Hidden Markov Model

Kezia Tumilaar[1], Yohanes Langi[2], Altien Rindengan[3]

[1)] Program Studi Matematika, FMIPA, Universitas Sam Ratulangi, kezykey@gmail.com
[2)]Program Studi Matematika, FMIPA, Universitas Sam Ratulangi, yarlangi@gmail.com
[3)]Program Studi Matematika, FMIPA, Universitas Sam Ratulangi, altien@unsrat.ac.id

Abstract

Hidden Markov Models (HMM) is a stochastic model and is essentially an extension of Markov Chain. In Hidden Markov Model (HMM) there are two types states: the observable states and the hidden states. The purpose of this research are to understand how hidden Markov model (HMM) and to understand how the solution of three basic problems on Hidden Markov Model (HMM) which consist of evaluation problem, decoding problem and learning problem. The result of the research is hidden Markov model can be defined as $\lambda = (A, B, \pi)$. The evaluation problem or to compute probability of the observation sequence given the model $P(O|\lambda)$ can solved by Forward-Backward algorithm, the decoding problem or to choose hidden state sequence which is optimal can solved by Viterbi algorithm and learning problem or to estimate hidden Markov model parameter $\lambda = (A, B, \pi)$ to maximize $P(O|\lambda)$ can solved by Baum – Welch algorithm. From description above Hidden Markov Model with state 3 can describe behavior from the case studies.

Key words: Decoding Problem, Evaluation Problem, Hidden Markov Model, Learning Problem

Hidden Markov Model

Abstrak

Hidden Markov Models (HMM) adalah sebuah model stokastik dan pada dasarnya merupakan perluasan dari rantai Markov. Dalam Hidden Markov Models (HMM), ada dua tipe state: state yang dapat diobservasi dan state yang tersembunyi. Tujuan dari penelitian ini adalah mengerti bagaimana Hidden Markov Models (HMM) dan memahami bagaimana solusi dari tiga masalah dasar pada Hidden Markov Models (HMM) yang terdiri dari evaluation problem, decoding problem dan learning problem. Hasil dari penelitian ini adalah Hidden Markov Models dapat didefinisikan sebagai $\lambda = (A, B, \pi)$. Evaluation Problem atau menghitung peluang observasi yang diberikan model $P(O|\lambda)$ dapat diselesaikan dengan algoritma forward-backward, decoding problem atau memilih urutan state tersembunyi yang optimal dapat diselesaikan dengan algoritma viterbi dan learning problem atau mengestimasi parameter Hidden Markov Models (HMM) $\lambda = (A, B, \pi)$ untuk memaksimumkan $P(O|\lambda)$ dapat diselesaikan dengan algoritma Baum-Welch. Dari deskripsi diatas, Hidden Markov Models (HMM) dengan state 3 mampu menggambarkan perilaku dari studi kasus.

Kata Kunci *: Decoding Problem, Evaluation Problem, Hidden Markov Model, Learning Problem*

1. Introduction

A stochastic process which is the simplest generalization of a sequence of independent random variables is a discrete-time Markov chain. The future probabilistic behaviour of the process depends only on the present state of the process and is not influenced by its past history, this is called the Markovian property. In every state Markov chain can be observed directly. But sometimes there is a sequence of a state that wants to be known but can't be observed directly but only can observed by observable state because it , is called the hidden Markov model. In this research, to understand the hidden Markov model wil be given case studies about weather change probabilities then solving three basic problems in hidden Markov models which consist evaluation problem , decoding problem and learning problem. In this research also using computer help.

2. Hidden Markov Models

A Hidden Markov Model (HMM) is a tool for representing probability distributions over sequences of observations [1].

