Factorial Hidden Markov Models(FHMM)

山下 滉

2017/10/20

1　Hidden Markov Model

1.1　Markov Chain

「コイントスして，表なら左へ一歩，裏なら右へ一歩」を延々と繰り返すという最も基本的な確率過程のことを**ランダムウォーク（酔歩）**という．左右の確率を変えたり前後左右へ動くようにしたりと，バリエーションはいろいろ考えられるが，ここでは単純に一次元で左右等確率の設定とする．形式的に書くと，$+1$ か -1 かが半々の確率で出る i.i.d[1]な確率変数たち $Z_1, Z_2, Z_3, ...$ を使って，

$$X_0 = 0, \quad X_t = X_{t-1} + Z_t \ (t = 1, 2, ...) \tag{1.1}$$

と表される X_t のことである [1].

ランダムウォークの場合，

$$P(X_{t+1} = x_{t+1} | X_t = x_t, X_{t-1} = x_{t-1}, ..., X_0 = x_0) = P(X_{t+1} = x_{t+1} | X_t = x_t) \tag{1.2}$$

が成り立つ．つまり，未来の状態は今の状態だけから定まり，過去の履歴（どこからどんな経路をたどって今の状態にたどりついたか）には無関係であった．現在の状態が一時点前の状態に依存して確率的に決まるような特性を**マルコフ性（Markov property）**という．そしてこのような確率過程を**マルコフ過程（Markov process）**という．その中でも特に，X_t がとり得る値が有限とおりなものを**マルコフ連鎖（Markov chain）**と呼ぶ.

マルコフモデルは複数の状態を持ち，ある状態から別の状態（元の状態も含む）へ一定の確率で遷移する．この確率を**遷移確率（transition probability）**という．また，現在の状態に依存した一定の確率で特定の**出力記号（output symbol）**を出力する．この確率を**出力確率（output probability）**という [1].

[1] 独立同一分布（independent and indentically distributed; i.i.d）

1.2　モデル概要

例として，2 種類のサイコロ (状態)ω_1,ω_2 を投げて出た目を観測する場合を考える．ここではサイコロの目 (出力記号) として奇数と偶数の 2 種類を考える．

初期確率 $\boldsymbol{\pi}$，遷移確率行列 $\boldsymbol{A}$，出力確率行列 $\boldsymbol{B}$ が以下のように与えられたとする．

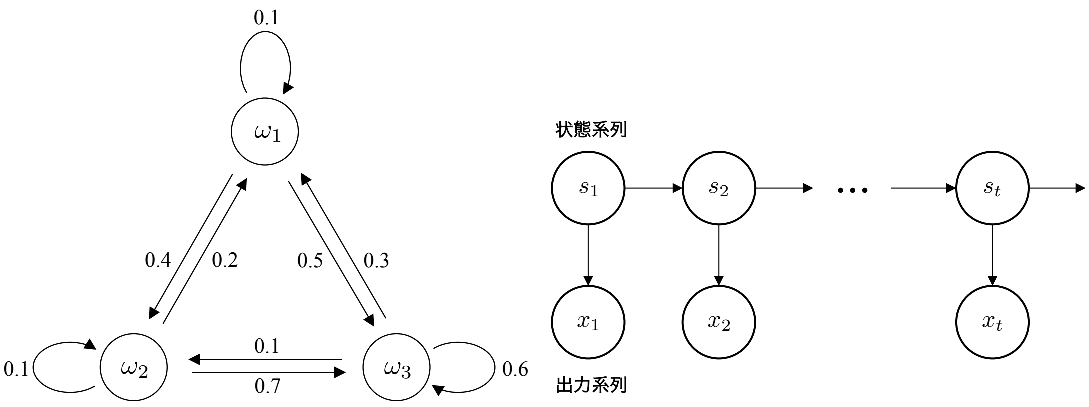
$$\boldsymbol{\pi} = \begin{array}{c} \begin{array}{cc}\omega_1 & \omega_2\end{array} \\ \begin{pmatrix}0.2 & 0.8\end{pmatrix}\end{array} \tag{1.3}$$

