BAB III

Hidden Markov Models (HMM)

3.1 Pendahuluan

Rantai Markov mempunyai *state* yang dapat diobservasi secara langsung. Namun pada beberapa situasi tertentu yang ditemukan di kehidupan nyata, beberapa faktor yang tidak dapat diobservasi *(hidden)* mempengaruhi perhitungan kemungkinan perpindahan *state*. Untuk memasukkan faktor-faktor seperti itu ke dalam perhitungan, dibutuhkan suatu model yang lebih "pintar" yaitu *Hidden Markov Model* (HMM). Sebuah HMM menggabungkan dua atau lebih rantai Markov, dengan hanya satu rantai saja yang *state*-nya dapat diobservasi sementara rantai lainnya tidak dapat diobservasi *(hidden)* yang mempengaruhi hasil dari *state* yang terobservasi. Biasanya pada HMM banyaknya *hidden state* dan probabilitas transisi dari *hidden state* tersebut diketahui.

Ada dua parameter yang berkaitan dengan setiap *state* pada HMM:

1. *Emission probabilities* : menjelaskan peluang dari *output* berbeda yang mungkin dari suatu *state*.
2. *Transition probabilities* : menjelaskan peluang perpindahan dari *state* saat ini ke *state* berikutnya.

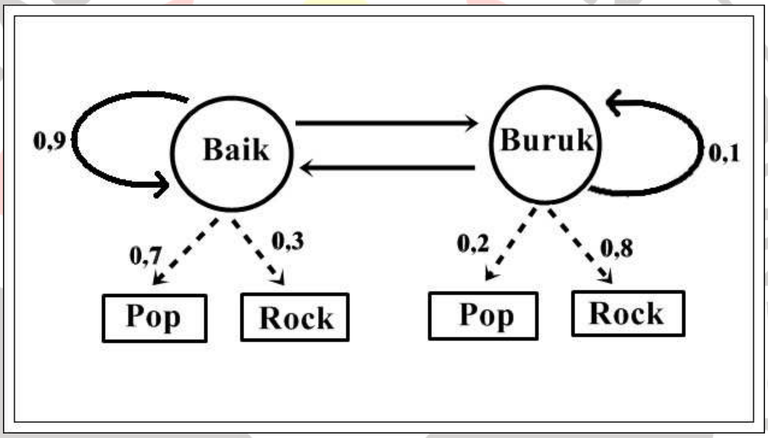
Contoh 3.1

Misalkan Bill menyukai dua jenis musik yaitu *Pop* dan *Rock*. Setiap hari Bill memilih satu jenis musik untuk ia dengarkan sepanjang hari tersebut. Setelah dilakukan pengamatan diperoleh informasi sebagai berikut: jika hari ini Bill mendengarkan jenis musik *Pop*, maka besok Bill akan mendengarkan musik *Pop* lagi dengan probabilitas 0,3, musik *Rock* dengan probabilitas 0,7. Sementara, jika hari ini Bill mendengarkan jenis musik *Rock*, maka besok Bill akan mendengarkan musik *Pop* dengan probabilitas 0,4, musik *Rock* dengan probabilitas 0,6

Permasalahan tersebut dapat diselesaikan dengan menggunakan rantai Markov dengan waktu dan ruang keadaan diskrit. Sementara untuk mengilustrasikan HMM, pada permasalahan di atas, misalkan ada informasi lain yang diketahui, yaitu bahwa jenis musik yang dipilih Bill bukan hanya dipengaruhi oleh pilihan lagu sebelumnya, tapi juga dipengaruhi oleh suasana hati Bill pada hari itu. Lebih jelasnya, jika suasana hati Bill baik, maka ia akan memilih jenis musik *Pop* dengan probabilitas 0,7, dan musik *Rock* dengan probabilitas 0,3. Sementara jika suasana hati Bill sedang buruk, ia akan memilih jenis musik *Pop* dengan probabilitas 0,2, musik *Rock* dengan probabilitas 0,8.

Permasalahannya adalah suasana hati Bill tidak terobservasi langsung, sehingga dikatakan bahwa suasana hati Bill “tersembunyi “ *(hidden)*. Namun keadaan tersembunyi ini berpengaruh terhadap perpindahan keadaan terobservasi. Pada contoh ini suasana hati Bill mempunyai pengaruh terhadap jenis musik yang dipilihnya.

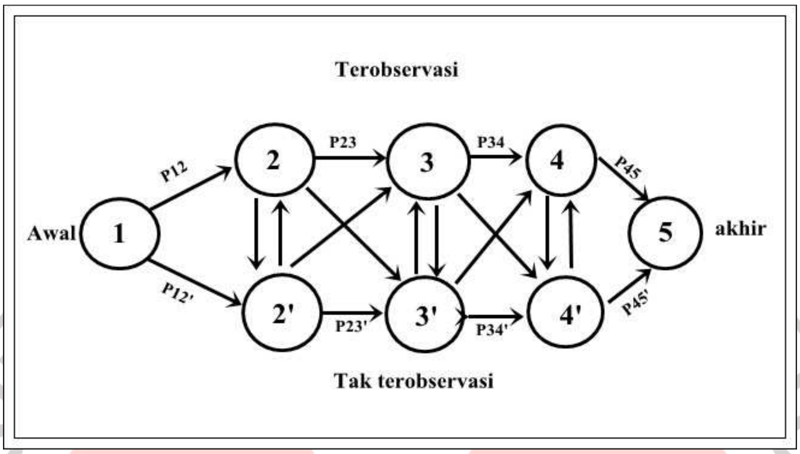
Selain informasi di atas, hal lain yang diketahui adalah peluang perubahan suasana hati Bill sebagai berikut, jika hari ini suasana hati Bill baik, maka besok suasana hati Bill akan baik dengan probabilitas 0,9, dan akan buruk dengan probabilitas 0,1. Sementara, jika hari ini suasana hati Bill buruk, maka besok suasana hati Bill akan baik dengan probabilitas 0,6, dan akan buruk dengan probabilitas 0,4. Untuk ilustrasi permasalahan di atas dapat dilihat pada gambar berikut,