A HMM is usually characterized by the following elements [2]:

1) N, the number of hidden states in the model. Donate the individual states as
$$S = \{s_1, s_2, \dots, s_N\}, \tag{1}$$
And the state at the length t as Q_t.
2) M, the number of distinct observation symbols per hidden state. Donate the individual symbols as
$$V = \{v_1, v_2, \dots, v_M\} \tag{2}$$
And the symbol at the length t as O_t.
3) The state transition probability matrix, $[A]_{ij} = \{a_{ij}\}$ where
$$a_{ij} = P(Q_{t+1} = s_j | Q_t = s_i), \quad 1 \le i, j \le N \tag{3}$$
4) The observation symbol probability matrix in hidden state j, $[B]_{jk} = \{b_j(v_k)\}$, where
$$b_j(v_k) = P(O_t = v_k | Q_t = s_j), \quad 1 \le j \le N, \quad 1 \le k \le M \tag{4}$$
5) The initial state distribution $\Pi = \{\pi_i\}$ where
$$\pi_i = P(Q_1 = s_i),\ 1 \le i \le N \tag{5}$$

Given appropriate values of N, M, A, B and π, the HMM can be used as a generator to give an observation sequence
$$O = \{O_1 O_2 \dots O_T\} \tag{6}$$
Where T is the number of observations in the sequence. For simplicity, using the compact notation
$\lambda = (A, B, \pi)$ (7)
To indicate the complete parameter set of the HMM.

In Markov Chain generally can't be observed directly, because it, is called the hidden Markov models. Hidden Markov models have two states namely observation state and hidden state. In bioinformatics, if given the DNA sequence ACAATGGT [3], the DNA sequence is an observation state and the hidden state can vary, depending on what information is to be obtained.

3. The Three Basic Problems for HMM

There are three basic problems of interest that must be solved for the model to be useful in real-world applications. These problems are the following [4]:

1) Evaluation Problem
Given the observation sequence $O = \{O_1 O_2 \dots O_T\}$, and $\lambda = (A, B, \pi)$, how do efficiently compute $P(O|\lambda)$ the probability of the observation sequence, given the model?
2) Decoding Problem
Given the observation sequence $O = \{O_1, O_2,, \dots, O_T\}$, and model $\lambda = (A, B, \pi)$, how to choose a corresponding state sequence $Q = q_1 q_2 \dots q_T$ which is optimal in some meaningful sense (i.e., best "explains" the observations)?
3) Learning Problem
How choose the model parameters $\lambda = (A, B, \pi)$ to maximize $P(O|\lambda)$?

4. Research Method

Research methods to use in this research is literature review as follows:
Studying about hidden Markov models and three basic problems on the Hidden Markov Model
.

1) Given example case studies about outside weather (hidden states) and mood sequence (observation state) where according to Markovian Property and Stochastic Process. The case studies adapted in *Hidden Markov Model: A Tutorial for the Course Computational Intelligence*.
2) Determine elements of hidden Markov model in case studies which consist the number of hidden states (N), the number of symbols observable in states (M), the state transition probability distribution (A), the observation symbol probability distribution (B) and the initial state distribution Π.
3) By elements of hidden Markov model in case study, generate mood sequence (observation states) for 7 days with software.

4) By elements of hidden Markov model in case study and caretaker mood sequence generated solve the three basic problems of hidden Markov model and the solution which consists evaluation problem solved by Forward – Backward algorithm, to solved decoding problem by Viterbi algorithm and as well as learning problem will be solved by Baum – Welch algorithm.

5. Results And Discussions

5.1 Hidden Markov Model

The formal definition of a Hidden Markov Model is a follows [5]:

$$\lambda = (A, B, \pi)$$

Elements of an Hidden Markov Model:

1) N is the number of hidden states, $S = \{s_1, s_2, \dots, s_N\}$ the state sequence of length t as Q_t.
2) M is the number of symbols observable in states $V = \{v_1, v_2, \dots, v_M\}$ the symbol at the length t as O_t.
3) A is the state transition probability distribution $A = \{a_{ij}\}$, $1 \le i, j \le N$
4) B is the Observation symbol probability distribution $B = \{b_j(v_k)\}$, $1 \le j \le N$, $1 \le k \le M$
5) Π is the initial state distribution $\pi = P(Q_1 = s_i)$, $1 \le i \le N$.

