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

名古屋工業大学
沢田慶　玉森聡　橋本佳
南角吉彦　徳田恵一

情報処理学会 第74回全国大会 2012年3月8日

はじめに (1/2)

- 統計的手法による画像認識
  - 固有顔・部分空間法などの統計的手法
    - 効率の良い特徴抽出
    - 認識対象の位置や大きさの変動に対応困難
      ⇒タスクに依存した正規化が必要
  - 画像認識タスクに依存しない手法が必要
- 分離型2次元格子HMM [倉田ら; ’06]
  - モデル構造に正規化処理を含む確率モデル
    ⇒認識対象の位置や大きさの変動に対応可能
  - 尤度最大化(maximum likelihood; ML)基準による推定
    - 尤度を最大にする確定的変数を点推定
      ⇒学習データが少量の場合に過学習

はじめに (2/2)

- ベイズ基準による分離型2次元格子HMM [沢田ら; ’11]
  - 事前情報を事前分布として利用可能
  - モデルパラメータを事後分布として表現
  - 積分計算が困難

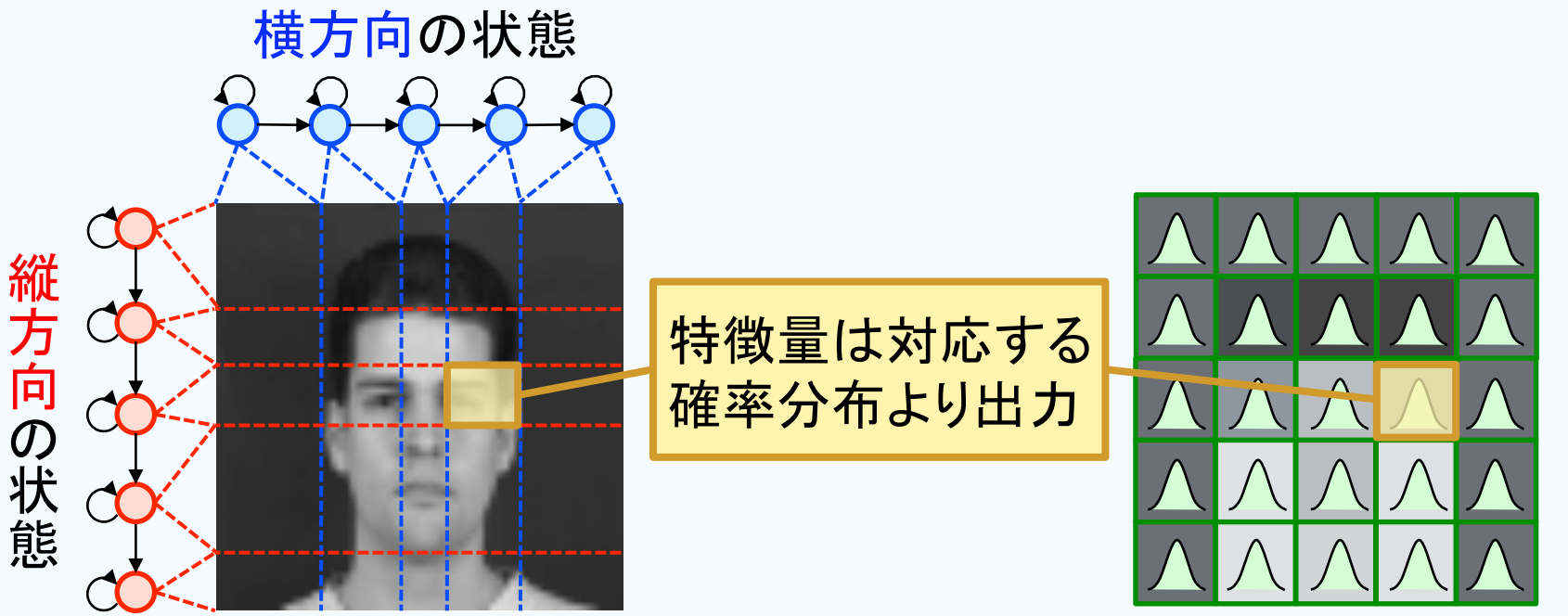
    ⇒変分ベイズ(variational Bayesian; VB)法により近似
    - 反復計算による逐次推定を行うため局所最適性の問題
- 確定的アニーリング(DA)EMアルゴリズム
  - 温度に依存する自由エネルギー関数の最小化
  - 各温度で最適化を確定的に実行
  - ML基準における局所最適性の問題を改善
  - VB法におけるアニーリング制御 [片平ら; ’07]

