EM ALGORITHMS OF GAUSSIAN MIXTURE MODEL AND HIDDEN MARKOV MODEL

Guorong Xuan Wei Zhang Peiqi Chai
Department of Computer Science, Tongji University
Shanghai 200092, P. R. of China
E-mail: grxuan@public1.sta.net.cn

ABSTRACT

The HMM (Hidden Markov Model) is a probabilistic model of the joint probability of a collection of random variables with both observations and states. The GMM (Gaussian Mixture Model) is a finite mixture probability distribution model. Although the two models have a close relationship, they are always discussed independently and separately. The EM (Expectation-Maximum) algorithm is a general method to improve the descent algorithm for finding the Maximum Likelihood Estimation. The EM of HMM and the EM of GMM have similar formula. Two points are proposed in this paper. One is that the EM of GMM can be regarded as a special EM of HMM. The other is that the EM algorithm of GMM based on symbol is faster in implementation than EM algorithm of GMM based on sample (or on observation) traditionally.

KEYWORDS:

Expectation-Maximum (EM), Hidden Markov Model (HMM), Gaussian Mixture Model (GMM), Maximum Likelihood Estimation (MLE), GMM based on sample, GMM based on symbol.

1. INTRODUCTION

The HMM (Hidden Markov Model) is a probabilistic model of the joint probability of a collection of random variables with both observations and states. It can be seen as a stationary Markov distribution. All probability distributions are assumed with Gaussian in this paper. The GMM (Gaussian Mixture Model) is a finite mixture probability distribution model. The EM (Expectation-Maximum) algorithm is a general method of finding the Maximum Likelihood Estimation (MLE) from an incomplete data and is actually an improved descent algorithm. The EM algorithm for MLE under both GMM and HMM is usually used [1] [2]. Although the two models have a close relationship, they are always discussed independently and separately.

The EM of HMM and the EM of GMM have similar formula. The EM of GMM can be regarded as a special EM of HMM in this paper. In addition, the EM algorithm of GMM based on symbol is proposed, which is faster in implementation than EM algorithm of GMM based on sample traditionally.

The following notations are used for next discussion.

T: number of samples $1 \le t \le T$
N: number of states $1 \le j,k \le N$
M: number of mixture components $1 \le m \le M$
L: number of symbols $1 \le l \le L$
O: observation or Sample $O = (o_1 \cdots o_t \cdots o_T)$
q : state $q = (q_1, \cdots q_t \cdots, q_T)$, $q_t \in \{1, \cdots, N\}$
V: symbol $V = (v_1 \cdots v_l \cdots v_L)$, $o_t \in \{v_1 \cdots v_l \cdots v_L\}$
A: state transition matrix, $A = \{a_{jk}\}$, $1 \le j,k \le N$
λ: parameter
λ_{ML}: parameter for Maximum Likelihood Estimation
$L(O|\lambda)$: Likelihood Function
$p(O|\lambda)$: joint probability of observations
$p(O,q|\lambda)$: joint probability of both observations & states
$Q(\lambda, \lambda^{(i-1)})$:EM Auxiliary Function
where superscript for iteration (i-1)
c_{jm}: mixture component (j state and m component)
$\mathrm{N}(o_t \mid \mu_{jm}, \Sigma_{jm})$: Gaussian Density Distribution
μ_{jm}: mean vector, Σ_{jm}: covariance matrix
γ_{jtm}: posterior probability
(t :sample, j :state and m: component)
α_{jtm}:forward partial observation probability in HMM
β_{jtm}: back partial observation probability in HMM
H: histogram of symbols in GMM $H = \{h_1 \cdots h_l \cdots h_L\}$
$P(V \mid \lambda)$: distribution by MLE based on symbol for GMM
w_{ml}: weight coefficient for GMM based on symbol
(m: component, l: symbol)

2. EM ALGORITHM OF HMM

Maximum Likelihood Estimation (MLE) of HMM using Likelihood Function directly is shown in Table 1

Table 1 MLE using Likelihood Function directly

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

$$\lambda_{ML}=\arg\max_{\lambda}\{L(O|\lambda)+constra\text{int}\}$$

where Constraint for HMM:

$$\sum_{m=1}^{M}c_{jm}=1 \quad 1\le j\le N,$$

$$\sum_{i=1}^{N}\pi_i=1, \qquad \sum_{j=1}^{N}a_{jk}=1 \quad 1\le k\le N$$

The MLE of HMM using EM Auxiliary Function [3] is shown in Table 2.