5.2 Case Studies

Suppose someone confined in a closed room for a few days and didn't know what was going on outside weather. If he was asked to guess the weather conditions outside the observation that he did was to look at the caretaker mood who comes in to the room whether happy or sad. Assume the weather sunny, rainy and cloudy. After the observations obtained the following information: if the outside weather conditions is sunny, the caretaker mood will be happy with probability 0.8 and sad with probability 0.2. If the outside weather conditions is rainy, the caretaker mood will be happy with probability 0.1 and sad with probability 0.9. Meanwhile, if the outside weather is cloudy , the caretaker mood will be happy with probability 0.6 and sad with probability 0.4. In addition to the above information, the other thing that is known is the transition probability of the outside weather , if sunny today then tomorrow will be sunny with a probability 0.8, it will rainy with probability 0.04 and cloudy with probability 0.16. If it's raining then tomorrow will be sunny with a probability 0.2, it will rain with probability 0.6 and cloudy with probability 0.2. Meanwhile, if the day is cloudy, then tomorrow will be sunny with a probability 0.1, it will rain with probability 0.4 and cloudy with probability 0.5. Assumed for the initial state of the case of the caretaker mood the sunny state with probability 0.8, the rainy state with probability 0 and the cloudy state with probability 0.2 [6]. Compute the probability of observation mood sequence $P(O|\lambda)$ for seven days, how outside weather which is optimal for seven days ahead and estimation the model $\lambda = A, B, \pi$ to maximize $P(O|\lambda)$.

5.3 Elements of Hidden Markov Model In Case Studies

The element of hidden Markov model in case studies as follows:

1) N is the number of hidden states. In the case of the caretaker mood who come into the room, where hidden state is sunny (1), rainy (2) and cloudy (3), so in this case N = 3.
2) M is the number of symbols observable in states. In the case of the caretaker mood who come into the room, where the observations are happy and sad that in this case M = 2.
3) $A = [a_{ij}]$ is the state transition probability matrix. In the case of the caretaker mood who come into the room, with state is sunny (1), rainy (2) and cloudy (3) there will be matrix transition sized 3x3

$$A = \begin{bmatrix} 0.8 & 0.04 & 0.16 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.4 & 0.5 \end{bmatrix}$$

4) $B = [b_{jk}]$ is the observation symbol probability matrix. In the case of the caretaker mood who come into the room, with observation happy and sad there will be matrix B sized 3x2

$$B = \begin{bmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \\ 0.6 & 0.4 \end{bmatrix}$$

5) $\pi = \{\pi_i\}$ is the initial state distribution. In the case of the caretaker mood who come into the room, $\pi(1) = P(sunny)$, $\pi(2) = P(rainy)$, $\pi(3) = P(cloudy)$. Assumed for the initial state of mood caretaker cases who come into the room, the sunny state with probability 0.8, the rainy state with probability 0 and the cloudy state with probability 0.2 , $\pi = \begin{bmatrix} 0.8 \\ 0 \\ 0.2 \end{bmatrix}$

5.4 Generate Mood Sequence with Software

To generate mood random sequence for 7 days, given transition matrix A and emission matrix B and have the result as follows:
MOODseq = 'Happy' 'Sad' 'Happy' 'Happy' 'Happy' 'Sad' 'Happy'

5.5 The Three Basic Problems in HMM and the Solution

5.5.1 The Three Basic Problems in HMM and the Solution (Mood Sequence for 3 Days) Using Software

1. *The Probability of the Mood Sequence (Evaluation Problem) solved by Forward Backward Algorithm*