$$\boldsymbol{A} = \begin{array}{cc} & \begin{array}{cc}\omega_1 & \omega_2\end{array} \\ \begin{array}{c}\omega_1\\ \omega_2\end{array} & \begin{pmatrix}0.7 & 0.3\\ 0.6 & 0.4\end{pmatrix}\end{array} \tag{1.4}$$

$$\boldsymbol{B} = \begin{array}{cc} & \begin{array}{cc}\text{奇数} & \text{偶数}\end{array} \\ \begin{array}{c}\omega_1\\ \omega_2\end{array} & \begin{pmatrix}0.5 & 0.5\\ 0.9 & 0.1\end{pmatrix}\end{array} \tag{1.5}$$

行列 $\boldsymbol{A}$ と初期確率 $\boldsymbol{\pi}$ により，遷移確率 π_i，a_{ij} を反映した状態遷移系列が得られ，行列 $\boldsymbol{B}$ により，各状態からの出力確率 b_{jk} を反映した出力記号系列が得られる．状態遷移と記号の出力の様子を図 1(a) に示す．

時刻 t における状態を s_t，出力記号を x_t として，グラフィカルモデルで表したものを図 1(b) に示す．このモデルにおいて，状態系列と出力系列の両方を観測することができるものを**マルコフモデル (Markov Model)**, 状態系列が観測できず出力系列のみを観測することができるものを**隠れマルコフモデル (Hidden Markov Model; HMM)** という．



(a) 状態遷移図　　(b) 状態空間モデル

図 1: マルコフモデル

1.3 確率モデル

Notation of HMM

N	: 状態数
M	: 出力記号の数
$s_t \in \{\omega_1, \omega_2, ..., \omega_N\}$	: 時点 t での状態
$x_t \in \{v_1, v_2, ..., v_M\}$	: 時点 t での観測結果 (出力記号)
$\mathbf{s} = s_1 s_2 \cdots s_t \cdots s_T$	: 状態系列
$\mathbf{x} = x_1 x_2 \cdots x_t \cdots x_T$	: 観測記号系列
π_i	: 初期状態 $(t = 1)$ が ω_i である確率 $P(s_1 = \omega_i)$
a_{ij}, $a(\omega_i, \omega_j)$	: 状態 ω_i から状態 ω_j への遷移確率 $P(s_t = \omega_j \| s_{t-1} = \omega_i)$
b_{jk}, $b(\omega_j, v_k)$	: 状態 ω_j で v_k を出力する確率 $P(x_t = v_k \| s_t = \omega_j)$
$\boldsymbol{\pi} = (\pi_1, \pi_2, ..., \pi_N)$	: π_i を成分としてもつ N 次元のベクトル
$\boldsymbol{A}$	: a_{ij} を (i, j) 成分としてもつ $N \times N$ の行列
$\boldsymbol{B}$	: b_{jk} を (j, k) 成分としてもつ $M \times M$ の行列

1.3.1 状態遷移確率

$$a_{ij} = P(s_t = \omega_j | s_{t-1} = \omega_i) \tag{1.6}$$

1.3.2 出力確率

$$b_{jk} = P(x_t = v_k | s_t = \omega_j) \tag{1.7}$$

1.4 パラメータ推定

隠れマルコフモデルでは，最尤推定を用いて観測系列よりパラメータ推定を行う．推定すべきパラメータ $\boldsymbol{\pi}$, $\boldsymbol{A}$, $\boldsymbol{B}$ をまとめて

$$\boldsymbol{\theta} = (\boldsymbol{\pi}, \boldsymbol{A}, \boldsymbol{B}) \tag{1.8}$$

とする．

最尤推定により最適なパラメータ $\boldsymbol{\theta}$ を求めることは，$P(\mathbf{x}; \boldsymbol{\theta})$ を $\boldsymbol{\theta}$ に関して最大化することである．そのためには EM アルゴリズムを適用し，Q 関数を最大化すれば良い．

$$Q(\boldsymbol{\theta}^0, \boldsymbol{\theta}) = \sum_{\mathbf{s}} P(\mathbf{s}|\mathbf{x}; \boldsymbol{\theta}^0) \log P(\mathbf{x}, \mathbf{s}; \boldsymbol{\theta}) \tag{1.9}$$

