BAB III

PENAKSIRAN PARAMETER-PARAMETER PADA HMM DENGAN ALGORITMA BAUM-WELCH

Pada bab ini akan dipaparkan algoritma Baum-Welch yang akan digunakan untuk menyelesaikan permasalahan HMM yang ketiga yaitu menaksir parameter-parameter pada HMM.

3.1 Algoritma Baum-Welch

Permasalahan ketiga dalam HMM berkaitan dengan mengestimasi 3 parameter pada HMM yaitu $\boldsymbol{A}, \boldsymbol{B}$, dan $\boldsymbol{\pi}$ sehingga terbentuk model baru $\hat{\lambda} = (\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{B}}, \widehat{\boldsymbol{\pi}})$ dimana $P(O|\hat{\lambda}) \geq P(O|\lambda)$ (Nisa Pandu, 2011). Dengan kata lain, permasalahan ketiga adalah masalah optimasi. Adapun permasalahan yang harus dipecahkan adalah mengestimasi model terbaik yang dapat menjelaskan suatu barisan observasi (O. C. Ibe, 2009).

Karena masalah yang ketiga merupakan optimasi, maka dicari metode terbaik untuk barisan observasi yang menentukan kriteria dalam tahap optimasi ini. Kriteria optimasi yang paling sering digunakan adalah kriteria kemungkinan terbesar (*maximum likelihood*) yang digunakan juga dalam mengestimasi parameter-parameter dalam HMM, artinya memaksimumkan peluang yang diberikan dari barisan observasi yang ada (O. C. Ibe, 2009). Sehingga solusinya dirumuskan:

$$\hat{\lambda} = \arg\max_{\lambda}\{P[O|\lambda]\}$$

Namun untuk menyelesaikan solusi di atas untuk masalah ini bukanlah hal yang mudah dan belum diketahui metode untuk menganalisis dan memperoleh λ yang memaksimumkan $P[O|\lambda]$, tapi dipilih model dengan parameter-parameter tertentu sehingga $P[O|\lambda]$ maksimum secara lokal (O. C. Ibe, 2009). Metode ini merupakan suatu solusi iteratif yang disebut algoritma Baum-Welch, dikenal juga sebagai algoritma maju-mundur (karena melibatkan perhitungan algoritma maju dan algoritma mundur dalam masalah HMM yang pertama yaitu menghitung peluang observasi). Selain itu algoritma Baum-Welch merupakan kasus khusus dari algoritma *Expectation Maximization* (EM) yang merupakan algoritma yang digunakan untuk mempelajari model-model probabilistik (melibatkan perhitungan peluang) dalam suatu kasus yang melibatkan keadaan-keadaan tersembunyi (O. C. Ibe, 2009).

Dalam algoritma Baum-Welch didefinisikan variabel maju dan variabel mundur, yang digunakan dalam algoritma maju dan algoritma mundur pada penyelesaian permasalahan HMM yang pertama yaitu menghitung peluang observasi (Nisa Pandu, 2011), yaitu sebagai berikut:

$$\begin{cases} \alpha_t(i) = P(O_1, O_2, \dots, O_t, X_T = i|\lambda) \\ \beta_t(i) = P(O_{t+1}, O_{t+2}, \dots, O_t, X_T = i|\lambda) \end{cases} \tag{3.1}$$

Dimana $t = 1,2, \dots, T; i = 1,2, \dots, N$ dan variabel $\alpha_t(i)$ menyatakan peluang keadaan i pada waktu t jika diketahui barisan keadaan terobservasi $\{O_1, O_2, \dots O_T\}$, dan $\beta_t(i)$ menyatakan peluang bersyarat dari bagian keadaan terobservasi $O_{t+1}, O_{t+2}, \ldots O_T$ yang diberikan model saat keadaan tersembunyi i berada pada waktu t. Perhitungannya sama seperti perhitungan pada pembahasan sebelumnya di bab II yaitu:

$$\alpha_t(i) = \pi_i b_j(O_1) \qquad 1 \le i \le N$$

$$\alpha_{t+1}(j) = \left\{\sum_{i=1}^{N} p_{ij}\,\alpha_t(i)\right\} b_j(O_{t+1}) \qquad 1 \le t \le T-1, 1 \le j \le N$$

$$\beta_T(i) = 1 \qquad 1 \le i \le N$$

$$\beta_T(i) = \sum_{j=1}^{N} p_{ij}\,\beta_{t+1}(j)\, b_j(O_{t+1}) \qquad 1 \le t \le T-1, 1 \le i \le N$$

Seperti dalam algoritma Viterbi (untuk menyelesaikan permasalahan HMM yang kedua yaitu menentukan barisan keadaan tersembunyi), didefinisikan variabel $\gamma_t(i)$sebagai berikut:

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{P[O|\lambda]} = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^{N} \beta_t(i)\,\alpha_t(i)}$$

Variabel tersebut menyatakan peluang dari keadaan tersembunyi i pada waktu t dengan seluruh barisan observasi dan model yang diberikan. Jumlah $\gamma_t(i)$ berdasarkan t memberikan angka harapan dari transisi untuk keadaan tersembunyi i pada waktu t dan keadaan tersembunyi j pada waktu $t+1$.