The probability of the mood sequence for 7 days, given transition matrix A and emission matrix B P(O|λ) using Forward-Backward algorithm have the result as follows:

```
 PSTATES  =    0.8532  0.7181  0.8160  0.7884  0.6799  0.5341  0.6168
               0.0070  0.1423  0.0253  0.0162  0.0316  0.2360  0.0664
               0.1398  0.1396  0.1587  0.1954  0.2885  0.2299  0.3168
logpseq   =  -4.7423
FORWARD =
        1.0000  0.8649  0.4649  0.6605  0.7240  0.7508  0.3676  0.6168
             0  0.0054  0.2659  0.0522  0.0275  0.0223  0.3536  0.0664
             0  0.1297  0.2691  0.2873  0.2485  0.2269  0.2788  0.3168
BACKWARD =
        1.0000   0.9865   1.5445   1.2353   1.0890   0.9055   1.4528   1.0000
        0.4928   1.2918   0.5352   0.4846   0.5891   1.4187   0.6675   1.0000
        0.6134   1.0778   0.5187   0.5526   0.7863   1.2716   0.8246   1.0000
Scale =  1.0000  0.7400   0.3037   0.5475   0.6272   0.6495   0.3415   0.5094
```

So when the caretaker who come into the room with mood sequence for the first day until the seventh day O = {Happy, Sad, Happy, Happy, Happy, Happy, Sad, Happy}, the logarithm of the probability for that's mood sequence is log 0.0000181 = -4,7423 so the probability for that's mood sequence is 0.0000181.

2. *Choose Hidden State Sequence which is Optimal (Decoding Problem) solved by Viterbi Algorithm*

The hidden state sequence which is optimal corresponding with mood sequence observation for 7 days and if given transition probability matrix A and emission matrix B using Viterbi algorithm and have the result as follows:
States = 'Sunny' 'Sunny' 'Sunny' 'Sunny' 'Sunny' 'Sunny' 'Sunny'
So, the hidden state sequence (in this case outside weather state which is optimal) for the first day until the seventh day is $q^* = \{$1(Sunny), 1(Sunny), 1(Sunny), 1(Sunny), 1(Sunny), 1(Sunny), 1(Sunny).

3. *Estimation of The Hidden Markov Model Parameter to Maximize $P(O|\lambda)$ (Learning Problem) solved by Baum – Welch Algorithm*

Estimation of the hidden Markov model parameter given the sequence observation for 7 days and transition probability matrix A and emission matrix B using Baum-Welch algorithm and have the result as follows:

estimateA =	0.2000	0.0000	0.8000
	1.0000	0.0000	0.0000
	0.0000	0.8000	0.2000
estimateB =	1.0000	0.0000	
	0.0000	1.0000	
	1.0000	0.0000	

5.5.2 The Three Basic Problems in HMM and the Solution (Mood Sequence for 3 Days)

The three basic problems in HMM and the solution as follows [7]:
The three basic problems in HMM and the solution with mood sequence for three days as follows:

1. *The Probability of the Observation Sequence (Evaluation Problem) solved by Forward – Backward Algorithm*

In the case of the caretaker mood who come into the room, after the observation, the caretaker mood observation in three days a row is happy, sad, happy. Compute the probability of the observation sequence O = {Happy, Sad, Happy) given the model λ = (A, B, π).

Solution:
Forward Algorithm
1) Initialization

$a_1(1) = \pi_1 b_1(O_1) = (0.8)(0.8) = 0.64$
$a_1(2) = \pi_2 b_2(O_1) = (0)(0.1) = 0$
$a_1(3) = \pi_3 b_3(O_1) = (0.2)(0.6) = 0.12$

2) Induction

Using equation (2)