Gambar 3.1 Ilustrasi Contoh Soal

3.2 Definisi HMM

HMM adalah sebuah proses stokastik ganda di mana salah satu prosesnya tidak dapat diobservasi (*hidden*). Proses yang tidak dapat diobservasi ini hanya dapat diobservasi melalui proses yang dapat diobservasi. Jika $X = \{X_1, X_2, \dots, \}$ adalah sebuah proses Markov, dan $O = \{O_1, O_2, \dots\}$ adalah sebuah fungsi dari X, maka X adalah sebuah HMM yang dapat diobservasi melalui O, atau dapat ditulis $O = f(X)$ untuk suatu fungsi f. Parameter X menyatakan *state process* yang tersembunyi (*hidden*), sementara parameter O menyatakan *observation process* yang dapat diobservasi. Untuk ilustrasi HMM dapat dilihat gambar 3.2 berikut,



Gambar 3.2 Ilustrasi HMM

HMM didefinisikan sebagai 5-*tuple* (5 pasangan di mana masing-masing anggota bisa berupa himpunan atau ukuran) sebagai berikut:

1) Banyaknya elemen keadaan tersembunyi *(hidden state)* pada model yang dinotasikan dengan N. Sebagai contoh, pada masalah pilihan musik Bill, di mana keadaan tersembunyinya adalah suasana hati baik, dan suasana hati buruk, sehingga pada kasus ini N=2. Atau dapat ditulis sebagai $X_t = i, i = 1$ (suasana hati baik), 2 (suasana hati buruk).
2) Matriks peluang tansisi $A = \{a_{ij}\}$ di mana a_{ij} adalah elemen dari A yang merupakan peluang bersyarat dari keadaan pada saat $t + 1$, jika diketahui keadaan X pada saat t, atau $a_{ij} = P(X_{t+1} = j | X_t = i)$ di mana $1 \leq i, j \leq N$. Karena itu A berukuran $N \times N$. Hal yang perlu dijadikan catatan adalah bahwa $a_{ij} \geq 0$ untuk setiap $1 \leq i, j \leq N$ dan $\sum_{j=1}^{N} a_{ij} = 1$ untuk setiap

$1 \leq i \leq N$. Artinya jumlah elemen masing-masing baris adalah 1.Untuk masalah pilihan jenis musik Bill, dengan keadaan $X_t = i, i = 1$ (suasana hati baik) dan 2(suasana hati buruk), akan ada matriks transisi berukuran 2×2.

$$A = \begin{pmatrix} 0,9 & 0,1 \\ 0,6 & 0,4 \end{pmatrix}$$

3) Banyaknya elemen keadaan yang terobservasi, M. M umumnya tetap, ditentukan oleh pengamat, tetapi M juga bisa dimisalkan sebagai variabel acak. Misalkan variabel acak dari suatu keadaan terobservasi adalah $K, k = 1,2, \dots, M$.Untuk masalah pilihan jenis musik Bill, banyaknya elemen hanya 2, yaitu musik *Pop*, dan musik *Rock*, sehingga $k = 1, 2$.

4) Distribusi peluang observasi pada saat *t*, pada keadaan *i*, yang biasa dikenal dengan matriks emisi $B = \{b_i(k)\}$, di mana

$$b_i^k = b_i(k) = P(O_t = k | X_t = i), 1 \leq i \leq N, 1 \leq k \leq M \tag{3.1}$$

K adalah observasi pada waktu ke-*t* bernilai *k*, jadi B adalah matriks berukuran $N \times M$, dan seperti pada matriks transisi A, jumlah elemen setiap baris adalah 1.

Pada contoh mengenai pilihan jenis musik Bill akan ada matriks B yang berukuran 2×2,

$$B = \begin{pmatrix} 0,7 & 0,3 \\ 0,2 & 0,8 \end{pmatrix}$$

5) Keadaan awal $\pi = \{\pi(i)\}$ di mana

$$\pi(i) = P(X_1 = i), 1 \leq i \leq N \tag{3.2}$$

Untuk masalah Bill,

$\pi(1) = P(suasana\ hati\ baik),\ \pi(2) = P(suasana\ hati\ buruk).$

Misalkan untuk keadaan awal pada contoh di atas suasana hati Bill bisa dimulai dalam suasana hati apapun, sehingga setiap jenis suasana hati memiliki peluang yang sama atau, $\pi = \begin{pmatrix} 0,5 \\ 0,5 \end{pmatrix}$

Istilah *tuple* di atas berkaitan dengan himpunan dan ukuran. Pada HMM himpunannya diwakili oleh variabel acak. Dari definisi di atas, cukup jelas bahwa dari nilai 5-*tuple* $(N, M, A, B\ dan\ \pi)$, terdapat tiga komponen yang merupakan ukuran (probabilitas), yaitu A, B, dan π. Akibatnya HMM lebih dikenal dengan notasi $\lambda = (A, B, \pi)$ dengan A berukuran $N \times N$ dan B berukuran $N \times M$.

3.3 Asumsi pada HMM

Ada tiga asumsi pokok yang dibutuhkan dalam analisis HMM, yaitu:

1. Asumsi Markov :

   Asumsi ini menyatakan bahwa keadaan berikutnya hanya dipengaruhi oleh keadaan saat ini. Model yang dihasilkan adalah HMM orde pertama. Pada beberapa kasus di kehidupan nyata, keadaan selanjutnya mungkin dipengaruhi oleh k keadaan sebelumnya, yang akan menghasilkan HMM orde ke-k yang lebih sulit untuk dianalisa daripada HMM orde pertama.