VB法を用いた分離型2次元格子HMM
の学習にアニーリング制御を適用

分離型2次元格子HMM

- 横方向と縦方向の状態遷移は独立
  - 状態が継続する長さにより画像の変動を表現

⇒認識対象の横方向と縦方向の伸縮に対応可能



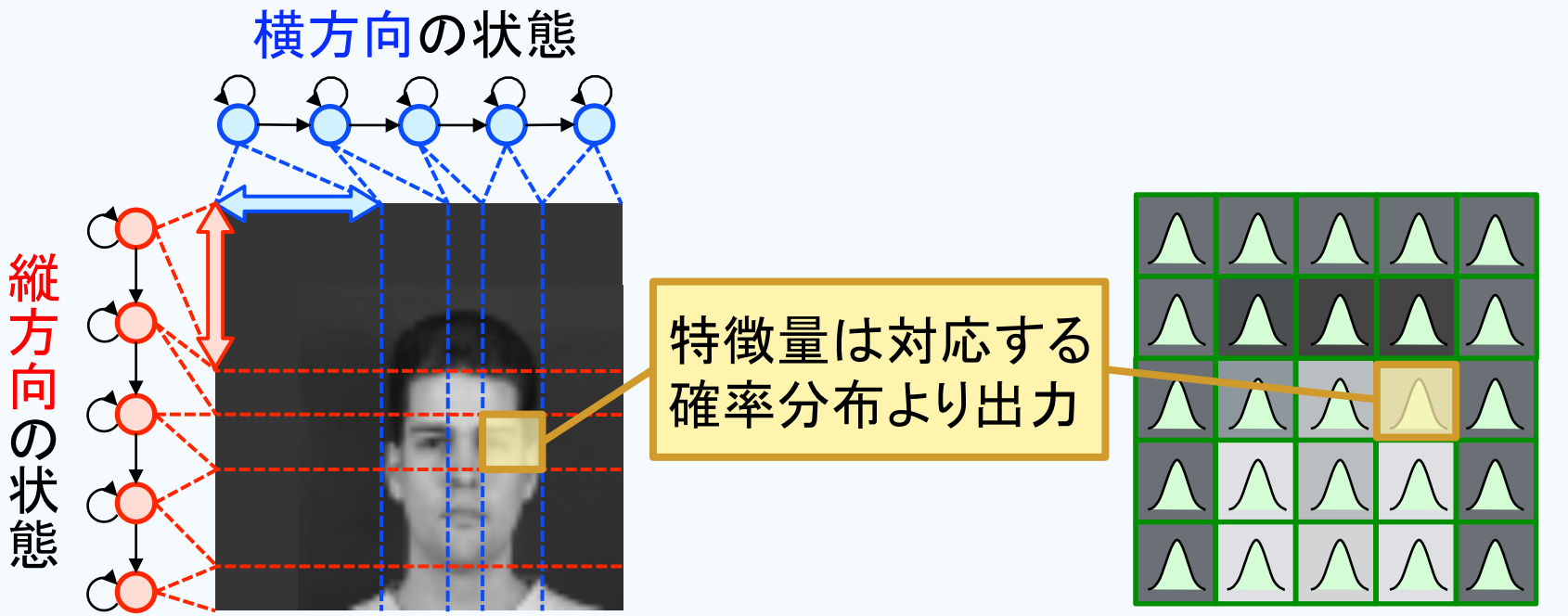
状態で画像を矩形領域に分割

状態の出力確率分布

分離型2次元格子HMM

- 横方向と縦方向の状態遷移は独立
  - 状態が継続する長さにより画像の変動を表現

⇒認識対象の横方向と縦方向の伸縮に対応可能



状態で画像を矩形領域に分割

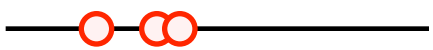
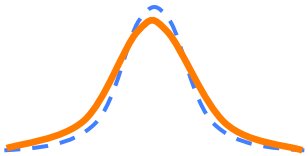
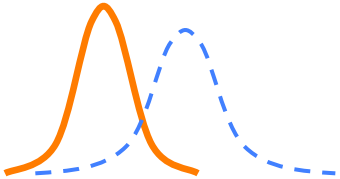
状態の出力確率分布