Table 2 MLE using EM Auxiliary Function

$E\ step:$

$$Q\left(\lambda,\lambda^{(i-1)}\right)=\sum_{q}P\left(O,q\mid\lambda^{(i-1)}\right)\log P(O,q\mid\lambda)$$

$$M\ step:\quad \lambda^{(i)}=\underset{\lambda}{\text{argmax}}\left[Q\left(\lambda,\lambda^{(i-1)}\right)+\text{constraint}\right]$$

$$\text{Final Result:}\quad \text{when}\quad i\to\infty,\quad \lambda^{(i)}\to\lambda_{ML}$$

The result formulas of MLE by using EM parameter estimation under Gaussian distribution of HMM based on sample is shown in Table 3.

Table 3 EM formula of HMM based on sample

$$c_{jm}^{(i)}=\frac{\sum_{t=1}^{T}\gamma_{jtm}^{(i)}}{\sum_{t=1}^{T}\sum_{m=1}^{M}\gamma_{jtm}^{(i)}},\qquad \mu_{jm}^{(i)}=\frac{\sum_{t=1}^{T}\gamma_{jtm}^{(i)}\cdot o_t}{\sum_{t=1}^{T}\gamma_{jtm}^{(i)}}$$

$$\Sigma_{jm}^{(i)}=\frac{\sum_{t=1}^{T}\gamma_{jtm}^{(i)}\cdot\left(o_t-\mu_{jm}^{(i)}\right)\left(o_t-\mu_{jm}^{(i)}\right)^t}{\sum_{t=1}^{T}\gamma_{jtm}^{(i)}}$$

$$\gamma_{jtm}^{(i)}=\left(\frac{\alpha_{jtm}^{(i-1)}\beta_{jtm}^{(i-1)}}{\sum_{j=1}^{N}\alpha_{jtm}^{(i-1)}\beta_{jtm}^{(i-1)}}\right)\cdot\left(\frac{c_{jm}^{(i-1)}\mathrm{N}\left(o_t\mid\mu_{jm}^{(i-1)},\Sigma_{jm}^{(i-1)}\right)}{\sum_{m=1}^{M}c_{jm}^{(i-1)}\mathrm{N}\left(o_t\mid\mu_{jm}^{(i-1)},\Sigma_{jm}^{(i-1)}\right)}\right)$$

$$\alpha_{jtm}^{(i-1)}=p\left(o_1o_2\cdots o_t,q_t=j\mid\mu_{jm}^{(i-1)},\Sigma_{jm}^{(i-1)}\right)$$

$$\beta_{jtm}^{(i-1)}=p\left(o_{t+1}o_{t+2}\cdots o_T,q_t=j\mid\mu_{jm}^{(i-1)},\Sigma_{jm}^{(i-1)}\right)$$

3. EM ALGORITHM OF GMM

3.1 EM Algorithm of GMM-1

In EM of GMM-1, all EM formulas of HMM can be used except number of states N=1 shown in Table 4. It means that the model is a strictly stationary model. The EM results of Table 4 can also be obtained by using Likelihood Function directly [4]. However, the EM is still a more elaborate method.

Table 4 EM formula of GMM-1based on sample

$$c_{m}^{(i)}=\frac{\sum_{t=1}^{T}\gamma_{tm}^{(i)}}{\sum_{t=1}^{T}\sum_{m=1}^{M}\gamma_{tm}^{(i)}},\qquad \mu_{m}^{(i)}=\frac{\sum_{t=1}^{T}\gamma_{tm}^{(i)}\cdot o_t}{\sum_{t=1}^{T}\gamma_{tm}^{(i)}}$$

$$\Sigma_{m}^{(i)}=\frac{\sum_{t=1}^{T}\gamma_{tm}^{(i)}\cdot\left(o_t-\mu_{m}^{(i)}\right)\left(o_t-\mu_{m}^{(i)}\right)^t}{\sum_{t=1}^{T}\gamma_{tm}^{(i)}}$$