2. Asumsi stasioneritas

Asumsi ini menyatakan bahwa peluang transisi dari suatu keadaan ke keadaan lainnya independen dengan waktu saat transisi itu terjadi. Sehingga untuk sembarang t_1 dan t_2 berlaku :

$$P\left(X_{t_1+1} = j \middle| X_{t_1} = i\right) = P\left(X_{t_2+1} = j \middle| X_{t_2} = i\right) = P_{ij} \tag{3.3}$$

3. Asumsi independensi/kebebasan

Jika diketahui suatu barisan observasi $O = O_1, O_2, \dots, O_T$ dan suatu barisan keadaan $X = X_1, X_2, \dots, X_T$. Maka pengamatan saat ini bersifat independen secara statistik dengan pengamatan sebelumnya. Atau dapat dinyatakan:

$$P(O|X,\lambda) = \prod_{t=1}^{T} P(O_t|X_t,\lambda) \tag{3.4}$$

3.4 Masalah-masalah Utama dalam HMM

3.4.1 Menghitung Peluang Observasi

Bila diketahui sebuah model $\lambda = (A, B, \pi)$ dan sebuah barisan observasi $O = \{O_1, O_2, \dots, O_T\}$, kemudian akan dihitung $P(O|\lambda)$ yang dapat ditulis sebagai,

$$P(O|\lambda) = \sum_{X} P(O|X,\lambda)P(X|\lambda)$$

Di mana $X = \{X_1, X_2, \dots, X_T\}$ adalah suatu barisan, $P(O|X,\lambda)$ adalah probabilitas barisan observasi O untuk suatu barisan *state*, X, dan $P(X|\lambda)$ merupakan probabilitas dari X bila diberikan sebuah model. Karena pada HMM barisan observasi diasumsikan independen, maka

$$P(O|X,\lambda) = \prod_{t=1}^{T} P(O_t|X_t,\lambda) = b_1(O_1).b_2(O_2).\dots.b_T(O_T).$$

$$P(X|\lambda) = \pi(1)a_{12}a_{23}\dots.a_{\mathrm{T-1,T}}$$

Sehingga diperoleh,

$$P(O|\lambda) = \sum_{X} P(O|X,\lambda)P(X|\lambda)$$

$$= \sum_{1,2,\dots,T} \pi(1)b_1(o_1)\,a_{12}b_2(o_2)a_{23}\dots a_{\mathrm{T-1,T}}b_T(o_T)$$

Untuk menghitung $P(O|\lambda)$ diperlukan $2T.N^T$ kali operasi perhitungan, dengan N^T adalah kemungkinan *hidden state* yang terjadi jika barisan observasi sepanjang T dan *hidden state*-nya sebanyak N. Sehingga meskipun untuk N dan T yang bernilai kecil, jumlah operasi perhitungan yang dibutuhkan secara komputasional akan sangat banyak. Contohnya untuk masalah perubahan suasana hati Bill, N=3 dan misalkan panjangnya barisan observasi, T=100, Maka operasi perhitungan yang harus dilakukan adalah sebanyak $2 \times 100 \times 2^{100}$ kali. Karena itulah dibutuhkan algoritma yang lebih efisien untuk menyelesaikan masalah *evaluation*. Algoritma yang banyak digunakan untuk penyelesaian masalah *evaluation* adalah algoritma maju *(forward algorithm)* dan algoritma mundur *(backward algorithm)*.

3.4.2 Menentukan Barisan Keadaan Tersembunyi

Permasalahan kedua pada HMM adalah *decoding problem*, yaitu menemukan barisan *state* terbaik (optimal) yang berasosiasi dengan barisan observasi O dari sebuah model λ yang juga telah diketahui. Barisan *state* yang optimal didefinisikan sebagai barisan *state* yang mempunyai probabilitas tertinggi dalam menghasilkan barisan observasi yang telah diketahui sebelumnya. Sehingga pada akhirnya diperoleh suatu barisan state X yang akan memaksimumkan $P(X|O,\lambda)$. Namun, untuk suatu barisan observasi sepanjang T dan N *hidden states*, akan dihasilkan sebanyak N^T barisan yang mungkin untuk X. Untuk contoh mengenai perubahan suasana hati Bill, dengan $N = 2$dan $T = 100$, akan ada 2^{100} barisan yang mungkin.

Misal, didefinisikan $\gamma_t(i)$ di mana $\gamma_t(i) = P(X_t = i|O,\lambda)$. Jika $\gamma_t(i)$ dijumlahkan terhadap i, karena $x_t = i$ merupakan partisi dari X maka menurut aturan Bayes mengenai partisi, hasilnya menjadi

$$\sum_{i=1}^{N} \gamma_t(i) = P(x_t = i|O,\lambda) = 1 \tag{3.5}$$

Sehingga, bisa dinyatakan bahwa *state* yang paling optimal untuk masing-masing *t* bisa diperoleh dari:

$$X_t^* = \arg\max_{1\le i\le N} \gamma_t(i) \tag{3.6}$$

Dengan demikian akan dihasilkan barisan *states* yang paling optimal yaitu, $X^* = X_1^*, X_2^*, \dots, X_T^*$ untuk suatu observasi $O = O_1, O_2, \dots, O_T$ yang diberikan. Sayangnya, pencarian barisan *states* yang paling optimal dengan cara tersebut, berpeluang menghasilkan barisan yang tidak valid, karena tidak mempertimbangkan probabilitas transisi *state*. Contohnya, apabila hasil dari perhitungan $X^* = \{X_1^* = i, X_2^* = j, X_3^* = k\}$, sementara diketahui bahwa proses tidak mungkin berpindah dari *state* j ke *state* k, atau $P(X_{t+1} = k|X_t = j) = 0$. Karena itu, untuk menghindari masalah tersebut, perlu digunakan suatu metode yang mempertimbangkan probabilitas transisi *state* pada proses pencarian barisan *state* yang paling optimal. Metode yang banyak digunakan untuk penyelesaian masalah ini antara lain algoritma Viterbi dan entropi.

.

3.4.3 Menaksir Parameter-parameter HMM

Permasalahan ketiga berkaitan dengan bagaimana menentukan estimasi 3 parameter HMM A, B, dan π sehingga terbentuk model baru $\hat{\lambda}(\hat{A}, \hat{B}, \hat{\pi})$ di mana $P(O|\hat{\lambda}) \geq P(O|\lambda)$. Dengan kata lain, permasalahan ketiga adalah masalah optimasi, dan permasalahan yang harus dipecahkan adalah mengestimasi model terbaik yang dapat menjelaskan suatu barisan observasi.

Untuk menyelesaikan permasalahan terakhir pada HMM ini, biasanya digunakan algoritma Baum-Welch yang merupakan kasus khusus dari algoritma EM (Ekspektasi Maksimum). Algoritma EM sendiri merupakan algoritma yang digunakan untuk mempelajari model-model probabilistik dalam suatu kasus yang melibatkan keadaan-keadaan tersembunyi.

