

変分ベイズ法を用いた分離型2次元格子HMM の学習におけるアニーリング制御の適用

沢田 慶[†] 玉森 聡[†] 橋本 佳[†] 南角 吉彦[†] 徳田 恵一[†]
[†]名古屋工業大学 大学院工学研究科 〒466-8555 名古屋市昭和区御器所町

1 はじめに

画像認識の一手法として、画像の位置や大きさの正規化処理が識別器に組み込まれた分離型2次元格子HMM (separable lattice 2-D hidden Markov model; SL2D-HMM) に基づく手法が提案されている [1]。従来 SL2D-HMM の学習には尤度最大化 (maximum likelihood; ML) 基準が用いられてきたが、近年変分ベイズ (variational Bayesian; VB) 法を用いたベイズ基準が適用され、学習データが少量の場合においても高い汎化能力が確認された [2]。VB 法は、事後分布の効率的な計算を行うことを可能とするが、反復計算による逐次推定を行うため局所最適性の問題が残されていた。そこで、本稿では VB 法を用いた SL2D-HMM の学習に確定的アニーリング EM (deterministic annealing expectation maximization; DAEM) アルゴリズムを適用し、局所最適性の問題の改善を目指す。

2 SL2D-HMM

SL2D-HMM は、HMM を2次元に拡張したモデルであり、横、縦方向の状態系列を持つ。横、縦方向の状態遷移はそれぞれ独立であり、モデルの出力は各方向の状態の組み合わせによって決定される。SL2D-HMM は、矩形を維持しながら横、縦方向の伸縮を表現可能であり、画像の位置や大きさの変動に対応することができる。SL2D-HMM の尤度関数 $P(O, S | \Lambda)$ は次式で定義される。

$$P(O, S | \Lambda) = P(O | S^{(1)}, S^{(2)}, \Lambda) \prod_{m=1}^2 P(S^{(m)} | \Lambda) \quad (1)$$

ここで、 O は2次元静止画像、 $S = \{S^{(1)}, S^{(2)}\}$ は横、縦方向の状態を表す隠れ変数、 Λ はモデルパラメータであり、初期状態確率、状態遷移確率、出力確率分布から構成される。

3 VB 法を用いた SL2D-HMM の学習

ベイズ基準による未知観測 X の識別には、次式で定義される予測分布 $P(X | O)$ を用いる。

$$P(X | O) = \int P(X | \Lambda) P(\Lambda | O) d\Lambda \quad (2)$$

ここで、 $P(\Lambda | O)$ は観測 O が与えられた下でのモデルパラメータ Λ の事後分布である。事後分布は事前分布 $P(\Lambda)$ と尤度関数 $P(O | \Lambda)$ を用いてベイズの定理から次式のように表される。

$$P(\Lambda | O) = \frac{P(O | \Lambda)P(\Lambda)}{P(O)} \quad (3)$$

しかし、隠れ変数を含むモデルにおいては、特殊な場合を除いて予測分布を解析的に求めることは困難であり何らかの近

似が必要となる。この問題に対し、近年事後分布の近似推定手法として VB 法が提案されている [3]。

VB 法は、全ての未知量を周辺化した対数周辺尤度を考える。そして、対数周辺尤度に対し任意の分布 $Q(S, \Lambda)$ を導入し、Jensen の不等式を用いることにより対数周辺尤度の下限 \mathcal{F} を定義する。

$$\begin{aligned} \ln P(O) &= \ln \sum_S \int P(O, S, \Lambda) d\Lambda \\ &\geq \sum_S \int Q(S, \Lambda) \ln \frac{P(O, S, \Lambda)}{Q(S, \Lambda)} d\Lambda \\ &= \mathcal{F}(Q) \end{aligned} \quad (4)$$

下限 \mathcal{F} は、 $Q(S, \Lambda)$ を変関数とする汎関数であり、 $Q(S, \Lambda)$ に関して下限 \mathcal{F} を最大化することにより、 $Q(S, \Lambda)$ は事後分布 $P(S, \Lambda | O)$ の近似分布となる。ここで、近似事後分布に対して制約条件 $Q(S, \Lambda) = Q(S^{(1)})Q(S^{(2)})Q(\Lambda)$ を与えることにより、実現可能な計算量で近似事後分布を推定することができる。

4 DAEM アルゴリズムの VB 法への適用

VB 法は、近似事後分布の効率的な計算を行うことを可能とするが、EM アルゴリズムと同様の反復計算による逐次推定を行うため局所最適性の問題が残されている。このような局所最適性の問題を改善する手法として DAEM アルゴリズムが提案されている [4]。本稿では、DAEM アルゴリズムを VB 法に適用するにあたり、3つの温度パラメータ $\beta_1, \beta_2, \beta_3$ を導入した自由エネルギー関数を定義する。

$$f(O, S, \Lambda) = P(O | S, \Lambda)^{\beta_1} P(S | \Lambda)^{\beta_2} P(\Lambda)^{\beta_3} \quad (5)$$

