

量子揺らぎとLatent Dynamics

佐藤 一誠

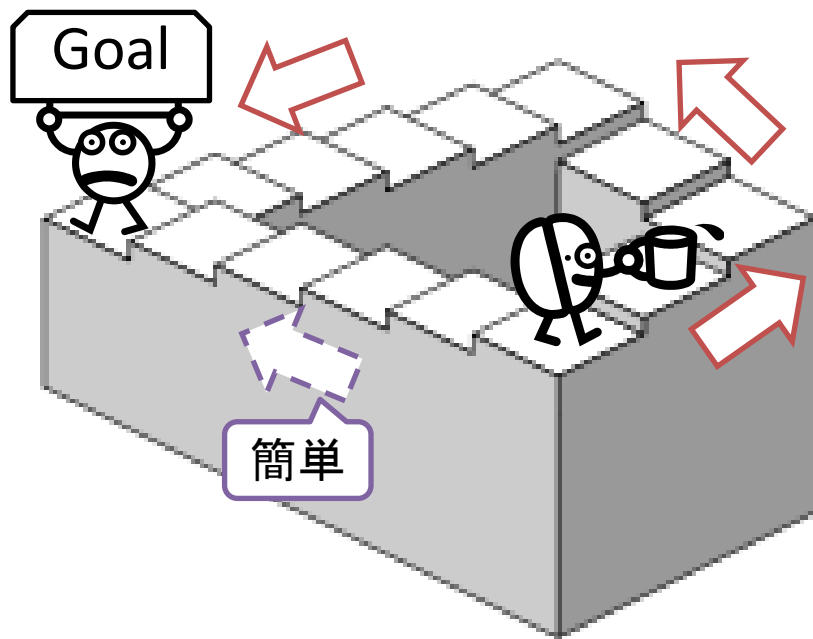
東京大学 情報基盤センター

- 参考文献(主に以下の論文について紹介)-

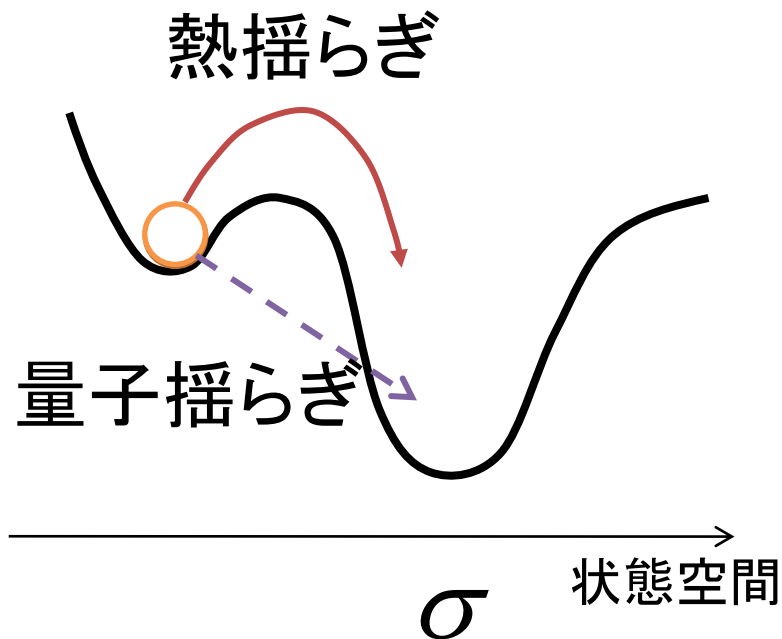
UAI2009: Quantum Annealing for Variational Bayes Inference

Joint work with Kenichi Kurihara, Shu Tanaka,
Seiji Miyashita and Hiroshi Nakagawa

