EM アルゴリズムのパラメータ推定についての補足

佐々木 稔

平成 15 年 8 月 1 日

確率競合モデルにおける EM アルゴリズムで，関数 $G_n(\underline{w}, w)$ を最大化するパラメータ w を直接計算できる．ここでは，そのパラメータの計算式を導出する．

パラメータとして $w = (a_h, b_h)$ となる確率競合モデルを考える．ここで，b_h は平均 ξ_h，分散 $(\sigma_h)^2$ の正規分布をもつパラメータ $b_h = (\xi_h, \sigma_h)$ とする．このとき，この競合モデルは以下の式で表される．

$$p(x_i, u|w) = \prod_{h=1}^{H} \{a_h q(x_i|b_h)\}^{u_h} \tag{1}$$

両辺の対数をとると，以下のようになる．

$$\log p(x_i, u|w) = \sum_{h=1}^{H} u_h \{\log a_h + \log q(x_i|b_h)\} \tag{2}$$

x_i と $\underline{w}$ が与えられたときの u_h の平均 $E_i(h)$ は，

$$\begin{aligned} E_i(h) &= E_i\{u_h|x_i, \underline{w}\} = \sum_{u_h=0,1} u_h p(u_h|x_i, \underline{w}) \\ &= \sum_{u_h=0,1} u_h \frac{p(u_h, x_i|\underline{w})}{p(x_i|\underline{w})} \\ &= \sum_{u_h=0,1} u_h \frac{\prod_{h=1}^{H}\{\underline{a_h} q(x_i|\underline{b_h})\}^{u_h}}{p(x_i|\underline{w})} \\ &= \frac{\underline{a_h} q(x_i|\underline{b_h})}{p(x_i|\underline{w})} \end{aligned} \tag{3}$$

である．これより，パラメータを $\underline{w}$ に固定したときの，x_i における $\log p(x_i, u|w)$ の値は，

$$\sum_{h=1}^{H} E_i(h)\{\log a_h + \log q(x_i|b_h)\} \tag{4}$$

となるので，すべての観測データ $x_i (i = 1, 2, \cdots, n)$ についての和を求め，それを最適化させる関数 $G_n(\underline{w}, w)$ とする．

$$G_n(\underline{w}, w) = \sum_{i=1}^{n} \sum_{h=1}^{H} E_i(h)\{\log a_h + \log q(x_i|b_h)\} \tag{5}$$

この関数 $G_n(\underline{w}, w)$ を，条件 $\sum_h a_h = 1, a_h > 0$ のもとで最大化を行う．

最大化を行うためには，ラグランジュの未定係数法を用いる．条件 $\sum_h a_h = 1, a_h > 0$ に対するラグランジュ係数を γ として，関数 $\mathcal{L}$ を考える．

$$\mathcal{L}(w, \gamma) = G_n(\underline{w}, w) - \gamma\{\sum_h a_h - 1\} \tag{6}$$

まず，a_h を求めるために，関数 $\mathcal{L}$ を a_h で偏微分して，その値を 0 とする．

$$\frac{\partial \mathcal{L}}{\partial a_h} = \sum_{i=1}^{n} E_i(h) \frac{1}{a_h} - \gamma = 0 \tag{7}$$

これを，a_h について解くと

$$a_h = \frac{1}{\gamma} \sum_{i=1}^{n} E_i(h) \tag{8}$$

このとき，

$$\begin{aligned} \sum_{h=1}^{H} a_h &= \frac{1}{\gamma} \sum_{h=1}^{H} \sum_{i=1}^{n} E_i(h) \\ &= \frac{1}{\gamma} \sum_{i=1}^{n} (\sum_{h=1}^{H} E_i(h)) = \frac{n}{\gamma} = 1 \end{aligned} \tag{9}$$

より，$\gamma = n$ となるので，求める a_h の値は

$$a_h = \frac{1}{n} \sum_{i=1}^{n} E_i(h) \tag{10}$$

となる．

次に，ξ_h を求めるために，$q(x_i|b_h)$ を以下の式で表現する．

$$q(x_i|b_h) = \frac{1}{(2\pi\sigma_h^2)^{M/2}} \exp(-\frac{\|x_i - \xi_h\|^2}{2\sigma_h^2}) \tag{11}$$

これを，$G_n(\underline{w}, w)$ に代入すると，以下の式が得られる．

$$\begin{aligned} G_n(\underline{w}, w) &= \sum_{i=1}^{n} \sum_{h=1}^{H} E_i(h)\{\log a_h - \log(2\pi\sigma_h^2)^{M/2} \\ &- \frac{\|x_i - \xi_h\|^2}{2\sigma_h^2}\} \end{aligned} \tag{12}$$

このとき，ξ_h に関して $G_n(\underline{w}, w)$ を最大化するためには，次式の $\tilde{G}$ を最大化すればよい．

$$\tilde{G} = -\frac{1}{2\sigma_h^2} \sum_{i=1}^{n} \sum_{h=1}^{H} E_i(h) \|x_i - \xi_h\|^2 \tag{13}$$

この $\tilde{G}$ を ξ_h で偏微分し，その値を 0 とする．

$$\frac{\partial \tilde{G}}{\partial \xi_h} = -\frac{1}{2\sigma_h^2}\sum_{i=1}^{n} E_i(h) \times (-2)(x_i - \xi_h) = 0 \tag{14}$$

これより，求める ξ_h の値は

$$\xi_h = \frac{\sum_{i=1}^{n} E_i(h) x_i}{\sum_{i=1}^{n} E_i(h)} \tag{15}$$

となる．

また，$(\sigma_h)^2$ を求めるために，式 (12) にある $(\sigma_h)^2$ を残したものを G^+ とすると，

$$\begin{aligned} G^+ &= \sum_{i=1}^{n}\sum_{h=1}^{H} E_i(h)\{\frac{M}{2}\log\frac{1}{(\sigma_h)^2} \\ &- \frac{1}{2\{(\sigma_h)^2\}^2}\|x_i - \xi_h\|^2\} \end{aligned} \tag{16}$$

この G^+ を $(\sigma_h)^2$ で偏微分し，その値を 0 とする．

$$\begin{aligned} \frac{\partial G^+}{\partial (\sigma_h)^2} &= \sum_{i=1}^{n} E_i(h)\{\frac{M}{2}(-\frac{1}{\{(\sigma_h)^2\}^2})(\sigma_h)^2 \\ &+ \frac{1}{2\{(\sigma_h)^2\}^2}\|x_i - \xi_h\|^2\} \\ &= \frac{1}{2\{(\sigma_h)^2\}^2}\sum_{i=1}^{n} E_i(h)\{-M(\sigma_h)^2 \\ &+ \|x_i - \xi_h\|^2\} = 0 \end{aligned} \tag{17}$$

これより，求める $(\sigma_h)^2$ の値は

$$(\sigma_h)^2 = \frac{\sum_{i=1}^{n} E_i(h)\|x_i - \xi_h\|^2}{M\sum_{i=1}^{n} E_i(h)} \tag{18}$$

となる．