3.5 Beberapa Metode Penyelesaian Masalah-masalah pada HMM

3.5.1 Menghitung Peluang Observasi dengan Algoritma Maju (*The Forward Algorithm*)

Algoritma ini adalah proses iterasi yang didasarkan pada perhitungan peluang bersyarat melalui sifat-sifat pada peluang. Dengan menggunakan definisi peluang bersyarat $P(O|\lambda)$ dapat dihitung, namun operasi perhitungan yang dibutuhkan akan bertambah banyak karena operasinya akan naik secara eksponensial, seiring dengan bertambah panjangnya barisan observasi yang ada. Algoritma *forward* menyimpan nilai yang telah dihitung pada iterasi sebelumnya, sehingga mereduksi $2T.N^T$ menjadi N^2T operasi. Algoritma ini akan sangat efisien ketika panjang barisan observasinya cukup besar.

Didefinisikan $\alpha_t(i)$ sebagai variabel maju, dimana:

$$\alpha_t(i) = P(O_1, O_2, \dots, O_t, X_t = i|\lambda) \tag{3.7}$$

dengan $\alpha_t(i)$ menyatakan total peluang observasi yang berakhir pada *state i* pada saat t di mana $t = 1,2, \dots, T$ jika diketahui suatu barisan observasi $\{O_1, O_2, \dots, O_t\}$. Menurut Rabiner (1989), secara umum algoritma maju terdiri atas tiga bagian, yaitu:

1. Tahap inisialisasi

$$\alpha_1(i) = \pi(i)b_i(O_1) \text{ di mana } 1 \le i \le N \tag{3.8}$$

Persamaan tersebut diperoleh dari definisi variabel maju dengan cara mensubstitusikan dua definisi parameter HMM yaitu $\pi(i) = P(X_t = i)$ dan $b_i(k) = P(O_t = k|X_t = i)$:

$$\begin{aligned}\alpha_1(i) &= P(O_1, X_1 = i|\lambda)\\ &= P(O_1|X_1 = i, \lambda)P(X_1 = i, \lambda)\\ &= \pi(i)P(O_1|X_1 = i, \lambda)\\ &= \pi(i)b_i(O_1)\end{aligned}$$

2. Tahap induksi

$$\alpha_{t+1}(j) = \{\sum_{i=1}^{N} \alpha_t(i)a_{ij}\}b_j(O_{t+1})\ j = 1, \dots, N,\ t = 1, \dots, T-1 \qquad (3.9)$$

Pada tahap ini akan dihitung nilai α pada saat $t > 1$, sama seperti pada tahap inisialisasi, pembuktian dilakukan dengan mensubstitusikan dua parameter HMM yaitu $b_i(k) = P(O_t = k|X_t = i)$ dan a_{ij} sehingga diperoleh:

$$\begin{aligned}
\alpha_{t+1}(j) &= P(O_1, O_2, \dots, O_t, O_{t+1}, X_{t+1} = j|\lambda) \\
&= P(O_1, O_2, \dots, O_t, O_{t+1}|X_{t+1} = j, \lambda)P(X_{t+1} = j|\lambda) \\
&= P(O_1, O_2, \dots, O_t|X_{t+1} = j, \lambda)P(O_{t+1}|X_{t+1} = j, \lambda)P(X_{t+1} = j|\lambda) \\
&= P(O_1, O_2, \dots, O_t, X_{t+1} = j|\lambda)\ P(O_{t+1}|X_{t+1} = j, \lambda) \\
&= P(O_1, O_2, \dots, O_t, X_{t+1} = j|\lambda)b_j(O_{t+1}) \\
&= \left[\sum_{i=1}^{N} P(O_1, O_2, \dots, O_t, X_t = i, X_{t+1} = j|\lambda)\right] b_j(O_{t+1}) \\
&= \left[\sum_{i=1}^{N} P(O_1, O_2, \dots, O_t, X_t = i|\lambda)\ P(X_{t+1} = j|O_1, O_2, \dots, O_t, X_t = i, \lambda)\right] b_j(O_{t+1}) \\
&= \left[\sum_{i=1}^{N} P(O_1, O_2, \dots, O_t, X_t = i|\lambda)\ P(X_{t+1} = j|X_t = i, \lambda)\right] b_j(O_{t+1}) \\
&= \left[\sum_{i=1}^{N} \alpha_t(i)a_{ij}\right] b_j(O_{t+1})
\end{aligned}$$

3. Tahap terminasi

Pada tahap ini adalah menjumlahkan semua peluang gabungan dari observasi dan *hidden state* bila diketahui sebuah model sehingga diketahui peluang marjinal dari observasi tersebut atau ditulis:

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i) \qquad (3.10)$$

Contoh 3.2

Perhatikan kembali masalah pemilihan jenis musik Bill, misalkan setelah dilakukan pengamatan pilihan jenis musik Bill selama 4 hari berturut-turut adalah *Pop*, *Rock*, *Pop*. Untuk permasalahan pertama pada HMM, berarti yang diminta adalah menghitung peluang bahwa model $\lambda = (A, B, \pi)$ menghasilkan barisan observasi $O = \{pop, rock, pop\}$. Diketahui:

$$A = \begin{pmatrix} 0,9 & 0,1 \\ 0,6 & 0,4 \end{pmatrix}, B = \begin{pmatrix} 0,7 & 0,3 \\ 0,2 & 0,8 \end{pmatrix}$$

dan misalkan proses dapat dimulai dengan dengan suasana hati baik dan suasana hati buruk dengan peluang yang sama yaitu $\pi = \begin{pmatrix} 0,5 \\ 0,5 \end{pmatrix}$.

Penyelesaian:

Permasalahan tersebut akan diselesaikan dengan menggunakan algoritma maju, di mana panjang barisan observasi, $T = 3$.