$$\gamma_{jtm}^{(i)}=\frac{\alpha_{jm}^{(i-1)}\beta_{jm}^{(i-1)}\gamma_{tm}^{(i)}}{\sum_{j=1}^{M}\alpha_{jm}^{(i-1)}\beta_{jm}^{(i-1)}},\ \gamma_{tm}^{(i)}=\left(\frac{c_{m}^{(i-1)}\mathrm{N}\left(o_t\mid\mu_{m}^{(i-1)},\Sigma_{m}^{(i-1)}\right)}{\sum_{m=1}^{M}c_{m}^{(i-1)}\mathrm{N}\left(o_t\mid\mu_{m}^{(i-1)},\Sigma_{m}^{(i-1)}\right)}\right)$$

m : sequence number of component

3.2 EM Algorithm of GMM-2

In EM of GMM-2, all EM formulas of HMM can be used except that both number of states is equal to number of categories (N=M) and state transition matrix is with the same values in any column. The GMM-2 is shown in Table 5. It means the model is a zero-order Markov model or named Independent Model (IM).

Table 5 EM formula of GMM-2 based on sample

$$c_{j}^{(i)}=\frac{\sum_{t=1}^{T}\gamma_{tj}^{(i)}}{\sum_{t=1}^{T}\sum_{j=1}^{M}\gamma_{tj}^{(i)}},\qquad \mu_{j}^{(i)}=\frac{\sum_{t=1}^{T}\gamma_{tj}^{(i)}\cdot o_t}{\sum_{t=1}^{T}\gamma_{tj}^{(i)}}$$

$$\Sigma_{j}^{(i)}=\frac{\sum_{t=1}^{T}\gamma_{tj}^{(i)}\cdot\left(o_t-\mu_{j}^{(i)}\right)\left(o_t-\mu_{j}^{(i)}\right)^t}{\sum_{t=1}^{T}\gamma_{tj}^{(i)}}$$

$$\gamma_{jtm}^{(i)}=\frac{\alpha_{jm}^{(i-1)}\beta_{jm}^{(i-1)}\gamma_{tj}^{(i)}}{\sum_{j=1}^{N}\alpha_{jm}^{(i-1)}\beta_{jm}^{(i-1)}},\ \gamma_{tj}^{(i)}=\left(\frac{c_{j}^{(i-1)}\mathrm{N}\left(o_t\mid\mu_{j}^{(i-1)},\Sigma_{j}^{(i-1)}\right)}{\sum_{j=1}^{M}c_{j}^{(i-1)}\mathrm{N}\left(o_t\mid\mu_{j}^{(i-1)},\Sigma_{j}^{(i-1)}\right)}\right)$$

j : sequence number of state

3.3 Comparisons

The final formulas of GMM-1 & GMM-2 are almost the same (Table 4, 5). The Comparisons are shown in Table 6. The parameter estimation of discrete probability density instead of distribution estimation of discrete probability density (Baum-Welch algorithm) is used in this paper.

Table 6 Comparisons for HMM with GMM-1 & GMM-2

	HMM	**GMM-1**	**GMM-2**
Random Process	First-order Markov	Stationary	Zero-order Markov
Hidden Parameter	states	Categories (only one state)	states (one category for each state)
Number of both States and mixture Components	N states, M Gaussian distribu-tion for each state	One state, M Gaussian distribution for this state	N states, one Gaussian distribution for each state
Fast Algorithm by histogram	No	Yes (see next)	Yes

4. EM algorithm of GMM based on symbol

The speed can be improved for GMM by the histogram because that the number of symbols is less than the number of samples (L<T). The EM formula of GMM-1 based on symbol is shown in Table 7.