尤度最大化 (ML) 基準

- 尤度が最大となるモデルパラメータ$\boldsymbol{\Lambda}_{\mathrm{ML}}$を点推定

学習データ $\boldsymbol{O}$ → $\boldsymbol{\Lambda}_{\mathrm{ML}} = \arg\max_{\boldsymbol{\Lambda}} P(\boldsymbol{O} \mid \boldsymbol{\Lambda})$

テストデータ $\boldsymbol{X}$ → 予測分布 $P(\boldsymbol{X} \mid \boldsymbol{\Lambda}_{\mathrm{ML}})$

学習データ量	多量	少量
学習データ		
推定される モデル		

学習データが少量の場合に過学習の問題

ベイズ基準

- モデルパラメータを事後分布として表現

事前情報 → 事前分布 $P(\boldsymbol{\Lambda})$

$$P(\boldsymbol{O}) = \sum_{\boldsymbol{S}} \int P(\boldsymbol{O}, \boldsymbol{S} \mid \boldsymbol{\Lambda}) P(\boldsymbol{\Lambda}) d\boldsymbol{\Lambda}$$

$\boldsymbol{S}$: 隠れ変数

学習データ $\boldsymbol{O}$ → 事後分布 $P(\boldsymbol{\Lambda} \mid \boldsymbol{O}) = \dfrac{P(\boldsymbol{O} \mid \boldsymbol{\Lambda}) P(\boldsymbol{\Lambda})}{\boxed{P(\boldsymbol{O})}}$

テストデータ $\boldsymbol{X}$ → 予測分布 $P(\boldsymbol{X} \mid \boldsymbol{O}) = \int P(\boldsymbol{X} \mid \boldsymbol{\Lambda}) P(\boldsymbol{\Lambda} \mid \boldsymbol{O}) d\boldsymbol{\Lambda}$

利点：事前情報が利用可能，モデルパラメータの周辺化

欠点：積分計算が困難 ⇒
- MCMC法 [Gilks, et al.; ’96]
- MAP法 [Gauvain, et al.; ’94]
- 変分ベイズ法 [Attias; ’99]

変分ベイズ (VB) 法 (1/2)

- 対数周辺尤度$\mathcal{L}(\boldsymbol{O})$の下限$\mathcal{F}$を定義

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

Jensenの不等式

$\boldsymbol{S}$: 状態系列

$Q(\boldsymbol{S}, \boldsymbol{\Lambda})$: 任意関数

- 下限$\mathcal{F}$の最大化による対数周辺尤度$\mathcal{L}(\boldsymbol{O})$の近似

$$
\mathcal{F}(Q) = \mathcal{L}(\boldsymbol{O}) - \mathrm{KL}(Q(\boldsymbol{S}, \boldsymbol{\Lambda}) \,||\, P(\boldsymbol{S}, \boldsymbol{\Lambda} \,|\, \boldsymbol{O}))
$$

$\mathrm{KL}(\cdot \,||\, \cdot)$: KL-divergence

  - 下限$\mathcal{F}$の最大化⇔KL-divergenceの最小化
  - 任意関数$Q(\boldsymbol{S}, \boldsymbol{\Lambda})$は事後分布$P(\boldsymbol{S}, \boldsymbol{\Lambda} \,|\, \boldsymbol{O})$を近似

変分ベイズ (VB) 法 (2/2)

- 近似事後分布の制約条件

$$P(\boldsymbol{S}, \boldsymbol{\Lambda} | \boldsymbol{O}) \approx Q(\boldsymbol{S}, \boldsymbol{\Lambda}) = Q(\boldsymbol{S}) Q(\boldsymbol{\Lambda})$$

$Q(\boldsymbol{S}, \boldsymbol{\Lambda})$: 近似事後分布
$Q(\boldsymbol{S}), Q(\boldsymbol{\Lambda})$: 変分事後分布

- 下限$\mathcal{F}$の最大化

VB E-step	$Q'(\boldsymbol{S}) = \arg\max_{Q(\boldsymbol{S})} \mathcal{F}$
VB M-step	$Q'(\boldsymbol{\Lambda}) = \arg\max_{Q(\boldsymbol{\Lambda})} \mathcal{F}$