1. Tahap inisialisasi

Pada tahap inisialisasi akan ada 2 buah α masing-masing untuk suasana hati baik (1), suasana hati biasa saja (2) dan suasana hati buruk,

$$\alpha_1(1) = \pi(1)b_1(pop) = (0,5)(0,7) = 0,35$$

$$\alpha_1(2) = \pi(2)b_2(pop) = (0,5)(0,2) = 0,1$$

2. Tahap induksi

Pada tahap ini akan dihitung $\alpha_2(j)\ dan\ \alpha_3(j)$

Tahap induksi untuk $t = 2$,

$$\alpha_2(j) = \left[\sum_{i=1}^{N} \alpha_1(i)a_{ij}\right] b_j(O_2)$$

$$\alpha_2(1) = [(\alpha_1(1)a_{11}) + (\alpha_1(2)a_{21}) +]b_1(O_2 = rock)$$

$$= [(0{,}35)(0{,}9) + (0{,}1)(0{,}6)](0{,}3) \; = 0{,}1125$$

$$\alpha_2(2) = [(\alpha_1(1)a_{12}) + (\alpha_1(2)a_{22})]b_2(O_2 = rock)$$

$$= [(0{,}35)(0{,}1) + (0{,}1)(0{,}4)](0{,}8) \; = 0{,}06$$

Tahap induksi untuk $t = 3$

$$\alpha_3(j) = \left[\sum_{i=1}^{N} \alpha_2(i)a_{ij}\right] b_j(O_3)$$

$$\alpha_3(1) = [(\alpha_2(1)a_{11}) + (\alpha_2(2)a_{21})]b_1(O_3 = pop)$$

$$= [(0{,}1125)(0{,}9) + (0{,}06)(0{,}6)](0{,}7) = 0{,}096075$$

$$\alpha_3(2) = [(\alpha_2(1)a_{12}) + (\alpha_2(2)a_{22})]b_2(O_3 = pop)$$

$$= [(0{,}1125)(0{,}1) + (0{,}06)(0{,}4)](0{,}2) = 0{,}00705$$

3. Tahap terminasi

Dari tahap terminasi maka algoritma menghasilkan penyelesaian sebagai berikut:

$$P(O = Pop, Rock, Pop|\lambda) = \sum_{i=1}^{N} \alpha_T(i)$$

$$= \alpha_3(1) + \alpha_3(2)$$

$$= 0{,}096075 + 0{,}00705$$

$$= 0{,}103125$$

3.5.2 Menghitung Peluang Observasi dengan Algoritma Mundur (*The Backward Algorithm*)

Langkah algoritma mundur hampirsama dengan algoritma maju. Namun bedanya, pada algoritma mundur inisialisasi didasarkan pada seluruh observasi yang ada. Jadi algoritma mundur mengganti $O_1, O_2, \dots, O_t$ pada persamaan (3.5) menjadi $O_{t+1}, O_{t+2}, \dots, O_T$.

$$\beta_t(i) = P(O_{t+1}, O_{t+2}, \dots, O_T | X_t = i, \lambda) \tag{3.11}$$

Tahap-tahap algoritma mundur dijelaskan sebagai berikut:

1. Tahap inisialisasi

$$\beta_T(i) = 1 \text{ untuk } i = 1,2, \dots, N \tag{3.12}$$

Pada tahap ini, dinyatakan $\beta_T(i) = 1$ karena diasumsikan i adalah *state* final, dan bernilai nol untuk i yang lainnya.

2. Tahap induksi

$$\beta_t(i) = \sum_{j=1}^{N} b_j(O_{t+1})\, \beta_{t+1}(j) a_{ij}$$

Untuk $t = T-1, T-2, \dots, 1$ dan $i = 1,2, \dots, N$

Bukti:

$\beta_t(i) = P(O_{t+1}, O_{t+2}, \dots, O_T | X_t = i, \lambda)$

$= \sum_{j=1}^{N} P(O_{t+1}, O_{t+2}, \dots, O_T, X_{t+1} = j | X_t = i, \lambda)$

$= \sum_{j=1}^{N} P(O_{t+1} | O_{t+2}, \dots, O_T, X_{t+1} = j, X_t = i, \lambda) P(O_{t+2}, \dots, O_T, X_{t+1}$

$= j | X_t = i, \lambda)$

$= \sum_{j=1}^{N} P(O_{t+1} | X_{t+1} = j, \lambda) P(O_{t+2}, \dots, O_T | X_{t+1} = j, X_t = i, \lambda)$

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

$$= \sum_{j=1}^{N} P(O_{t+1}|\,X_{t+1} = j,\lambda)P(O_{t+2},\dots,O_T|X_{t+1} = j,\lambda)P(X_{t+1} = j|X_t = i,\lambda)$$

$$= \sum_{j=1}^{N} P(O_{t+1}|\,X_{t+1} = j,\lambda)\beta_{t+1}(j)\,P(X_{t+1} = j|X_t = i,\lambda)$$

$$= \sum_{j=1}^{N} b_j(O_{t+1})\beta_{t+1}(j)a_{ij} \qquad (3.13)$$

Untuk $t = T-1, T-2, \dots, 1$ dan $i = 1,2,\dots,N$

3. Tahap Terminasi

$$P(O|\lambda) = \sum_{i=1}^{N} b_i\,(1)\pi(i)\beta_1(i) \qquad (3.14)$$

Contoh 3.3

Tinjau kembali masalah pemilihan jenis musik Bill yang dipengaruhi oleh suasana hatinya. Asumsikan bahwa setelah dilakukan pengamatan, diperoleh informasi bahwa jenis musik yang dipilih Bill selama tiga hari berturut-turut adalah,$Pop, Rock, Pop$. Akan dicari probabilitas bahwa model yang dimiliki akan menghasilkan barisan tersebut.

Penyelesaian:

Contoh 3.3 akan diselesaikan dengan menggunakan algoritma mundur, dengan $T = 3$.