Ph.D Thesis (2011): Quantum annealing in Statistical Machine Learning



ペンローズの階段
出典: Wikipedia



古典系とは異なるパスが
量子系で見つければ
簡単に状態遷移できる？

潜在変数モデル＋量子揺らぎ＋変分ベイズ法
による学習を考える

もくじ

- 密度行列
- 量子揺らぎ＋変分ベイズ法
- 実験

潜在変数の表現

潜在変数: $z_i = 2$

e.g. K-means, Iterative Conditional Mode, Sampling

確率ベクトル: $q(z_i) = \left(\frac{3}{4}, \frac{1}{4} \right)$

e.g. EM algorithm, 変分ベイズ法

密度行列: $\Phi = \begin{pmatrix} \frac{3}{4} & 0 \\ 0 & \frac{1}{4} \end{pmatrix}$ を考える

密度行列 Φ

- self-adjoint positive-semidefinite matrix
- trace one

$$\forall \sigma \in H, \sigma^T \Phi \sigma \geq 0, \text{Tr}\{\Phi\} = 1$$

H : ヒルベルト空間

密度行列と不確かさ

$$(p_1, p_2, p_3, p_4)^T$$



$\sigma^{(l)}$: indicator vector

$$\sigma^{(2)} = (0, 1, 0, 0)^T$$

$$\begin{pmatrix} p_1 & 0 & 0 & 0 \\ 0 & p_2 & 0 & 0 \\ 0 & 0 & p_3 & 0 \\ 0 & 0 & 0 & p_4 \end{pmatrix} = \sum_{i=1}^4 p_i \sigma^{(i)} \sigma^{(i)T}$$

Introduce
uncertainty



Density matrix

$$\Phi = \sum_i p_i \underbrace{\mathbf{u}_i \mathbf{u}_i^T}_{\text{State vectors}}$$

p_i Mixture proportions

例えば

$$\begin{pmatrix} \frac{3}{4} & 0 \\ 0 & \frac{1}{4} \end{pmatrix} \begin{array}{l} \text{古典系では指標ベクトルによる} \\ \text{状態ベクトルの系でのみ考えている} \end{array}$$

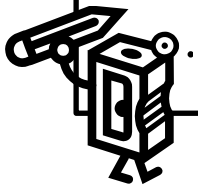
↙ ↘

$$= \frac{3}{4} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} + \frac{1}{4} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \end{pmatrix}$$
$$= \frac{1}{2} \begin{pmatrix} \frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{pmatrix} \begin{pmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \end{pmatrix} + \frac{1}{2} \begin{pmatrix} \frac{\sqrt{3}}{2} \\ -\frac{1}{2} \end{pmatrix} \begin{pmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}$$

Born Rule

A density matrix Φ assigns a probability to the unit vector σ given by

$$p(\sigma) = \sigma^T \Phi \sigma$$



- Gleason's theorem (1957) -

量子論的確率測度 μ に対して、対応する密度行列 Φ が1つ存在する

密度行列の非対角項の効果

$$\sigma^* = \arg \max p(\sigma) = \sigma^T \Phi \sigma$$

を考える

古典系: Φ が対角行列のとき

σ^* は Φ の対角要素の最大値に対応する

指標ベクトル e.g. (1,0), (0,1) ...

状態が確定する

密度行列の非対角項の効果

$$\sigma^* = \arg \max p(\sigma) = \sigma^T \Phi \sigma$$

を考える

量子系: Φ が非対角行列のとき

σ^* は Φ の最大固有値に対応する

確率
振幅

固有ベクトル e.g. $(\sqrt{3}/2, 1/2), (1/2, -\sqrt{3}/2) \dots$

状態が確定しない

手法 \ 不确实性	Mixture Proportions	State Vectors
EM [Dempster,'77]	✓	×
DA [Ueda+,'97]	✓	×
VB [Attias,'99]	✓	×
This work	✓	✓