ここで，$\boldsymbol{\theta}$ は $\boldsymbol{\theta}^0$ を更新した結果得られる新しいパラメータである．Q 関数を最大化するには，パラメータを繰り返し更新すればよい．

式 (1.9) を式変形すると，

$$Q(\boldsymbol{\theta}^0, \boldsymbol{\theta}) = \sum_{\mathbf{s}} P(\mathbf{s}|\mathbf{x};\boldsymbol{\theta}^0) \log P(s_1) + \sum_{\mathbf{s}} P(\mathbf{s}|\mathbf{x};\boldsymbol{\theta}^0) \sum_{t=1}^{n-1} \log a(s_t, s_{t+1}) + \sum_{\mathbf{s}} P(\mathbf{s}|\mathbf{x};\boldsymbol{\theta}^0) \sum_{t=1}^{n} \log b(s_t, x_t) \tag{1.10}$$

$$= Q(\boldsymbol{\theta}^0, \boldsymbol{\pi}) + Q(\boldsymbol{\theta}^0, \boldsymbol{A}) + Q(\boldsymbol{\theta}^0, \boldsymbol{B}) \tag{1.11}$$

が得られる．ここで，

$$Q(\boldsymbol{\theta}^0, \boldsymbol{\pi}) \stackrel{\text{def}}{=} \sum_{\mathbf{s}} P(\mathbf{s}|\mathbf{x};\boldsymbol{\theta}^0) \log P(s_1) \tag{1.12}$$

$$Q(\boldsymbol{\theta}^0, \boldsymbol{A}) \stackrel{\text{def}}{=} \sum_{\mathbf{s}} P(\mathbf{s}|\mathbf{x};\boldsymbol{\theta}^0) \sum_{t=1}^{n-1} \log a(s_t, s_{t+1}) \tag{1.13}$$

$$Q(\boldsymbol{\theta}^0, \boldsymbol{B}) \stackrel{\text{def}}{=} \sum_{\mathbf{s}} P(\mathbf{s}|\mathbf{x};\boldsymbol{\theta}^0) \sum_{t=1}^{n} \log b(s_t, x_t) \tag{1.14}$$

と定義した．上の 3 式は，それぞれパラメータ $\boldsymbol{\pi}$, $\boldsymbol{A}$, $\boldsymbol{B}$ のみを含む．したがって，$Q(\boldsymbol{\theta}^0, \boldsymbol{\theta})$ を $\boldsymbol{\theta}$ について最大化するには，$Q(\boldsymbol{\theta}^0, \boldsymbol{\pi})$, $Q(\boldsymbol{\theta}^0, \boldsymbol{A})$, $Q(\boldsymbol{\theta}^0, \boldsymbol{B})$ を，それぞれパラメータ $\boldsymbol{\pi}$, $\boldsymbol{A}$, $\boldsymbol{B}$ について最大化すればよい．ここで，$\boldsymbol{\theta}^0$ は更新前のパラメータであるので，最大化に当たっては定数とみなしてよい．

本手法は EM アルゴリズムに則っているので，パラメータ $\boldsymbol{\pi}$, $\boldsymbol{A}$, $\boldsymbol{B}$ を適当な初期値に設定し，各パラメータの更新式を反復的に計算することにより，より良い推定値が得られることが保証されている．この計算方法は**バウム・ウェルチアルゴリズム (Baum-Welch algorithm)** と呼ばれている [2,3]．ただし，得られる解は最適解である保証はなく，一般的には局所的最適解である．

式の導出など詳しい情報は，私の過去のゼミの資料や，参考文献 [2]8 章，参考文献 [4]13 章 2 節を参照してほしい．

2 Continuous Hidden Markov Model

2.1 モデル概要

Continuous Hidden Markov Model は，出力記号が離散値ではなく連続値である隠れマルコフモデルである．1 章で述べた隠れマルコフモデルでは，各状態から出力記号ごとに出力確率があったが，出力記号が連続値をとると，出力確率を設定することができない．そこで Continuous Hidden Markov Model では，出力が連続確率密度分布に従うとし，各状態から各連続確率密度分布の選択確率を定めることで，出力を連続値とする．

