_T|s_t = i, \lambda)P(s_t = i|\lambda) = \alpha_t(i)\beta_t(i)$ Таким образом $\gamma_t(i) = \alpha_t(i)\beta_t(i)/P(obs|\lambda)$. Знаменатель вычислялся ранее $P(obs|\lambda) = \sum_{i=1}^{N} P(o_1, ..., o_T, s_t = i|\lambda) = \sum_{i=1}^{N} \alpha_T(i)$

$ksi_t(i, j) = P(s - t = i, s_{t+1} = j|obs, \lambda)$ - вероятность находиться в состоянии i в момент времени t и в состоянии j в момент времени t+1 при условии наблюдений obs и моделиλ $P = \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_t(i)A_{ij}B_j(y_{t+1})\beta_{t+1}(i)$

1 Алгоритм Баума-Вэлша

Алгоритм Баума-Велша занимается обучаением скрытой марковской модели (далее СММ). Алгоритм итеративно изменяет модель $\lambda = (A, B, \pi)$ таким образом, чтобы вероятность $P(obs|\lambda)$ была максимальной.

Пусть λ - текущая модель, а $\widehat{\lambda}$ - кандидат стать новой моделью, необходимо найти такое $\widehat{\lambda}$, чтобы $P(obs|\widehat{\lambda}) \geq P(obs, \lambda)$ или, что эквивалентно $\log P(obs|\widehat{\lambda}) \geq \log P(obs, \lambda)$. Введем вспомогательную функцию $Q(\widehat{\lambda}|\lambda) = \mathbb{E}\left[\log P(obs, S|\widehat{\lambda})|obs, \lambda\right]$ по определению условного мат ожидания $\sum_s P(S|obs, \lambda) \cdot \log\left[P(obs, S|\widehat{\lambda})\right]$. Можно так же доказать, что задча поиска $\widehat{\lambda} = \arg\max_{\lambda} \sum_s P(obs, S|\lambda)$ эквивалентна задаче поиска $\arg\max_{\lambda} Q(\widehat{\lambda}|\lambda)$