$t = 2,\ O_2 = sad$
$a_2(j) = \left[\sum_{i=1}^{N} a_1(i) a_{ij}\right] b_j(O_2)$
$a_2(1) = [(0.64)(0.8) + (0)(0.2) + (0.12)(0.1)]0.2 = 0.1048$
$a_2(2) = [(0.0256) + (0) + (0.048)]0.9 = 0.06624$
$a_2(3) = [(0.1024) + (0) + (0.06)]0.4 = 0.06496$
$t = 3,\ O_3 = Happy$
$a_3(j) = \left[\sum_{i=1}^{N} a_2(i) a_{ij}\right] b_j(O_3)$
$a_3(1) = [(0.08384) + (0.013248) + (0.006496)]0.8 = 0.08286$
$a_3(2) = [(0.004192) + (0.03974) + (0.02598)]0.1 = 0.00699$
$a_3(3) = [(0.016768) + (0.013248) + (0.03248)]0.6 = 0.03749$

3)Termination

$$P(O = Happy, Sad, Happy|\lambda) = \sum_{i=1}^{N} a_T(i)$$
$$= a_3(1) + a_3(2) + a_3(3)$$
$$= 0.08286 + 0.00699 + 0.03749 = 0.13$$

Backward Algorithm:
1)Initialization : $\beta_T(i) = 1$
$\beta_3(1) = \beta_3(2) = \beta_3(3) = 1$

2)Induction :

$t = 2, O_3 = Happy$
$\beta_2(i) = \sum_{i=1}^{N} b_j(O_3)\, \beta_3(j) a_{ij}$
$\beta_2(1) = [(0.64) + (0.004) + (0.096)] = 0.74$
$\beta_2(2) = [(0.16) + (0.06) + (0.12)] = 0.34$
$\beta_2(3) = [(0.08) + (0.04) + (0.3)] = 0.42$

$t = 1, O_2 = Sad$
$\beta_1(i) = \sum_{i=1}^{N} b_j(O_2)\, \beta_2(j) a_{ij}$
$\beta_1(1) = [(0.1184) + (0.01224) + (0.02688)] = 0.15752$
$\beta_1(2) = [(0.0296) + (0.1836) + (0.0336)] = 0.2468$
$\beta_1(3) = [(0.0148) + (0.0544) + (0.189)] = 0.2582$
$P\big(O = Happy,\ Sad,\ Happy\big|\lambda\big) = \sum_{i=1}^{N} \beta_1(i)\pi(i) b_i(O_1)$
$= 0.10081 + 0 + 0.03098 = 0.13$

2. *Choose Hidden State Sequence which is Optimal (Decoding Problem) solved by Viterbi Algorithm*

In the case of the caretaker mood who come into the room, after the observation, the caretaker mood sequence in three days a row is happy, sad, happy and given the model λ = (A, B, π). Choose hidden state sequence which is optimal
Solutions:
1)Initialization
$\delta_1(i) = \pi_i b_i(O_1)$
$\psi_1(i) = 0$
$\delta_1(1) = \pi_1 b_1(O_1) = (0.8)(0.8) = 0.64$
$\delta_1(2) = \pi_2 b_2(O_1) = (0)(0.1) = 0$
$\delta_1(3) = \pi_3 b_3(O_1) = (0.2)(0.6) = 0.12$
$\psi_1(1) = \psi_1(2) = \psi_1(3) = 0$
2)Recursion
$\delta_t(j) = \max_{1 \le i \le N} \left[\delta_{t-1}(i) a_{ij}\right] b_j(O_t)$
$\psi_t(j) = \operatorname{argmax}_{1 \le i \le N} \left[\delta_{t-1}(i) a_{ij}\right]$
For t=2, $O_2 = 2(Sad)$
$\delta_2(1) = max\{(0.512), (0), (0.012)\}0.2 = 0.1024$
$\psi_2(1) = 1(Sunny)$
$\delta_2(2) = max\{(0.0256), (0), (0.048)\}0.9 = 0.0432$
$\psi_2(2) = 1(Cloudy)$
$\delta_2(3) = max\{(0.1024), (0), (0.06)\}0.4 = 0.04096$
$\psi_2(3) = 1(Sunny)$
For t=3, $O_3 = 1(Happy)$
$\delta_3(1) = max\{(0.08192), (0.00864), (0.00409)\}0.8 = 0.06553$
$\psi_3(1) = 1(Sunny)$
$\delta_3(2) = max\{(0.00409), (0.02592), (0.01638)\}0.1 = 0.00259$
$\psi_3(2) = 2(Rainy)$
$\delta_3(3) = max\{(0.01638), (0.00864), (0.02048)\}0.6 = 0.01228$
$\psi_3(3) = 3(Cloudy)$
3)Termination
$P^* = \max_{1 \le i \le N} [\delta_T(i)]$ $\qquad q_T^* = \operatorname{argmax}_{1 \le i \le N} [\delta_T(i)]$
$P^* = \max\{\delta_3(1), \delta_3(2), \delta_3(3)\}$
$= \max\{(0.06553), (0.00259), (0.01228)\} = 0.06553$
$q_3^* = argmax\{\delta_3(1), \delta_3(2), \delta_3(3)\} = 1\ (Sunny)$
4)Backtracking
$q_t^* = \psi_{t+1}(q_{t+1}^*)$
$q_2^* = \psi_3(q_3)^* = \psi_3(1) = Sunny$ $\qquad q_1^* = \psi_2(q_2)^* = \psi_2(1) = Sunny$