Table 7 EM formula of GMM-1 based on symbol

$$c_m^{(i)} = \frac{\sum_{l=1}^{L} w_{ml}^{(i)}}{\sum_{l=1}^{L}\sum_{m=1}^{M} w_{ml}^{(i)}}, \qquad \mu_m^{(i)} = \frac{\sum_{l=1}^{L} w_{ml}^{(i)} \cdot v_l}{\sum_{l=1}^{L} w_{ml}^{(i)}}$$

$$\Sigma_m^{(i)} = \frac{\sum_{l=1}^{L} w_{ml}^{(i)} \cdot \left(v_l - \mu_m^{(i)}\right)\left(v_l - \mu_m^{(i)}\right)^t}{\sum_{l=1}^{L} w_{ml}^{(i)}}$$

$$w_{ml}^{(i)} = \left[\frac{h_l}{T}\right] \cdot \left(\frac{c_m^{(i-1)} \mathrm{N}\left(v_l \mid \mu_m^{(i-1)}, \Sigma_m^{(i-1)}\right)}{\sum_{m=1}^{M} c_m^{(i-1)} \mathrm{N}\left(v_l \mid \mu_m^{(i-1)}, \Sigma_m^{(i-1)}\right)} \right)$$

The conditional probability:

$$\mathrm{N}\left(v_l \mid \mu_m^{(i-1)}, \Sigma_m^{(i-1)}\right) = \xi_m^{(i-1)} \cdot \exp\left[-\frac{1}{2} D_{ml}^{(i-1)} - \frac{1}{2}\log\left|\Sigma_m^{(i-1)}\right|\right]$$

where

$$\xi_m^{(i-1)} = 1 \Big/ \left(\sum_{l=1}^{L} \exp\left[-\frac{1}{2} D_{ml}^{(i-1)} - \frac{1}{2}\log\left|\Sigma_m^{(i-1)}\right|\right] \right)$$

$$D_{ml}^{(i-1)} = \sum_{l=1}^{L} \left[\left(v_l - \mu_m^{(i-1)}\right)^t \left(\Sigma_m^{(i-1)}\right)^{-1} \left(v_l - \mu_m^{(i-1)}\right) \right]$$

The MLE based on symbol is faster than MLE based on observation in the implementation, because we calculate the same symbols only once (with its histogram).

5. EXAMPLE OF EM BASED ON SYMBOL

A GMM example of EM algorithm based on symbol of image "Lena" with 3D symbol V (RGB) is given. The number of samples (observation O) is $T = 512 \times 512 = 2^{18}$ and the number of symbols is $L < 32 \times 32 \times 32 = 2^{15}$ (quantized from 256 to 32 gray level, L is the number of symbols with non-zero values in histogram). The symbol set is sparse, and only $L = 3657 < 2^{12}$ symbols are left. The histogram H and the distribution $P(V \mid \lambda)$ by MLE for GMM are shown by projections in Figure 1-6.

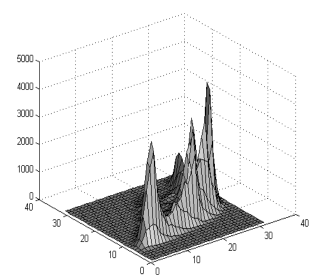


Figure 1 projection 1 of histogram H

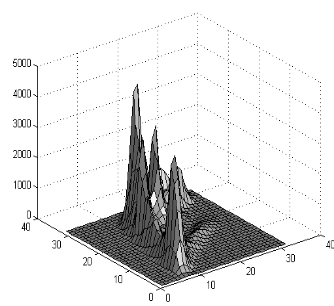


Figure 2 projection 2 of histogram H

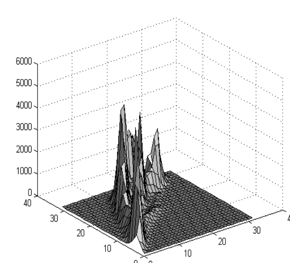


Figure 3 projection 3 of histogram H

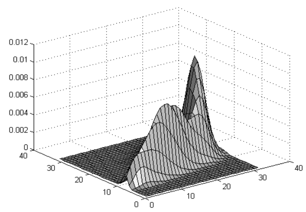


Figure 4 projection 1 of $P(V \mid \lambda)$ by MLE

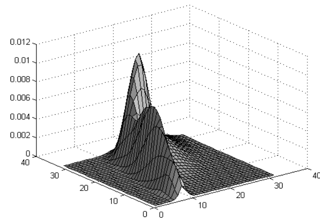


Figure 5 projection 2 of $P(V \mid \lambda)$ by MLE

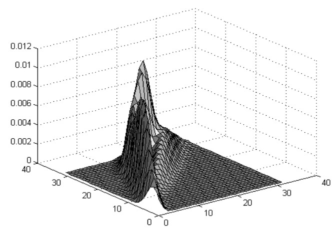


Figure 6 projection 3 of $P(V \mid \lambda)$ by MLE

The results of parameters by MLE of GMM based on symbol are listed as Table 8.