Kemudian didefinisikan sebuah variabel baru $\xi_t(i,j)$ dimana $\xi_t(i,j)$ adalah peluang proses berada pada keadaan tersembunyi-i pada waktu t dan berada pada keadaan tersembunyi-j pada waktu t bila diketahui barisan observasi dan model:

$$\xi_t(i,j) = P(X_t = i, X_{t+1} = j|O,\lambda) \qquad (3.2)$$

Dengan menggunakan definisi peluang bersyarat dan aturan Bayes, maka variabel $\xi_t(i,j)$ dapat dinyatakan sebagai:

$$\xi_t(i,j) = P(X_t = i, X_{t+1} = j|O,\lambda)$$

$$= \frac{P(X_t = i, X_{t+1} = j, O|\lambda)}{P(O|\lambda)}$$

$$= \frac{P(O_1, O_2, \dots, O_t, X_t = i|\lambda)P(X_{t+1} = j|X_t = i)P(O_{t+1}|X_{t+1} = j) \\ P(O_{t+2}, \dots, O_t, X_t = j|\lambda)}{P(O|\lambda)}$$

$$= \frac{\alpha_t(i)a_{ij}\, b_j(O_{t+1})\beta_{t+1}(j)}{P(O|\lambda)}$$

Dengan diperoleh nilai $\xi_t(i,j)$, bisa dihitung peluang proses berada pada keadaan tersembunyi *i* pada waktu *t* , $\gamma_t(i)$ dengan menjumlahkan $\xi_t(i,j)$ atas *j*.

$$\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i,j) \quad (3.3)$$

Karena diketahui dari hasil sebelumnya bahwa $\gamma_t(i)$ merupakan peluang proses berada pada keadaan tersembunyi *i* pada waktu *t*, maka penaksir parameter π :

$$\hat{\pi}(i) = \gamma_1(i) \quad (3.4)$$

Sementara untuk penaksir a_{ij} adalah:

$$\hat{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)}$$

Penaksir tersebut diperoleh dengan membagi jumlah transisi dari keadaan tersembunyi *i* ke keadaan tersembunyi *j* dengan total seluruh transisi dari keadaan tersembunyi *i*.

Begitu juga dengan penaksir $b_i(j)$ yaitu:

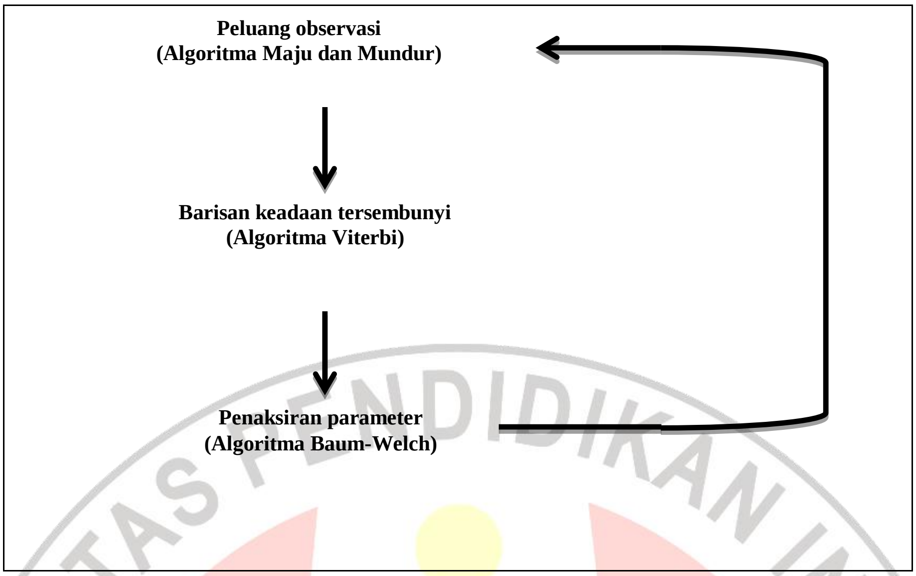
$$\hat{b}_i(j) = \frac{\sum_{t=1, O_t=j}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)}$$

Yang diperoleh dengan membagi jumlah keadaan tersembunyi yang menghasilkan observasi *j* pada saat proses berada pada keadaan tersembunyi *i* dengan jumlah seluruh proses yang berada pada keadaan tersembunyi *i*.