So, when the caretaker come into the room with mood sequence happy, sad, happy the hidden state sequence (in this case outside weather state which is optimal) is q^* ={1(Sunny),1(Sunny), 1(Sunny)}.

3. *Estimation Hidden Markov Model Parameter to Maximize $P(O|\lambda)$ (Learning Problem) solved by Baum – Welch Algorithm*

In the case of the caretaker mood who come into the room, after the observation, the caretaker mood sequence in three days a row is happy, sad, happy. Estimation parameter $\bar{\lambda} = \bar{A}, \bar{B}, \bar{\pi}$ to maximize $P(O|\lambda)$.
Solution:
Unknown $\xi_t(i,j) = \frac{a_t(1)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)}{P(O|\lambda)}$
The result as follows:

For t = 1	For t = 2
$\xi_1(1,1) = \frac{(0.64)(0.8)(0.2)(0.74)}{0.13} = 0.58289$	$\xi_2(1,1) = \frac{(0.1048)(0.8)(0.8)(1)}{0.13} = 0.51593$
$\xi_1(1,2) = \frac{(0.64)(0.04)(0.9)(0.34)}{0.13} = 0.06025$	$\xi_2(1,2) = \frac{(0.1048)(0.04)(0.1)(1)}{0.13} = 0.00322$
$\xi_1(1,3) = \frac{(0.64)(0.16)(0.4)(0.42)}{0.13} = 0.13233$	$\xi_2(1,3) = \frac{(0.1048)(0.16)(0.6)(1)}{0.13} = 0.07739$
$\xi_1(2,1) = \frac{(0)(0.2)(0.2)(0.74)}{0.13} = 0$	$\xi_2(2,1) = \frac{(0.06624)(0.2)(0.8)(1)}{0.13} = 0.08152$
$\xi_1(2,2) = \frac{(0)(0.6)(0.9)(0.34)}{0.13} = 0$	$\xi_2(2,2) = \frac{(0.06624)(0.6)(0.1)(1)}{0.13} = 0.03057$
$\xi_1(2,3) = \frac{(0)(0.2)(0.4)(0.42)}{0.13} = 0$	$\xi_2(2,3) = \frac{(0.06624)(0.2)(0.6)(1)}{0.13} = 0.06114$
$\xi_1(3,1) = \frac{(0.12)(0.1)(0.2)(0.74)}{0.13} = 0.01366$	$\xi_2(3,1) = \frac{(0.06496)(0.1)(0.8)(1)}{0.13} = 0.03997$
$\xi_1(3,2) = \frac{(0.12)(0.4)(0.9)(0.34)}{0.13} = 0.11298$	$\xi_2(3,2) = \frac{(0.06496)(0.4)(0.1)(1)}{0.13} = 0.01998$
$\xi_1(3,3) = \frac{(0.12)(0.5)(0.4)(0.42)}{0.13} = 0.07753$	$\xi_2(3,3) = \frac{(0.06496)(0.5)(0.6)(1)}{0.13} = 0.14991$