Table 8 The parameters by MLE of GMM

Priori probabilities (Mixture Components)

$$c_1 = 0.3896$$
$$c_2 = 0.2629$$
$$c_2 = 0.3475$$

Mean vectors:

$$\mu_1 = [13.9415 \quad 6.5279 \quad 10.2281]^t$$
$$\mu_2 = [22.9980 \quad 18.8678 \quad 20.3639]^t$$
$$\mu_3 = [26.0142 \quad 15.8115 \quad 15.7451]^t$$

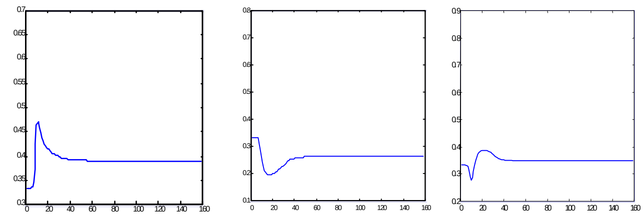
Covariance matrices:

$$\Sigma_1 = \begin{bmatrix} 46.6749 & 21.8882 & 8.8357 \\ 21.8882 & 11.2503 & 5.0199 \\ 8.8357 & 5.0199 & 3.1143 \end{bmatrix}$$

$$\Sigma_2 = \begin{bmatrix} 44.3546 & 43.0801 & 20.0916 \\ 43.0801 & 45.9463 & 25.1790 \\ 20.0916 & 25.1790 & 20.7511 \end{bmatrix}$$

$$\Sigma_3 = \begin{bmatrix} 5.1412 & 4.2196 & 1.4446 \\ 4.2196 & 8.4047 & 5.1727 \\ 1.4446 & 5.1727 & 5.4384 \end{bmatrix}$$

Each priori probability (Mixture Component) with number of iteration (i) is shown in Figure 7. After about 50 iterations (150 iterations listed in Figure 7), the parameters are convergent with proper initial value. Since L<<T, the speed of implementation on symbol is faster. The speed on symbol is T/L times faster than that on sample, that is $T/L = (512 \times 512)/3647 \approx 72$ in this example.



$c_1^{(i)}$ $c_2^{(i)}$ $c_3^{(i)}$

Figure 7 priori probability in procedure of iteration

6. ACKNOWLEDGEMENT

This study is funded by National Fundamental Research Programs (973-G1998030419). The authors wish to thank the Committee of NFRP of China.

7. CONCLUSSIONS

1) The EM of GMM (GMM-1 or GMM-2) can be regarded as a special EM of HMM. HMM is a first-order Markov model. The final formulas of GMM-1 & GMM-2 are almost the same.
2) GMM-1 is a strict stationary model. All the formulas of EM of HMM can be used in EM of GMM-1 except number of state N=1.
3) GMM-2 is a zero-order Markov model or named Independent Model (IM). All the formulas of EM of HMM can be used in EM of GMM-2 except that both N=M and state transition matrix is with the same values in any column.
4) EM algorithm based on symbol is faster than EM algorithm based on sample in GMM implementation. An example of EM algorithm based on symbol is given in this paper.

8. REFERENCES

[1] L. Rabiner and B. H. Juang, "*Fundamentals of Speech Recognition*", Prentice-Hall PTR, 1993.

[2] J. A. Bilmes ,"A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixed and Hidden Markov Models", *U.C. Berkely, TR-97-021*, April 1998.

[3] A. P. Dempster, N. M Land, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm", *Journal of the Royal Statistical Society*, Series B, Vol. 39, No. 1, 1977 , pp. 1-38

[4] R. O. Duda, P. E. Hart and D. G. Stork, "*Pattern Classification*", John Wiley & Sons, INC. 2001.