2.2 確率モデル

Notation of Continuous HMM

N	: 状態数
M	: 状態内の確率分布の数 (クラスタ数)
$s_t \in \{\omega_1, \omega_2, ..., \omega_N\}$	: 時点 t での状態
$m_t \in \{c_1, c_2, ..., c_M\}$	: 時点 t でのクラスタ
$x_t \in \{v_{11}, ..., v_{1M}, v_{N1}, ..., v_{NM}\}$	: 時点 t での観測結果 (出力記号)
$\mathbf{s} = s_1 s_2 \cdots s_t \cdots s_T$	: 状態系列
$\mathbf{m} = m_1 m_2 \cdots m_t \cdots s_T$	: クラスタ系列
$\mathbf{x} = x_1 x_2 \cdots x_t \cdots x_T$	: 観測記号系列
π_i	: 初期状態 $(t = 1)$ が ω_i である確率 $P(s_1 = \omega_i)$
a_{ij}, $a(\omega_i, \omega_j)$	: 状態 ω_i から状態 ω_j への遷移確率 $P(s_t = \omega_j \| s_{t-1} = \omega_i)$
g_{jk}	: 状態 ω_j からクラスタ c_k を出力する確率 $P(m_t = c_k \| s_t = \omega_j)$
μ_{ik}	: クラスタ v_{ik} の平均
σ_{ik}^2	: クラスタ v_{ik} の分散
$\boldsymbol{\pi} = (\pi_1, \pi_2, ..., \pi_N)$	: π_i を成分としてもつ N 次元のベクトル
$\boldsymbol{A}$	: a_{ij} を (i, j) 成分としてもつ $N \times N$ の行列
$\boldsymbol{G}$	: g_{jk} を (j, k) 成分としてもつ $N \times M$ の行列
$\boldsymbol{\mu} = (\mu_{11}, ..., \mu_{1M}, \mu_{N1}, ..., \mu_{NM})$	: μ_{ik} を成分としてもつ NM 次元のベクトル
$\boldsymbol{\sigma} = (\sigma_{11}, ..., \sigma_{1M}, \sigma_{N1}, ..., \sigma_{NM})$	: σ_{ik} を成分としてもつ NM 次元のベクトル

2.2.1 状態遷移確率

$$a_{ij} = P(s_t = \omega_j | s_{t-1} = \omega_i) \tag{2.1}$$

2.2.2 出力確率

$$P(x_t|s_t=\omega_j) = \sum_{k=1}^{M} P(m_t=c_k|s_t=\omega_j)P(x_t|s_t=\omega_j, m_t=c_k) \tag{2.2}$$

$$= \sum_{k=1}^{M} g_{jk}\cdot\phi(x_t,\mu_{jk},\sigma_{jk}^2) \tag{2.3}$$

ここで,

$$g_{jk} = P(m_t=c_k|s_t=\omega_j) \tag{2.4}$$

$$\phi(x_t,\mu_{jk},\sigma_{jk}^2) = P(x_t|s_t=\omega_j, m_t=c_k) \tag{2.5}$$

とした.

g_{jk} は,状態 ω_j からその状態内のクラスタ c_k が選択される確率である.また,g_{jk} には次の性質がある.

$$\sum_{k=1}^{M} g_{jk} = 1 \tag{2.6}$$

$\phi(\cdot)$ はカーネル関数と呼ばれ,正規分布が用いられる.

$$\phi(x_t,\mu_{jk},\sigma_{jk}^2) = \frac{1}{\sqrt{2\pi\sigma_{jk}^2}}\exp\left(-\frac{(x_t-\mu_{jk})^2}{2\sigma_{jk}^2}\right) \tag{2.7}$$

2.3 パラメータ推定

推定すべきパラメータをまとめて

$$\boldsymbol{\theta} = (\boldsymbol{\pi}, \boldsymbol{A}, \boldsymbol{G}, \boldsymbol{\mu}, \boldsymbol{\sigma}) \tag{2.8}$$

とする.