Of these results can be search value $\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i,j)$
For t=1

$$\gamma_1(1) = [0.58289 + 0.06025 + 0.13233] = 0.8$$
$$\gamma_1(2) = [0 + 0 + 0] = 0$$
$$\gamma_1(3) = [0.01366 + 0.11298 + 0.07753] = 0.2$$

For t=2

$$\gamma_2(1) = [0.51593 + 0.00322 + 0.07739] = 0.59$$
$$\gamma_2(2) = [0.08152 + 0.03057 + 0.06114] = 0.17$$
$$\gamma_2(3) = [0.03997 + 0.01998 + 0.14991] = 0.21$$

Then, by using the calculations results can be found HMM estimation is $\bar{\lambda} = \bar{A}, \bar{B}, \bar{\pi}$.

$$\bar{\pi} = \begin{bmatrix} \gamma_1(1) \\ \gamma_1(2) \\ \gamma_1(3) \end{bmatrix} = \begin{bmatrix} 0.8 \\ 0 \\ 0.2 \end{bmatrix}$$

$\bar{\pi}$ is estimation of the initial state probability matrix. So that $P(O|\bar{\lambda}) \geq P(O|\lambda)$ fulfilled, initial state probability when process is in the “sunny” state is 0.8, the initial state probability estimation when process is in the “rainy” state is 0 and the initial state probability estimation when process is in the “cloudy” state is 0.2.

$$\bar{A} = \begin{bmatrix} \frac{\sum_{t=1}^{T} \xi_t(1,1)}{\sum_{t=1}^{T} \gamma_t(1)} & \frac{\sum_{t=1}^{T} \xi_t(1,2)}{\sum_{t=1}^{T} \gamma_t(1)} & \frac{\sum_{t=1}^{T} \xi_t(1,3)}{\sum_{t=1}^{T} \gamma_t(1)} \\ \frac{\sum_{t=1}^{T} \xi_t(2,1)}{\sum_{t=1}^{T} \gamma_t(2)} & \frac{\sum_{t=1}^{T} \xi_t(2,2)}{\sum_{t=1}^{T} \gamma_t(2)} & \frac{\sum_{t=1}^{T} \xi_t(2,3)}{\sum_{t=1}^{T} \gamma_t(2)} \\ \frac{\sum_{t=1}^{T} \xi_t(3,1)}{\sum_{t=1}^{T} \gamma_t(3)} & \frac{\sum_{t=1}^{T} \xi_t(3,2)}{\sum_{t=1}^{T} \gamma_t(3)} & \frac{\sum_{t=1}^{T} \xi_t(3,3)}{\sum_{t=1}^{T} \gamma_t(3)} \end{bmatrix} = \begin{bmatrix} 0.8 & 0.05 & 0.15 \\ 0.5 & 0.2 & 0.3 \\ 0.1 & 0.3 & 0.6 \end{bmatrix}$$

$\bar{A}$ is estimation of the transitions matrix A. So that $P(O|\bar{\lambda}) \geq P(O|\lambda)$ fulfilled, transition probability from "sunny" state to "sunny" state is 0.8, to "rainy" state is 0.05 and "cloudy" state is 0.15. Transition probability from "rainy" state to "sunny" state is 0.5, to "rainy" state is 0.2 and "cloudy" state is 0.3. Transition probability from "cloudy" state to "sunny" state is 0.1, to "rainy" state is 0.3 and "cloudy" state is 0.6.