交互に最適化

- 事後分布の近似分布を推定
  - ベイズ基準を近似可能でML基準より高い汎化能力
  - EMアルゴリズムと同様の反復計算による逐次推定

    ⇒初期値に依存した局所最適性の問題

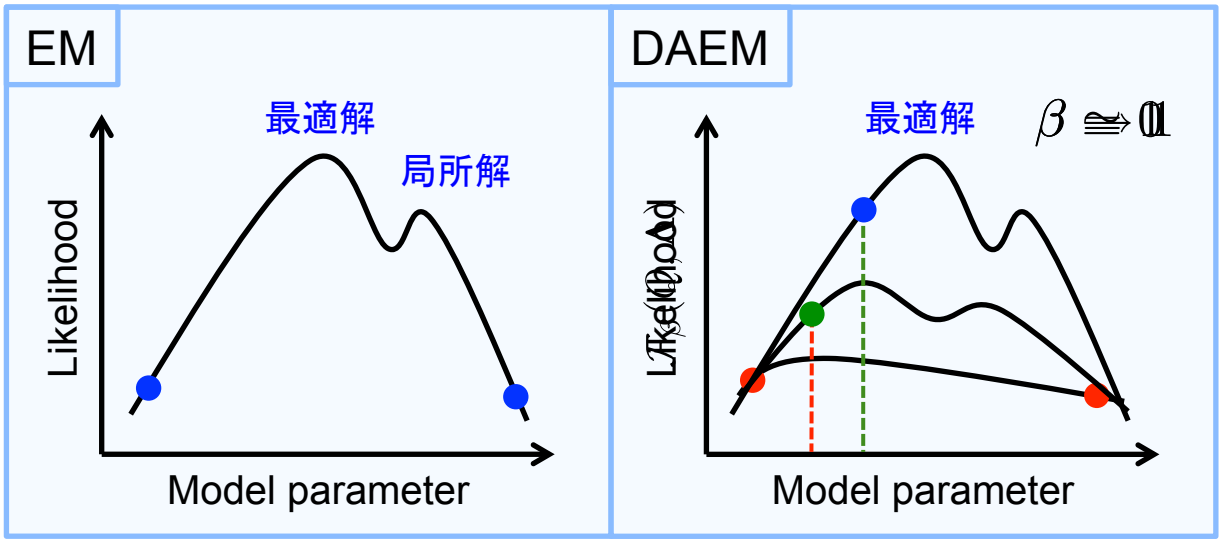
確定的アニーリング(DA)EMアルゴリズム

- 自由エネルギー関数に温度パラメータを導入した下限

$$\mathcal{F}_\beta(Q,\boldsymbol{\Lambda}) = \sum_{\boldsymbol{S}} Q(\boldsymbol{S}) \ln \frac{P(\boldsymbol{O},\boldsymbol{S}\,|\,\boldsymbol{\Lambda})^\beta}{Q(\boldsymbol{S})}$$

β ：温度パラメータ

- 温度パラメータによるアニーリング制御
  - $\beta \simeq 0$ ⇒ 下限 $\mathcal{F}_\beta$ は唯一の極大値
  - $\beta \to 1$ ⇒ 下限 $\mathcal{F}_\beta$ は複数の極大値
  - $\beta = 1$ ⇒ 本来の尤度関数



初期値に
依存しにくい
推定が可能

VB法におけるアニーリング制御の適用

- 自由エネルギー関数に3つの温度パラメータを導入

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

β_1 : 出力確率分布の温度パラメータ　β_2 : 状態遷移確率の温度パラメータ

β_3 : 事前分布の温度パラメータ

- 自由エネルギー関数の下限$\mathcal{F}_{\beta_1,\beta_2,\beta_3}$を定義

$$\mathcal{F}_{\beta_1,\beta_2,\beta_3}(Q) = \sum_{\boldsymbol{S}} \int Q(\boldsymbol{S}, \boldsymbol{\Lambda}) \ln \frac{P(\boldsymbol{O} \,|\, \boldsymbol{S}, \boldsymbol{\Lambda})^{\beta_1} P(\boldsymbol{S} \,|\, \boldsymbol{\Lambda})^{\beta_2} P(\boldsymbol{\Lambda})^{\beta_3}}{Q(\boldsymbol{S}, \boldsymbol{\Lambda})} d\boldsymbol{\Lambda}$$

- アニーリング制御における温度スケジュール
  - 各温度パラメータに個別の温度スケジュール

    ⇒モデル構造に適したアニーリング制御が可能

局所最適性の問題を改善可能

顔画像認識実験 実験条件 (1/2)