1 Tahap inisialisasi

$$\beta_3(1) = \beta_3(2) = 1$$

2 Tahap induksi

Untuk $t = 2$, $O_3 = 2(rock)$

$$\beta_2(1) = [(b_1(O_3)\beta_3(1)a_{11} + (b_2(O_3)\beta_3(2)a_{12}]$$

$$= [(0{,}7)(1)(0{,}9) + (0{,}2)(1)(0{,}2)] = 0{,}65$$

$$\beta_2(2) = [(b_1(O_3)\beta_3(1)a_{21} + (b_2(O_3)\beta_3(2)a_{22}]$$

$$= [(0{,}7)(1)(0{,}6) + (0{,}2)(0{,}1)(0{,}4)] = 0{,}5$$

Untuk $t = 1$, $O_2 = 2(rock)$

$$\beta_1(1) = [(b_1(O_2)\beta_2(1)a_{11} + (b_2(O_2)\beta_2(2)a_{12}]$$

$$= [(0{,}3)(0{,}65)(0{,}9) + (0{,}8)(0{,}5)(0{,}1)] = 0{,}2155$$

$$\beta_1(2) = [(b_1(O_2)\beta_2(1)a_{21} + (b_2(O_2)\beta_2(2)a_{22}]$$

$$= [(0{,}3)(0{,}65)(0{,}6) + (0{,}8)(0{,}5)(0{,}4)] = 0{,}277$$

3 Tahap terminasi

$$P(O = Pop, Pop, Rock|\lambda) = \sum_{i=1}^{N} \beta_1(i)\,\pi(i) b_i(O_1)$$

$$= \beta_1(1)\pi(1)b_1(O_1) + \beta_1(2)\pi(2)b_2(O_1)$$

$$= 0{,}075425 + 0{,}0277 = 0{,}103125$$

Hasil dari algoritma mundur konsisten dengan solusi yang diperoleh dari algoritma maju, yaitu 0,103125.

3.5.3 Menentukan Barisan Keadaan Tersembunyi dengan Menggunakan Algoritma Viterbi

Algoritma Viterbi diperkenalkan oleh Andrew J. Viterbi pada tahun 1967. Algoritma ini pertama kali digunakan untuk menyelesaikan masalah pengkodean yang rumit, namun akhir-akhir ini algoritma Viterbi telah banyak digunakan untuk mempermudah penyelesaian masalah pada bidang-bidang lain. Salah satunya, algoritma Viterbi digunakan dalam HMM untuk mencari barisan keadaan tersembunyi yang paling optimal dari suatu barisan observasi.

Didefinisikan,

$$\arg\max_y \{z\} \tag{3.15}$$

Yaitu, argumen *y* yang bersesuaian dengan nilai maksimum dari *z*. Algoritma Viterbi memaksimalkan $P(X,O)$ dan probabilitas bersyarat $P(X|O)$ secara bersamaan berdasarka fakta bahwa

$$arg\max_X\{P(X|O,\lambda)\} = \arg\max_X\left\{\frac{P(X,O|\lambda)}{P(O|\gamma)}\right\}$$

Algoritma Viterbi mendefinisikan:

$$\delta_t(i) = \max_{X_1,X_2,\ldots,X_{t-1}} P(O_1,O_2,\ldots,O_t,X_1,X_2,\ldots,X_{t-1},X_t = i|\lambda) \tag{3.16}$$

dan

$$\psi_t(j) = arg\max_{i\le 1\le N}\{\delta_{t-1}(i)a_{ij}\} \tag{3.17}$$

Variabel $\delta_t(i)$ menyatakan probabilitas terbesar sepanjang *t* observasi pertama dan berakhir pada *state i*. Sehingga $\delta_t(i)$ merupakan probabilitas dari barisan *state* yang paling optimal untuk barisan observasi secara parsial. Sementara $\psi_t(j)$ menyimpan *state* sebelumnya yang akan membentuk barisan *state* yang paling optimal.

Algoritma Viterbi terdiri atas empat tahap:

1. Tahap inisialisasi

Pada saat t=1,

$$\delta_1(i) = P(X_1 = i, O_1)$$

$$= P(O_1|X_1 = i)P(X_1 = i)$$

Dengan mensubstitusi asumsi awal pada HMM yaitu $b_i(k) = P(O_t = k|X_t = i)$ dan $\pi(i) = P(X_t = i)$ diperoleh:

$$\delta_1(i) = b_i(O_1)\pi(i)$$

Pada tahap ini

$$\psi_1(i) = 0$$

2. Tahap rekursi

Pada tahap rekursi,

$$\delta_t(j) = \max_{X_1,X_2,\dots,X_{t-1}} P(O_1, O_2, \dots, O_{t-1}, O_t, X_1, X_2, \dots, X_{t-1}, X_t = j|\lambda)$$

$$= \max_{X_1,X_2,\dots,X_{t-1}} P\{(O_t|O_1, O_2, \dots, O_t, X_1, X_2, \dots, X_{t-1}, X_t = j, \lambda)$$

$$P(O_1, O_2, \dots, O_{t-1}, X_1, X_2, \dots, X_{t-1}, X_t = j, \lambda)\}$$

$$= \max_{X_1,X_2,\dots,X_{t-1}}\{P(O_t|X_t = j, \lambda)P(O_1, O_2, \dots, O_{t-1}, X_1, X_2, \dots, X_{t-1}, X_t = j, \lambda)\}$$

$$= P(O_t|X_t = j, \lambda) \max_{X_1,X_2,\dots,X_{t-2}} \max_{1\le i\le N}\{P(O_1, O_2, \dots, O_{t-1}, X_1, X_2, \dots, X_{t-2}, X_{t-1} = i, X_t = j, \lambda)\}$$

$$= b_j(O_t) \max_{X_1,X_2,\dots,X_{t-2}} \max_{1\le i\le N}\{P(X_t = j|X_{t-1} = i)P(O_1, O_2, \dots, O_{t-1}, X_1, X_2, \dots, X_{t-2}, X_{t-1} = i, \lambda)\}$$

$$= b_j(O_t) \max_{1\le i\le N}\left\{P(X_t = j|X_{t-1} = i) \max_{X_1,X_2,\dots,X_{t-2}} P(O_1, O_2, \dots, O_{t-1}, X_1, X_2, \dots, X_{t-2}, X_{t-1} = i, \lambda)\right\}$$

$$= b_j(O_t) \max_{1\le i\le N}\{P(X_t = j|X_{t-1} = i)\delta_{t-1}(i)\}$$

$$= b_j(O_t) \max_{1 \le i \le N} \{a_{ij} \delta_{t-1}(i)\}$$

3. Tahap terminasi

$$P^* = \max_{1 \le i \le N} \{\delta_T(i)\} \quad (3.18)$$

$$X_T{}^* = arg \max_{1 \le i \le N} \{\delta_T(i)\} \quad (3.19)$$

4. Tahap *backtracking*

$$X_T{}^* = \psi_{t+1}(X_{t+1}{}^*) \, , t = T-1, T-2, \dots, 1 \quad (3.20)$$

Tahap *backtracking* memungkinkan barisan *state* yang paling optimal ditemukan dari titik terakhir yang disimpan pada tahap rekursi.