$$\bar{B} = \begin{bmatrix} \frac{\sum_{t=1,O_t=1}^{T} \gamma_t(1)}{\sum_{t=1}^{T} \gamma_t(1)} & \frac{\sum_{t=1,O_t=2}^{T} \gamma_t(1)}{\sum_{t=1}^{T} \gamma_t(1)} \\ \frac{\sum_{t=1,O_t=1}^{T} \gamma_t(2)}{\sum_{t=1}^{T} \gamma_t(2)} & \frac{\sum_{t=1,O_t=2}^{T} \gamma_t(2)}{\sum_{t=1}^{T} \gamma_t(2)} \\ \frac{\sum_{t=1,O_t=1}^{T} \gamma_t(3)}{\sum_{t=1}^{T} \gamma_t(3)} & \frac{\sum_{t=1,O_t=2}^{T} \gamma_t(3)}{\sum_{t=1}^{T} \gamma_t(3)} \end{bmatrix} = \begin{bmatrix} 0.6 & 0.4 \\ 0 & 1 \\ 0.5 & 0.5 \end{bmatrix}$$

$\bar{B}$ is estimation of the emission matrix B. So that $P(O|\bar{\lambda}) \geq P(O|\lambda)$ fulfilled, the probability of caretaker with "happy" mood when the outside weather is "sunny" state is 0.6, when the outside weather is "rainy" is 0 and when the outside weather is "cloudy" state is 0.5 . Then, the probability of caretaker with "sad" mood when the outside weather is "sunny" state is 0.4, when the outside weather is "rainy" state is 1 and when the outside weather is "cloudy" state is 0.5.

6. Conclusions

Based of the research and literature studies that have been done about Hidden Markov Model, it can be concluded as follows:

1. Hidden Markov Model consists of five elemen which consists of N is the number of hidden states, M is the number of observable state, A is the transition probability matrix, B is the observable probability matrix or emission matrix and π is the initial state probability matrix. So, Hidden Markov Model can be defined as $\lambda = (A, B, \pi)$.
2. Hidden Markov model have three basic problems and solution for the every problem as follows the first, evaluation problem or to compute observation sequence probability $P(O|\lambda)$ can solved by Forward-Backward algorithm. The second, decoding problem or to choose the hidden state sequence which is optimal $Q^* = \{Q_1^*, Q_2^*, \dots, Q_T^*\}$ can solved by Viterbi algorithm. The third, learning problem or to estimate parameter of the hidden Markov model $\lambda = (A, B, \pi)$ to maximize $P(O|\lambda)$ can solved by Baum-Welch algorithm. From description above Hidden Markov Model with state 3 can describe behaviour from the case studies .

7. Reference

[1] Ghahramani, Z. 2001. An Introduction Hidden Markov Models and Bayesian Network. *International Journal of Pattern Recognition and Artifizial Intelligence*. 15(1) : 9 - 42.

[2] Rabiner,L.R. 1989. A Tutorial on Hidden Markov Models and Selected Aplications in Speech Recognition, *Proceedings of The IEEE*, 77(2) : 257-286.

[3] Fonzo, V.D., F.A. Pentini and V. Parisi. 2007. Hidden Markov Models in Bioinformatics. *Bentham Science Publishers Ltd*. 2(1) : 49 – 61.

[4] Ching, W.K and M. Ng. 2006. *Markov Chains: Models, Algorithms and Applications*. Springer Science+Business Media, Inc. USA.

[5] Singh, B, N. Kapur and P. Kaur 2012. Speech Recognition with Hidden Markov Model : A Review. *International Journal of Advanced Reasearch in Computer Science and Software Engineering*. 2(3):401-403.

[6] Resch, B. 2015. Hidden Markov Model : *A Tutorial for Course Computational Intelligence*. http://www.igi.tugraz.at/lehre/CI. [February 19th , 2015].

[7] Universitas Pendidikan Indonesia. 2015. Hidden Markov Model : http://a-reasearch.upi.edu/operator/upload/ta_mtk_0706664_chapter3.pdf. [February 5th , 2015].