ここで、 β_1 は出力確率分布に対する温度パラメータ、 β_2 は横、縦方向の初期状態確率と状態遷移確率に対する温度パラメータ、 β_3 は事前分布に対する温度パラメータである。3つの温度パラメータを導入することで、自由エネルギー関数の下限 $\mathcal{F}_{\beta_1, \beta_2, \beta_3}$ は次式のように定義される。

$$\begin{aligned} \mathcal{F}_{\beta_1, \beta_2, \beta_3}(Q) &= \sum_S \int Q(S, \Lambda) \\ &\times \ln \frac{P(O | S, \Lambda)^{\beta_1} P(S | \Lambda)^{\beta_2} P(\Lambda)^{\beta_3}}{Q(S, \Lambda)} d\Lambda \end{aligned} \quad (6)$$

温度パラメータが $\beta_k \simeq 0$ ($k = 1, 2, 3$) であるとき、下限 $\mathcal{F}_{\beta_1, \beta_2, \beta_3}$ は唯一の極大値を持つ。そのため、初期値に依存することなく下限 $\mathcal{F}_{\beta_1, \beta_2, \beta_3}$ の最適値を推定することができる。温度パラメータ β_k の値を増加させると下限 $\mathcal{F}_{\beta_1, \beta_2, \beta_3}$ には最適値以外の極大値が現われる。ここで、温度パラメータ β_k をゆるやかに変化させることで最適値である極大値を推定しやすくなる。温度パラメータが $\beta_k = 1$ であるとき、下限 $\mathcal{F}_{\beta_1, \beta_2, \beta_3}$ は下限 \mathcal{F} と一致するが、温度パラメータをゆるやかに変化させることで初期値に依存しにくい推定を行うことができる。また、式 (5) で定義したように、周辺尤度の各要素に個別の温度パラメータを設定することにより、モデル構造に適したアニーリング制御を行うことができる。

A training algorithm based on variational Bayesian method using deterministic annealing process for separable lattice 2-D HMMs

Kei Sawada[†] Akira Tamamori[†] Kei Hashimoto[†]

Yoshihiko Nankaku[†] Keiichi Tokuda[†]

[†]Department of Scientific and Engineering Simulation, Nagoya Institute of Technology, Gokiso-cho, Syowa-ku, Nagoya, 466-8555, Japan

表 1: 実験条件

データベース	XM2VTS
使用画像	64×64, 256 階調, グレースケール
学習データ	1 名あたり 6 枚 × 100 名分
テストデータ	1 名あたり 2 枚 × 100 名分
状態数	8×8, 16×16, 24×24, 32×32, 40×40, 48×48, 56×56, 64×64
事前分布	全学習データから学習した UBM
調整パラメータ	500, 1000, 2000, 3000, 4000

表 2: VB 法における VB-EM と VB-DAEM の尤度

状態数	8×8	16×16	24×24	32×32
VB-EM	-15293	-14237	-14166	-14169
VB-DAEM	-15239	-14169	-14004	-13936
状態数	40×40	48×48	56×56	64×64
VB-EM	-14814	-15864	-15864	-17292
VB-DAEM	-14380	-14803	-15584	-17292

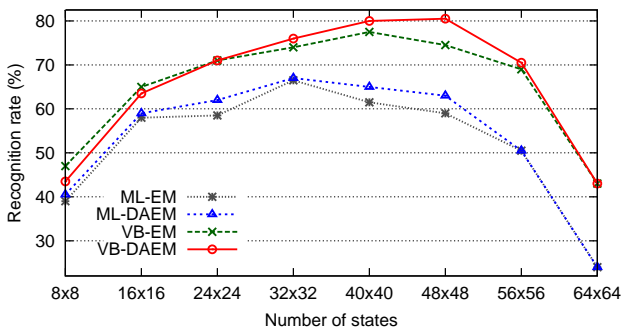


図 1: ML 法と VB 法における EM と DAEM の認識率

5 顔画像認識実験

本稿では提案法の有効性を確認するために XM2VTS データベース [5] を用いて顔画像認識実験を行った。実験条件を表 1 に示す。本実験では DAEM アルゴリズムにおける温度スケジュールを $\beta_k(i) = \left(\frac{i}{20}\right)^{\alpha_k}$ ($i = 1 \dots 20$) と設定した。ここで、 $\alpha_k = 2^n$ とし、各温度で自由エネルギー関数の下限が収束するまで繰り返し学習を行う。

VB 法における EM アルゴリズム (VB-EM) と DAEM アルゴリズム (VB-DAEM) の対数周辺尤度の比較を行った。表 2 に学習時における画像 1 枚あたりの平均対数周辺尤度を示す。また、温度スケジュールは $\alpha_1 = 1, \alpha_2 = 1, \alpha_3 = \frac{1}{64}$ と設定した。ここで、 $\alpha_3 = \frac{1}{64}$ と設定したのは、予備実験により良い結果が得られたためである。表 2 より、VB-DAEM の尤度は VB-EM より高い値となっていることがわかる。特に 40×40 状態と 48×48 状態において、VB-DAEM の尤度は大きく増加している。ただし、64×64 状態は隠れ変数を含まないため、VB-DAEM と VB-EM は同一の結果となる。以上の結果より、DAEM は VB 法の局所最適性の問題を改善していることが確認できた。