Contoh 3.4:

Perhatikan masalah pemilihan jenis musik Bill seperti yang diilustrasikan pada Gambar 3.2. Setelah dilakukan pengamatan, jenis musik yang dipilih Bill selama tiga hari berturut-turut adalah: $Pop, rock, pop$. Permasalahan kedua pada HMM untuk permasalahan ini adalah bagaimana menentukan barisan *hidden states* yang optimal pada hal ini yaitu suasana hati Bill yang paling mungkin menyebabkan Bill memilih jenis musik sesuai dengan barisan observasi tersebut.

Penyelesaian:

1. Tahap inisialisasi

Dengan $O_1 = 1(pop)$

$$\delta_1(1) = \pi(1) b_1(O_1) = (0{,}5)(0{,}7) = 0{,}35$$

$$\delta_1(2) = \pi(2) b_2(O_1) = (0{,}5)(0{,}2) = 0{,}1$$

$$\psi_1(1) = \psi_1(2) = 0$$

2. Tahap rekursi

Untuk $t = 2, O_2 = 2(rock)$

$$\delta_2(1) = \max\{\delta_1(1)a_{11}, \delta_1(2)a_{21},\} \, b_1(O_2 = 2)$$

$$= \max\{(0{,}35)(0{,}9), (0{,}1)(0{,}6)\} \, 0{,}3$$

$$= \max\{(0{,}315), (0{,}06)\} \, (0{,}3) = 0{,}09455$$

$$\psi_2(1) = 1(suasana\ hati\ baik)$$

$$\delta_2(2) = \max\{\delta_1(1)a_{12}, \delta_1(2)a_{22}\} \, b_2(O_2 = 2)$$

$$= \max\{(0{,}35)(0{,}1), (0{,}1)(0{,}4)\}(0{,}8)$$

$$= \max\{(0{,}035), (0{,}04)\} \, (0{,}8) = 0{,}032$$

$$\psi_2(2) = 2(suasana\ hati\ buruk)$$

Untuk $t = 3, O_3 = 1(pop)$

$$\delta_3(1) = \max\{\delta_2(1)a_{11}, \delta_2(2)a_{21}\} \, b_1(O_3 = 1)$$

$$= \max\{(0{,}09455)(0{,}9), (0{,}032)(0{,}06)\} \, (0{,}7)$$

$$= \max\{(0{,}085095), (0{,}0192)\} \, (0{,}7) = 0{,}0595665$$

$$\psi_3(1) = 1(suasana\ hati\ baik)$$

$$\delta_3(2) = \max\{\delta_2(1)a_{12}, \delta_2(2)a_{22}\} \, b_2(O_3 = 1)$$

$$= \max\{(0{,}09455)(0{,}1), (0{,}032)(0{,}4)\}(0{,}2)$$

$$= \max\{(0{,}009455), (0{,}0128)\} \, (0{,}2) = 0{,}00256$$

$$\psi_3(2) = 2(suasana\ hati\ buruk)$$

3. Tahap terminasi

$$P^* = \max\{\delta_3(1), \delta_3(2)\}$$

$$= \max\{(0{,}0595665), (0{,}00256)\} = 0{,}0595665$$

$$X_3{}^* = arg \max\{\delta_3(1), \delta_3(2)\} = 1(suasana\ hati\ baik$$

4. Tahap *backtracking*

$$X_t^{\ *} = \psi_{t+1}(X_{t+1}^{\ *})$$

$$X_2^{\ *} = \psi_3(X_3^{\ *}) = \psi_3(1) = 1(suasana\ hati\ baik)$$

$$X_1^{\ *} = \psi_2(X_2^{\ *}) = \psi_2(1) = 1(suasana\ hati\ baik)$$

Jadi, saat Bill memilih jenis musik dengan urutan $Pop, Rock, Pop$ barisan keadaan tersembunyi (dalam hal ini suasana hati) yang paling optimal adalah: $X^* = \{1(suasana\ hati\ baik), 1(suasana\ hati\ baik), 1(suasana\ hati\ baik)\}$.

3.5.4 Penaksiran Parameter-parameter HMM dengan Algoritma Baum-Welch

Algoritma Baum-Welch juga dikenal sebagai algoritma maju-mundur dengan variabel maju dan variabel mundurnya didefinisikan sebagai:

$$\begin{cases} \alpha_t(i) = P(O_1, O_2, \dots, O_t, X_T = i|\lambda) \\ \beta_t(i) = P(O_{t+1}, O_{t+2}, \dots, O_t, X_T = i|\lambda) \end{cases} \tag{3.21}$$

Kemudian didefinisikan sebuah variabel baru $\xi_t(i,j)$ di mana $\xi_t(i,j)$ adalah probabilitas proses berada pada *state-i* pada waktu *t* dan berada pada *state-j* pada waktu *j* bila diketahui barisan observasi dan model:

$$\xi_t(i,j) = P(X_t = i, X_{t+1} = j|O,\lambda) \tag{3.22}$$

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 *state i* pada waktu t , $\gamma_t(i)$ dengan menjumlahkan $\xi_t(i,j)$ atas *j*.

$$\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i,j) \tag{3.23}$$

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

$$\hat{\pi}(i) = \gamma_1(i) \tag{3.24}$$

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 *state i* ke *state j* dengan total seluruh transisi dari *state 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 *state* yang menghasilkan observasi *j* pada saat proses berada pada *state i* dengan jumlah seluruh proses yang berada pada *state i*.

Contoh 3.5

Perhatikan kembali masalah pemilihan musik yang dilakukan oleh Bill. Dengan menggunakan informasi yang sama dengan contoh-contoh sebelumnya, akan dicari penaksir parameter unuk $\hat{\lambda} = (\hat{A}, \hat{B}, \hat{\pi})$.