Jelasnya algoritma ini adalah sebagai berikut:

1. Dapatkan estimasi (perkiraan) untuk distribusi keadaan awal untuk keadaan tersembunyi i, sebagai frekuensi harapan dengan keadaan tersembunyi i, saat $t = 1$ yaitu: $\hat{\pi}_i = \gamma_t(i)$
2. Dapatkan estimasi (perkiraan) untuk $\hat{a}_{ij}$ dan $\hat{b}_i(j)$ seperti yang didefinisikan pada pembahasan sebelumnya.
3. Gunakan model $\lambda = (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi})$ yang dimiliki untuk menghitung $\hat{a}_{ij}$ dan $\hat{b}_i(j)$. Kemudian estimasi kembali model baru $\hat{\lambda} = (\hat{\boldsymbol{A}}, \hat{\boldsymbol{B}}, \hat{\boldsymbol{\pi}})$. Lakukan langkah ini sampai diperoleh $\hat{a}_{ij}$ dan $\hat{b}_i(j)$ yang konvergen.
4. Jika $P[O|\hat{\lambda}] - P[O|\lambda] < \delta$, hentikan iterasi, dimana δ merupakan suatu nilai yang ditentukan sebelumnya.

Langkah-langkah dalam algoritma Baum-Welch di HMM, digambarkan dalam gambar sebagai berikut:



Gambar 3.1 Langkah-langkah algoritma Baum-Welch

Berdasarkan gambar 3.1, maka algoritma Baum-Welch adalah langkah iteratif. Karena algoritma ini digunakan untuk memaksimumkan peluang observasi, maka algoritma Baum-Welch akan berhenti saat peluang observasi sudah maksimum. Sehingga taksiran untuk elemen-elemen dari parameter-parameter $(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{B}}, \widehat{\boldsymbol{\pi}})$ memungkinkan konvergen pada suatu nilai tertentu.

Parameter-parameter baru dari setiap iterasi yang dihasilkan kemungkinan akan mempengaruhi peluang observasi dan barisan keadaan tersembunyi. Sehingga selain akan diteliti kekonvergenan parameter-parameter HMM ini, permasalahan pertama dan kedua dalam HMM juga harus diteliti kembali setelah diperoleh parameter-parameter yang baru.

Contoh 3.5

Perhatikan kembali kasus permainan komputer Viola. Dengan menggunakan informasi yang sama dengan contoh-contoh sebelumnya, akan dicari penaksir parameter unuk $\hat{\lambda} = (\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{B}}, \widehat{\boldsymbol{\pi}})$.

Penyelesaian:

Diketahui,

$$\xi_t(i,j) = \frac{\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)}{P(O|\lambda)}$$

Hasilnya adalah sebagai berikut:

Tabel 3.1 Hasil perhitungan $\xi_t(i,j)$

t	1	2	3
$\xi_1(1,1)$	0,1242	0,1493	0,3642
$\xi_1(1,2)$	0,1183	0,0439	0,1152
$\xi_1(1,3)$	0,2122	0,0346	0,086
t	1	2	3
$\xi_1(2,1)$	0,0596	0,013041	0,13708
$\xi_1(2,2)$	0,1234	0,1023	0,09424
$\xi_1(2,3)$	0,1616	0,0588	0,0514
$\xi_1(3,1)$	0,0456	0,256	0,0845
$\xi_1(3,2)$	0,0787	0,1364	0,0484
$\xi_1(3,3)$	0,079	0,0602	0,02025

Dari hasil tersebut dapat dicari nilai $\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i,j)$, hasilnya adalah sebagai berikut:

Tabel 3.2 Hasil perhitungan $\gamma_t(i)$

t	1	2	3
$\gamma_t(1)$	0,4547	0,2278	0,5654
$\gamma_t(2)$	0,3446	0,1741	0,2827
$\gamma_t(3)$	0,2033	0,4526	0,1532

Kemudian, dengan menggunakan hasil dari perhitungan-perhitungan tersebut dapat dicari penaksir parameter HMM yaitu $\hat{\lambda} = (\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{B}}, \widehat{\boldsymbol{\pi}})$. Penaksir inilah yang nantinya akan menghasilkan $P(O|\hat{\lambda}) \geq P(O|\lambda)$. Berikut hasil perhitungan untuk penaksir parameter-parameter HMM:

$$\hat{\boldsymbol{\pi}} = \begin{bmatrix} \gamma_1(1) \\ \gamma_1(2) \\ \gamma_1(3) \end{bmatrix} = \begin{bmatrix} 0{,}4547 \\ 0{,}3446 \\ 0{,}2033 \end{bmatrix}$$