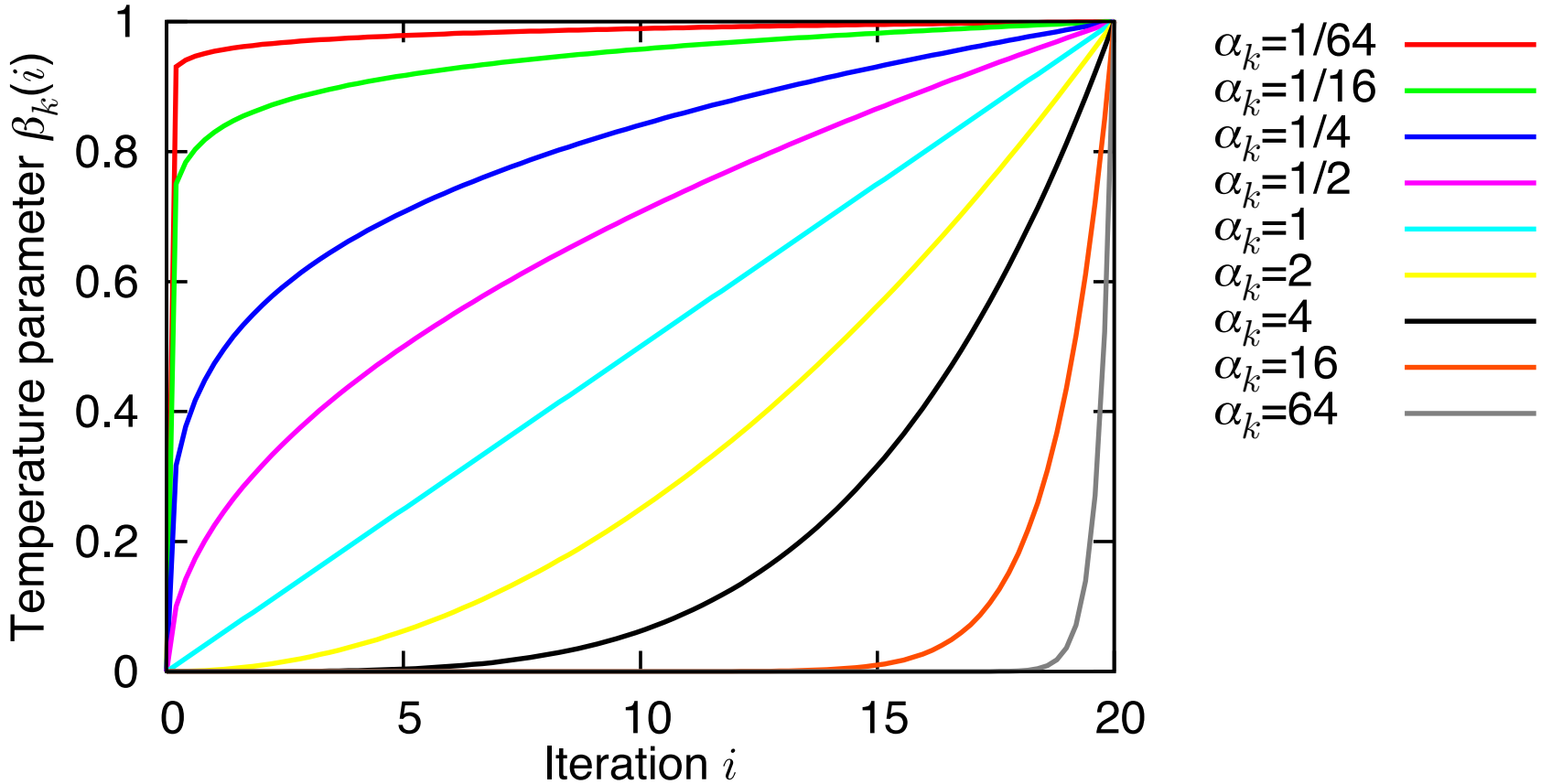
データベース	XM2VTS
画像サイズ	64×64
階調数	256, グレースケール
学習データ	1名あたり6枚 × 100名分
テストデータ	1名あたり2枚 × 100名分
状態数	8×8, 16×16, 24×24, 32×32, 40×40, 48×48, 56×56, 64×64
事前分布	全学習データから学習したモデル
調整パラメータτ	500, 1000, 2000, 3000, 4000

顔画像認識実験　実験条件 (2/2)