HMM と同様に,EM アルゴリズムに基づいて Q 関数を定義し,最大化することによりパラメータを推定する.

$$Q(\boldsymbol{\theta}^0, \boldsymbol{\theta}) = \sum_{\mathbf{s}}\sum_{\mathbf{m}} P(\mathbf{s},\mathbf{m}|\mathbf{x};\boldsymbol{\theta}^0)\log P(\mathbf{x},\mathbf{s},\mathbf{m};\boldsymbol{\theta}) \tag{2.9}$$

この後の式展開など詳しい情報は,参考文献 [5],参考文献 [6]p71-p73 を参照.

3　Factorial Hidden Markov Models

この章では Factorial Hidden Markov Models について説明する．EM アルゴリズムや変分推論 [7] などを用いたパラメータ推定など，更に詳細な内容については参考文献 [8] を参照されたい．

3.1　モデル概要

Factorial Hidden Markokv Models（FHMM）[8] は 1997 年に提案された，複数の隠れ変数のマルコフ連鎖に対応した HMM であり，音声解析や自然言語処理，モーションキャプチャによる運動データの復元など提案されてから今日まで多岐にわたり利用されている [9–12]．FHMM では，互いに独立な複数の隠れ変数のマルコフ連鎖があり，ある時刻 t における観測変数は，時刻 t におけるすべての隠れ変数の状態に依存する．図 2 に対応するグラフィカルモデルを示す．このグラフィカルモデルでは，時刻 t における m 個目のマルコフ連鎖の隠れ変数を $S_t^{(m)}$，観測変数を x_t としている．

3.2　確率モデル

まず，通常の HMM の場合を考える．状態系列

$$\mathbf{s} = s_1 s_2 \cdots s_t \cdots s_T \tag{3.1}$$

と観測系列

$$\mathbf{x} = x_1 x_2 \cdots x_t \cdots x_T \tag{3.2}$$

を考える．このとき，式 (**??**) は，

$$P(\mathbf{x}, \mathbf{s}) = P(s_1)P(x_1|s_1)\prod_{t=2}^{T} P(s_t|s_{t-1})P(x_t|s_t) \tag{3.3}$$

と書ける．ただし，状態数を K としたとき，

$$s_t \in \{1, ..., K\} \tag{3.4}$$

である．このとき，HMM の遷移確率行列 $P(s_t|s_{t-1})$ は $K \times K$ になる．また，出力記号の数を D としたとき，HMM の出力確率行列 $P(x_t|s_t)$ は $K \times D$ になる．出力確率行列 $P(x_t|s_t)$ が連続値をとる場合，正規分布（ガウス分布）や混合正規分布，またはニューラルネットワークのような多くの異なる形でモデル化することができる．

次に，FHMM の場合を考える．FHMM では複数の隠れ変数のマルコフ連鎖があるので，時点 t における隠れ変数は，

$$s_t = s_t^{(1)}, s_t^{(2)}, \cdots, s_t^{(m)}, \cdots, s_t^{(M)} \tag{3.5}$$

と表せる．ここで，隠れ変数のマルコフ連鎖の数を M とした．各状態変数 $s_t^{(m)}$ は，$K^{(M)}$ 個の状態をとる．しかし本誌では簡略化のため，すべての m に対して $K^{(M)} = K$ とする．このとき，遷移確率行列 $P(_t|s_{t-1})$ を次のように表す．