Nilai di atas merupakan taksiran peluang awal. Artinya agar nilai $P(O|\hat{\lambda}) \geq P(O|\lambda)$ terpenuhi, maka peluang proses berada pada keadaan suasana hati senang 0,4547, sedih 0,3446 dan kesal 0,2033.

$$\hat{\boldsymbol{A}} = \begin{bmatrix} \frac{\sum_{t=1}^{T} \xi_t(1,1)}{\sum_{t=1}^{T} \gamma_t(1)} & \frac{\sum_{t=1}^{T} \xi_t(1,2)}{\sum_{t=1}^{T} \gamma_t(1)} & \frac{\sum_{t=1}^{T} \xi_t(1,3)}{\sum_{t=1}^{T} \gamma_t(1)} \\ \frac{\sum_{t=1}^{T} \xi_t(2,1)}{\sum_{t=1}^{T} \gamma_t(2)} & \frac{\sum_{t=1}^{T} \xi_t(2,2)}{\sum_{t=1}^{T} \gamma_t(2)} & \frac{\sum_{t=1}^{T} \xi_t(2,2)}{\sum_{t=1}^{T} \gamma_t(2)} \\ \frac{\sum_{t=1}^{T} \xi_t(3,1)}{\sum_{t=1}^{T} \gamma_t(3)} & \frac{\sum_{t=1}^{T} \xi_t(3,2)}{\sum_{t=1}^{T} \gamma_t(3)} & \frac{\sum_{t=1}^{T} \xi_t(3,3)}{\sum_{t=1}^{T} \gamma_t(3)} \end{bmatrix}$$

$$= \begin{bmatrix} 0{,}511 & 0{,}222 & 0{,}267 \\ 0{,}262 & 0{,}399 & 0{,}339 \\ 0{,}477 & 0{,}326 & 0{,}197 \end{bmatrix}$$

Matriks tersebut merupakan penaksir untuk matriks transisi A. Matriks $\hat{A}$ menggambarkan bahwa untuk mencapai nilai $P(O|\hat{\lambda}) \geq P(O|\lambda)$ maka peluang transisi dari suasana hati senang ke senang adalah sebesar 0,511, dari suasana hati senang ke sedih sebesar 0,222, dari suasana hati senang ke kesal sebesar 0,267, dari suasana hati sedih ke senang sebesar 0,262, dari suasana hati sedih ke sedih sebesar 0,399, dari suasana hati sedih ke kesal sebesar 0,339, dari suasana hati kesal ke senang sebesar 0,477, dari suasana hati kesal ke sedih sebesar 0,326, dan dari suasana hati kesal ke kesal sebesar 0,197.

$$\widehat{\boldsymbol{B}} = \begin{bmatrix} \frac{\sum_{t=1,O_t=1}^{T} \gamma_t(1)}{\sum_{t=1}^{T} \gamma_t(1)} & \frac{\sum_{t=1,O_t=2}^{T} \gamma_t(1)}{\sum_{t=1}^{T} \gamma_t(1)} \\ \frac{\sum_{t=1,O_t=1}^{T} \gamma_t(2)}{\sum_{t=1}^{T} \gamma_t(2)} & \frac{\sum_{t=1,O_t=2}^{T} \gamma_t(2)}{\sum_{t=1}^{T} \gamma_t(2)} \\ \frac{\sum_{t=1,O_t=1}^{T} \gamma_t(3)}{\sum_{t=1}^{T} \gamma_t(3)} & \frac{\sum_{t=1,O_t=2}^{T} \gamma_t(3)}{\sum_{t=1}^{T} \gamma_t(3)} \end{bmatrix} = \begin{bmatrix} 0,635 & 0,365 \\ 0,57 & 0,43 \\ 0,75 & 0,25 \end{bmatrix}$$

Matriks tersebut merupakan penaksir untuk matriks emisi B. Untuk mencapai nilai $P(O|\hat{\lambda}) \geq P(O|\lambda)$ maka peluang Viola menang saat suasana hatinya senang adalah sebesar 0,635, peluang Viola menang saat suasana hatinya sedih adalah sebesar 0,57, peluang Viola menang saat suasana hatinya kesal adalah sebesar 0,75, peluang Viola kalah saat suasana hatinya senang adalah sebesar 0,365, peluang Viola kalah saat suasana hatinya sedih adalah sebesar 0,43, peluang Viola kalah saat suasana hatinya kesal adalah sebesar 0,25. Karena algoritma ini merupakan langkah iteratif maka langkah-langkah tersebut diulang sampai selisih antara elemen yang bersesuaian dalam setiap parameter mendekati nol (konvergen ke nol).