温度スケジュール

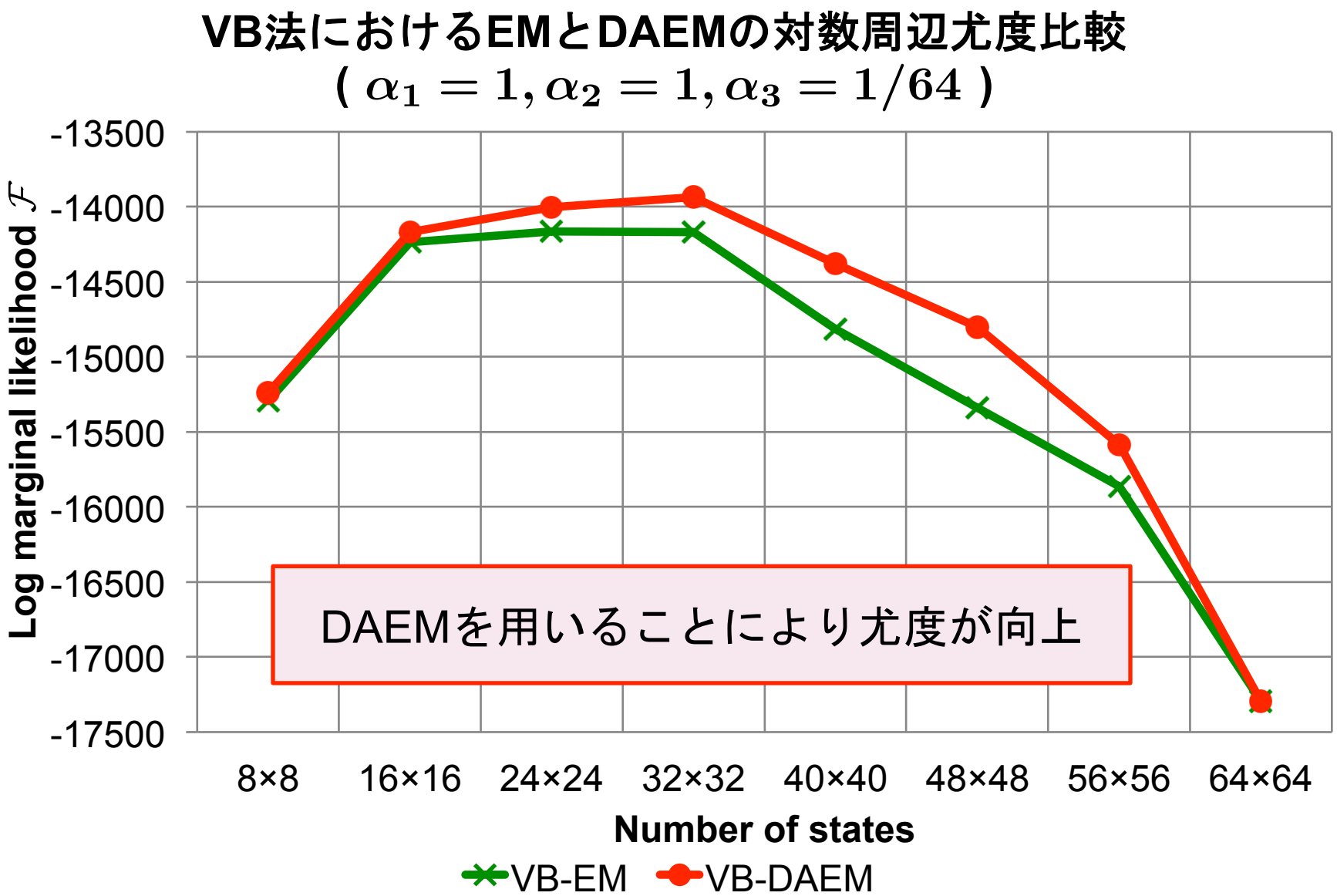
$$\beta_k(i) = \left(\frac{i}{20}\right)^{\alpha_k} \quad (k = 1, 2, 3 \quad i = 1 \ldots 20 \quad \alpha_k = 2^n)$$