$$P(s_t|s_{t-1}) = \prod_{m=1}^{M} P(s_t^{(m)}|s_{t-1}^{(m)}) \tag{3.6}$$

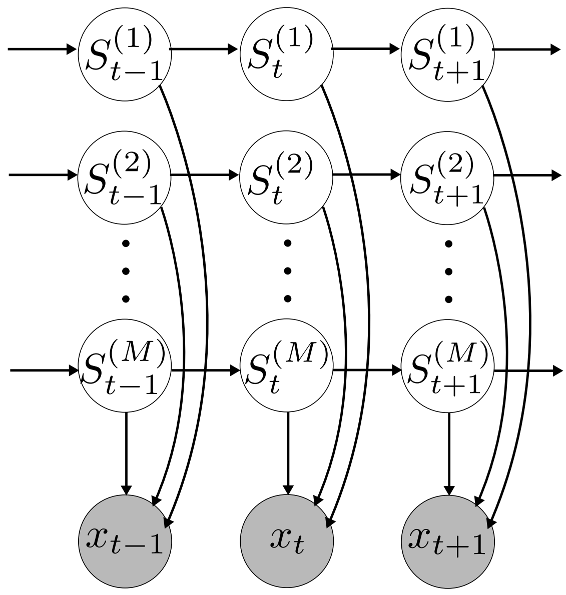


図 2: FHMM のグラフィカルモデル

図 2 を見ると，時点 t における観測値 x_t は時点 t におけるすべての隠れ変数に依存していることがわかる．連続値をとる観測値のために，これらの依存を解決する手法の 1 つとして線形ガウスモデル（Liner Gaussian Model）がある．これは，観測値 x_t が正規乱数ベクトルになっており，その平均が状態変数の線形関数になっているものである．ここで，状態変数を 1-of-K 表現で表すこととする．1-of-K 表現とは，K 番目の要素が 1 でそれ以外の要素はすべて 0 であるベクトル表現である．

$D \times 1$ の観測値ベクトルの確率分布は，

$$P(x_t|s_t) = |C|^{-\frac{1}{2}}(2\pi)^{-\frac{D}{2}} \exp\left\{-\frac{1}{2}(x_t-\mu_t)^\top C^{-1}(x_t-\mu_t)\right\} \tag{3.7}$$

で表される．ここで，

$$\mu_t = \sum_{m=1}^{M} W^{(m)} s_t^{(m)} \tag{3.8}$$

とする．各行列 $W^{(m)}$ は $D \times K$ の行列であり，その列は各 $s_t^{(m)}$ を決める平均に寄与するものであり，C は $D \times D$ の分散・共分散行列，$^\top$ は転置，そして $|\cdot|$ は行列式を表す．

参考文献

[1] 平岡和幸堀玄. プログラミングのための確率統計. オーム社, 2009.

[2] 石井健一郎上田修功. 続・わかりやすいパターン認識: 教師なし学習入門. オーム社, 2014.

[3] Stephen E Levinson, Lawrence R Rabiner, and Man Mohan Sondhi. An introduction to the application of the theory of probabilistic functions of a markov process to automatic speech recognition. *Bell System Technical Journal, The*, Vol. 62, No. 4, pp. 1035–1074, 1983.

[4] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[5] Javier R. Movellan. Tutorial on hidden markov models.

[6] 確率モデルによる音声認識. 確率モデルによる音声認識. 電子情報通信学会 (コロナ社), 1988.

[7] Michael I. Jordan, Zoubin Ghahramani, Tommi S. Jaakkola, and Lawrence K. Saul. An introduction to variational methods for graphical models. *Mach. Learn.*, Vol. 37, No. 2, pp. 183–233, November 1999.

[8] Zoubin Ghahramani and Michael I. Jordan. Factorial hidden markov models. *Mach. Learn.*, Vol. 29, No. 2-3, pp. 245–273, November 1997.

[9] Jean-Louis Durrieu. and Jean-Philippe Thiran. Source/filter factorial hidden markov model, with application to pitch and formant tracking. Vol. 21, pp. 2541–2553, August 2013.

[10] Beth Logan and Pedro J. Moreno. Factorial hidden markov models for speech recognition: Preliminary experiments, 1997.

[11] Dana Kulic Dongheui Lee and Yoshihiko Nakamura. Missing motion data recovery using factorial hidden markov models. Robotics and Automation, 2008. ICRA 2008. IEEE International Conference., May 2008.

[12] Anjan Nepal and Alexander Yates. Factorial hidden markov models for learning representations of natural language. *CoRR*, 2013.