Penyelesaian:

Diketahui,

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

Untuk $t = 1$,

$$\xi_1(1,1) = \frac{\alpha_1(1)a_{11}b_1(O_2)\beta_2(1)}{P(O|\lambda)} = \frac{(0{,}35)(0{,}9)(0{,}3)(0{,}65)}{0{,}103125} = 0{,}5956363$$

$$\xi_1(1,2) = \frac{\alpha_1(1)a_{12}b_2(O_2)\beta_2(2)}{P(O|\lambda)} = \frac{(0{,}35)(0{,}1)(0{,}8)(0{,}5)}{0{,}103125} = 0{,}1357575$$

$$\xi_1(2,1) = \frac{\alpha_1(2)a_{21}b_1(O_2)\beta_2(1)}{P(O|\lambda)} = \frac{(0{,}1)(0{,}6)(0{,}3)(0{,}65)}{0{,}103125} = 0{,}1134545$$

$$\xi_1(2,2) = \frac{\alpha_1(2)a_{22}b_2(O_2)\beta_2(2)}{P(O|\lambda)} = \frac{(0{,}1)(0{,}4)(0{,}8)(0{,}5)}{0{,}103125} = 0{,}1551515$$

Untuk $t = 2$,

$$\xi_2(1,1) = \frac{\alpha_2(1)a_{11}b_1(O_3)\beta_3(1)}{P(O|\lambda)} = \frac{(0{,}1125)(0{,}9)(0{,}7)(1)}{0{,}103125} = 0{,}6872727$$

$$\xi_2(1,2) = \frac{\alpha_2(1)a_{12}b_2(O_3)\beta_3(2)}{P(O|\lambda)} = \frac{(0{,}1125)(0{,}1)(0{,}2)(1)}{0{,}103125} = 0{,}2183936$$

$$\xi_2(2,1) = \frac{\alpha_2(2)a_{21}b_1(O_3)\beta_3(1)}{P(O|\lambda)} = \frac{(0,06)(0,6)(0,7)(1)}{0,103125} = 0,2443636$$

$$\xi_2(2,2) = \frac{\alpha_2(2)a_{22}b_2(O_3)\beta_3(2)}{P(O|\lambda)} = \frac{(0,06)(0,4)(0,2)(1)}{0,103125} = 0,0465454$$

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

Untuk $t = 1$,

$$\gamma_1(1) = [0,5956363 + 0,1357575] = 0,731394$$

$$\gamma_1(2) = [0,1134545 + 0,1551515] = 0,268606$$

Untuk $t = 2$,

$$\gamma_2(1) = [0,6872727 + 0,2183936] = 0,9056656$$

$$\gamma_2(2) = [0,2443636 + 0,0465454] = 0,290909$$

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

$$\hat{\pi} = \begin{bmatrix} \gamma_1(1) \\ \gamma_1(2) \end{bmatrix} = \begin{bmatrix} 0,731394 \\ 0,268606 \end{bmatrix}$$

Nilai di atas merupakan taksiran peluang awal. Artinya agar nilai $P(O|\hat{\lambda}) \geq P(O|\lambda)$ terpenuhi, maka probabilitas proses berada pada *state* “Suasana hati baik” adalah sebesar 0,731394 dan taksiran peluang awal bahwa proses berada pada “Suasana hati buruk adalah sebesar 0,268606.

$$\hat{A} = \begin{bmatrix} \frac{\sum_{t=1}^{T} \xi_t(1,1)}{\sum_{t=1}^{T} \gamma_t(1)} & \frac{\sum_{t=1}^{T} \xi_t(1,2)}{\sum_{t=1}^{T} \gamma_t(1)} \\ \frac{\sum_{t=1}^{T} \xi_t(2,1)}{\sum_{t=1}^{T} \gamma_t(2)} & \frac{\sum_{t=1}^{T} \xi_t(2,2)}{\sum_{t=1}^{T} \gamma_t(2)} \end{bmatrix} = \begin{bmatrix} \frac{1,282909}{1,6370594} & \frac{0,354151}{1,6370594} \\ \frac{0,3578181}{0,559515} & \frac{0,2016996}{0,559515} \end{bmatrix} = \begin{bmatrix} 0,78 & 0,32 \\ 0,64 & 0,36 \end{bmatrix}$$

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 probabilitas transisi dari "Suasana hati baik" ke "Suasana hati baik" adalah sebesar 0,78, dari "Suasana hati baik" ke "Suasana hati buruk" sebesar 0,32, dari "Suasana hati buruk"ke "Suasana hati baik" sebesar 0,64, dan dari "Suasana hati buruk" ke "Suasana hati buruk" sebesar 0,36.

$$\hat{B} = \begin{bmatrix} \frac{\sum_{t=1,O_t=1}^{T} \gamma_t(1)}{\sum_{t=1}^{T} \gamma_t(1)} & \frac{\sum_{t=1,O_t=2}^{T} \gamma_t(1)}{\sum_{t=1}^{T} \gamma_t(1)} \\ \frac{\sum_{t=1,O_t=1}^{T} \gamma_t(2)}{\sum_{t=1}^{T} \gamma_t(2)} & \frac{\sum_{t=1,O_t=2}^{T} \gamma_t(2)}{\sum_{t=1}^{T} \gamma_t(2)} \end{bmatrix} = \begin{bmatrix} \frac{0,7313938 + 0}{1,6370595} & \frac{0 + 0,9056656}{1,6370595} \\ \frac{0,268606 + 0}{0,559515} & \frac{0 + 0,290909}{0,559515} \end{bmatrix}$$

$$= \begin{bmatrix} 0,45 & 0,55 \\ 0,48 & 0,52 \end{bmatrix}$$

Matriks tersebut merupakan penaksir untuk matriks emisi B. Untuk mencapai nilai $P(O|\hat{\lambda}) \geq P(O|\lambda)$ maka probabilitas Bill memilih musik *Pop* saat suasana hatinya baik adalah sebesar 0,45, probabilitas Bill memilih musik *Rock* saat suasana hatinya baik seb esar 0,55, probabilitas Bill memilih musik *Pop* saat suasana hatinya buruk adalah sebesar 0,48, dan probabilitas Bill memilih musik *Rock* saat suasana hatinya buruk adalah sebesar 0,52.