ML 法における EM アルゴリズム (ML-EM) と DAEM アルゴリズム (ML-DAEM)、VB 法における VB-EM と VB-DAEM による認識率の比較を行った。図 1 に各手法の認識率を示す。図 1 より、状態数が少ない場合は ML、VB 法ともに EM と DAEM の認識率の差はあまりないが、状態数が多い場合には DAEM は EM より高い認識率が得られた。特に VB 法において認識率が高くなる 40×40 状態と 48×48 状態において DAEM を用いることで認識率は大きく向上した。状態数が少ない場合に VB-DAEM が VB-EM と比べ低い認識率となった原因としては、VB 法で用いる事前分布 $P(\Lambda)$

表 3: 温度スケジュールの影響 (40×40 状態, $\alpha_3 = \frac{1}{64}$)

	$\alpha_2 = \frac{1}{4}$		$\alpha_2 = \frac{1}{2}$		$\alpha_2 = 1$	
	尤度	(%)	尤度	(%)	尤度	(%)
$\alpha_1 = \frac{1}{4}$	-14618	76.5	-14608	77.0	-14601	77.0
$\alpha_1 = \frac{1}{2}$	-14464	80.5	-14450	81.5	-14440	82.0
$\alpha_1 = 1$	-14398	81.0	-14390	80.5	-14380	80.0
$\alpha_1 = 2$	-14409	82.0	-14401	81.5	-14390	81.5
$\alpha_1 = 4$	-14409	82.0	-14404	82.0	-14398	81.5

を VB-EM で最適化したため、VB-DAEM に対して事前分布 $P(\Lambda)$ が最適ではなかった可能性が考えられる。

温度スケジュールの影響を調査するため、VB-DAEM において α_1 と α_2 に個別の温度スケジュールを用いて実験を行った。状態数は 40×40 状態、 $\alpha_3 = \frac{1}{64}$ と設定した。表 3 に α_1 と α_2 に個別の温度スケジュールを用いた場合の尤度と認識率を示す。表 3 より、 α_1 と α_2 に個別の温度スケジュールを用いることにより尤度と認識率に大きな影響を与えることがわかった。出力確率分布に対する温度パラメータ β_1 は $\alpha_1 = 2, 4$, 初期状態確率と状態遷移確率に対する温度パラメータ β_2 は $\alpha_2 = \frac{1}{4}, \frac{1}{2}$ を用いることで高い認識率が得られた。また、出力確率分布に対する温度パラメータ β_1 は初期状態確率と状態遷移確率に対する温度パラメータ β_2 より尤度と認識率に与える影響が大きかったことがわかった。特に $\alpha_1 = \frac{1}{4}$ では尤度と認識率が低くなっているが、これは β_1 の値が 0 から 1 に急激に変化するため、EM と近い学習が行われたためと考えられる。以上の結果より、周辺尤度の各要素に対して個別に適切な温度スケジュールを与えることが重要であることがわかった。

6 むすび

本稿では、VB 法を用いた SL2D-HMM の学習にアニーリング制御を適用した。顔画像認識実験より、VB 法に DAEM アルゴリズムを適用することで局所最適性の問題を改善することができ、対数周辺尤度と認識率が向上した。また、温度パラメータに個別の温度スケジュールを用いることで、モデル構造に適したアニーリング制御を行うことができることを示した。今後の課題としては、よりモデル構造に適した温度スケジュールの調査や、適切な温度スケジュールを自動で設定することが挙げられる。

謝辞 本研究の一部は堀科学芸術振興財団および人工知能研究振興財団の助成による。

参考文献

- [1] D. Kurata, Y. Nankaku, K. Tokuda, T. Kitamura, and Z. Ghahramani, "Face recognition based on separable lattice HMMs," ICASSP, vol.5, pp.737-740, 2006.
- [2] 沢田慶, 玉森聡, 橋本佳, 南角吉彦, 徳田恵一, "変分ベイズ法を用いた分離型 2 次元格子 HMM に基づく顔画像認識," PRMU, vol.111, no.317, pp.125-130, 2011.
- [3] H. Attias, "Inferring parameters and structure of latent variable models by variational Bayes," UAI, pp.21-30, 1999.
- [4] K. Katahira, K. Watanabe, and M. Okada, "Deterministic annealing variant of variational Bayes method," J. Phys.: Conf. Ser, vol.95, p.012015, 2008.
- [5] K. Messer, J. Mates, J. Kitter, J. Luetttin, and G. Maitre, "XM2VTSDB: The extended M2VTS database," AVBPA, pp.72-77, 1999.