$k = 1$: 出力確率分布の変数
$k = 2$: 状態遷移確率の変数
$k = 3$: 事前分布の変数



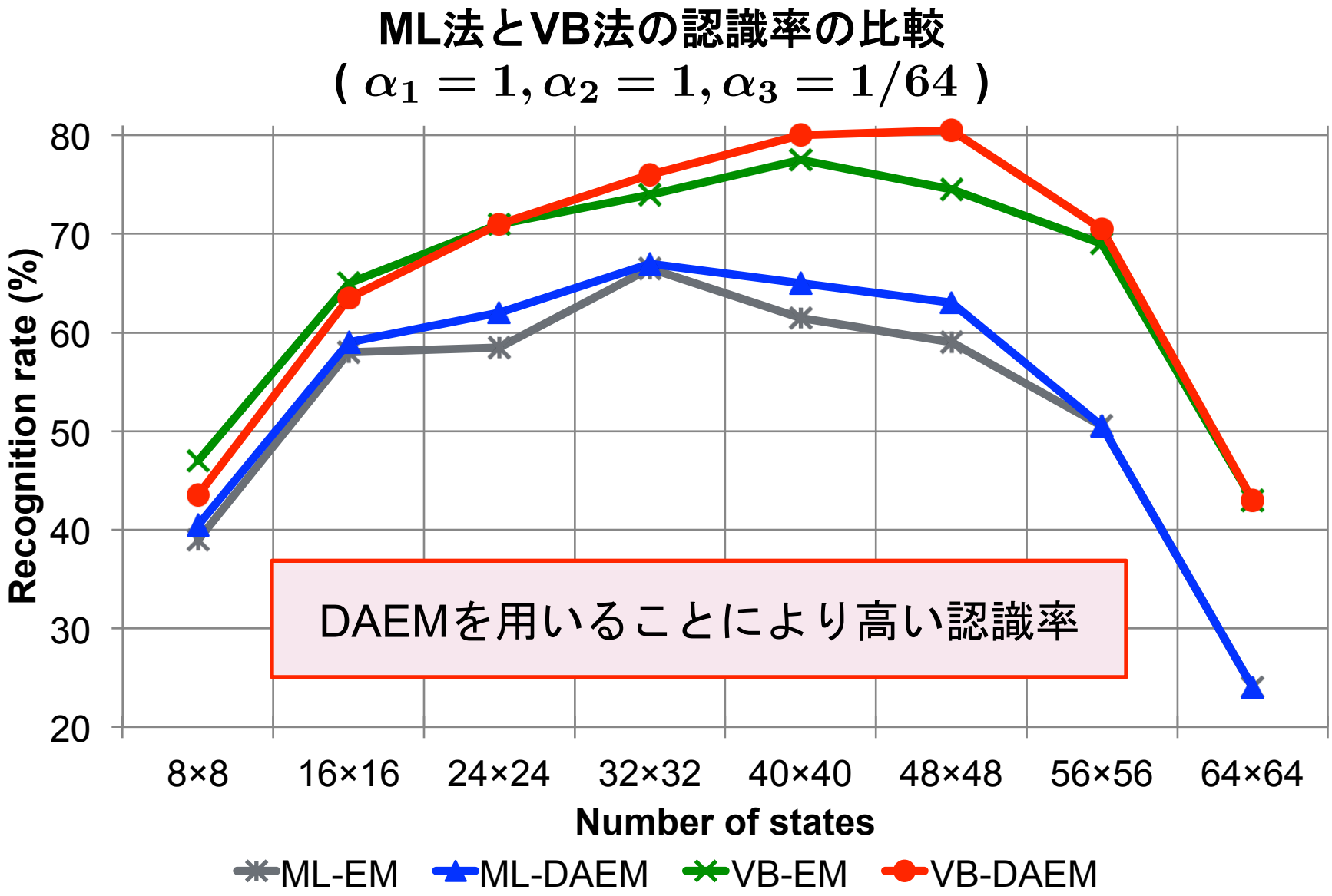
α_1 : 出力確率のスケジュールを調整　α_2 : 状態遷移のスケジュールを調整　α_3 : 事前分布のスケジュールを調整

対数周辺尤度の比較



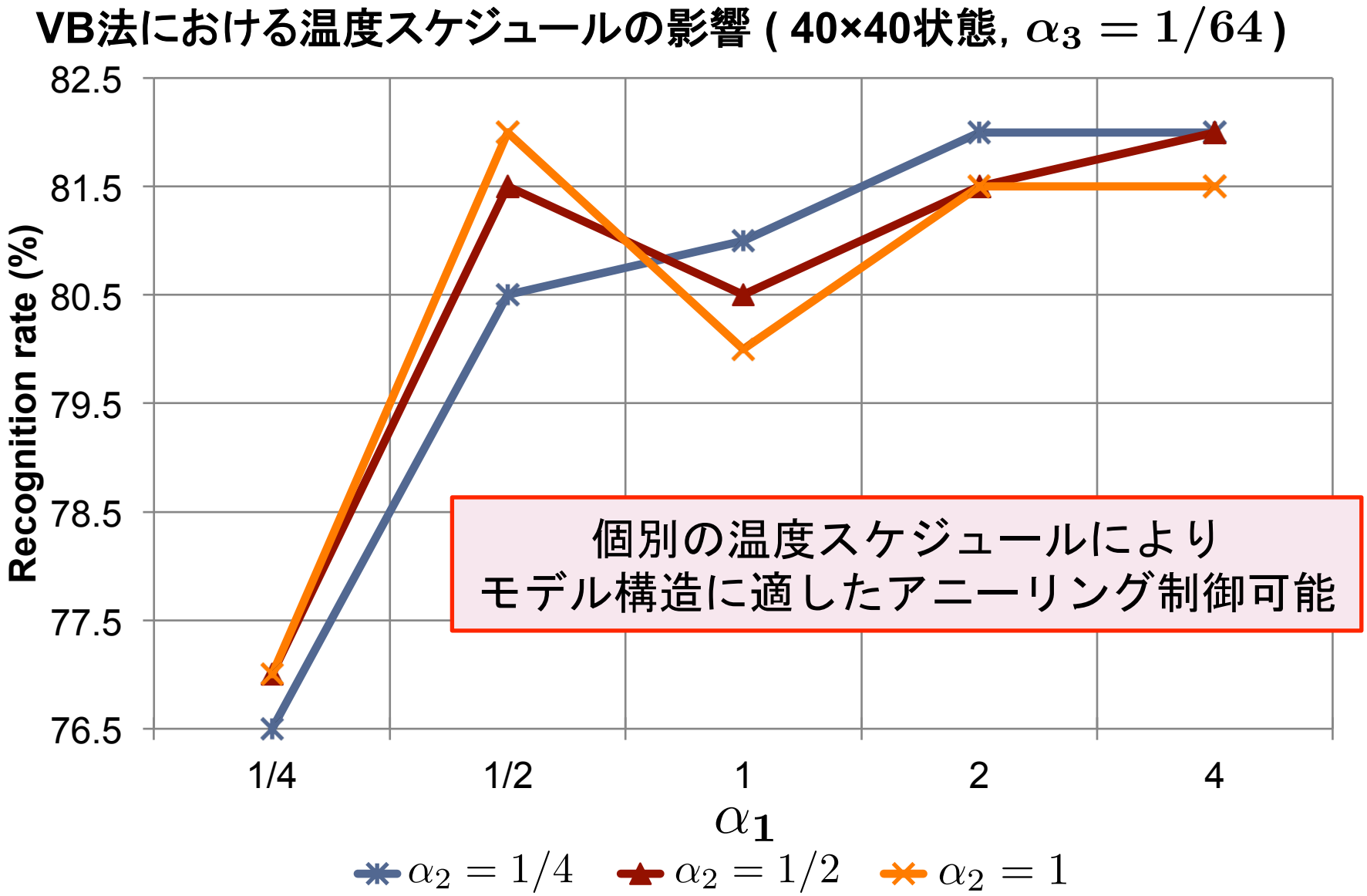
α_1 : 出力確率のスケジュールを調整　α_2 : 状態遷移のスケジュールを調整　α_3 : 事前分布のスケジュールを調整

認識率の比較



α_1 : 出力確率のスケジュールを調整　α_2 : 状態遷移のスケジュールを調整　α_3 : 事前分布のスケジュールを調整

温度スケジュールの影響



個別の温度スケジュールにより
モデル構造に適したアニーリング制御可能

α_1 : 出力確率のスケジュールを調整　α_2 : 状態遷移のスケジュールを調整　α_3 : 事前分布のスケジュールを調整

むすび

- 変分ベイズ法を用いた分離型2次元格子HMMの学習にアニーリング制御
  - アニーリング制御を適用した学習アルゴリズムを導出
    - 3つの温度パラメータを導入

      ⇒各温度パラメータに個別の温度スケジュールが可能
  - アニーリング制御により高い対数周辺尤度と認識率

    ⇒局所最適性の問題を改善
  - モデル構造に適した温度スケジュールが存在

- 今後の課題
  - 温度スケジュールの最適化に関する検討
  - 事前分布の検討