もくじ

- 密度行列
- 量子揺らぎ＋変分ベイズ法
- 実験

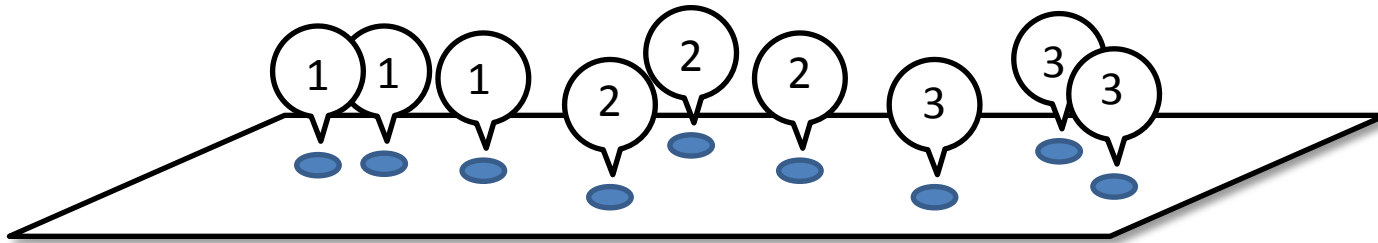
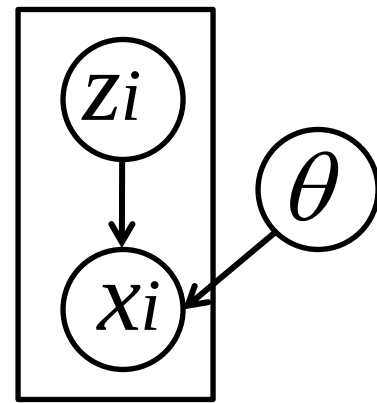
準備

● : データ点 x_i

○ : 潜在変数

N : データ数

取りうる値 $\{1, 2, \dots, K\}$



各データ点の ○ のとり方を σ で表現する

σ : K^N 次元の指標ベクトル $L=K^N$ とする

密度行列によるモデル化

エネルギー関数:

$$E[\sigma^{(\ell)}] = -\log p(\mathbf{x}, \sigma^{(\ell)})$$

(古典系)ハミルトニアン:

$$\mathcal{H}_c = \begin{pmatrix} E[\sigma^{(1)}] & & & 0 \\ & E[\sigma^{(2)}] & & \\ & & \ddots & \\ 0 & & & E[\sigma^{(L)}] \end{pmatrix}$$

密度行列:

$$\frac{1}{Z} e^{-\mathcal{H}_c} \quad Z = \text{Tr}\{e^{-\mathcal{H}_c}\}$$

量子系ハミルトニアン

$\mathcal{H} = \mathcal{H}_c + \mathcal{H}_q$, n 個のデータが K 個の状態を取るとき

$$\mathcal{H}_q = \sum_{i=1}^n \sigma_{xi}, \quad \sigma_{xi} = \left(\bigotimes_{j=1}^{i-1} \mathbb{E}_K \right) \otimes \sigma_x \otimes \left(\bigotimes_{l=i+1}^n \mathbb{E}_K \right),$$

$\sigma_x = \Gamma(\mathbb{E}_K - 1_K)$,
※ $K=2$ のとき[Kadowaki,1998]と同じ
 $K>2$ の時はパウリ行列が使えなくなるので
経路積分の計算を頑張る必要がある

$n=2, K=2$ のとき

$$\begin{pmatrix} -\log p(\mathbf{x}_{1:n}, \sigma^{(1)}) & -\Gamma & -\Gamma & 0 \\ -\Gamma & -\log p(\mathbf{x}_{1:n}, \sigma^{(2)}) & 0 & -\Gamma \\ -\Gamma & 0 & -\log p(\mathbf{x}_{1:n}, \sigma^{(3)}) & -\Gamma \\ 0 & -\Gamma & -\Gamma & -\log p(\mathbf{x}_{1:n}, \sigma^{(4)}) \end{pmatrix}$$

ハミルトニアンが作れればあとは・・・

シュレディンガー方程式：

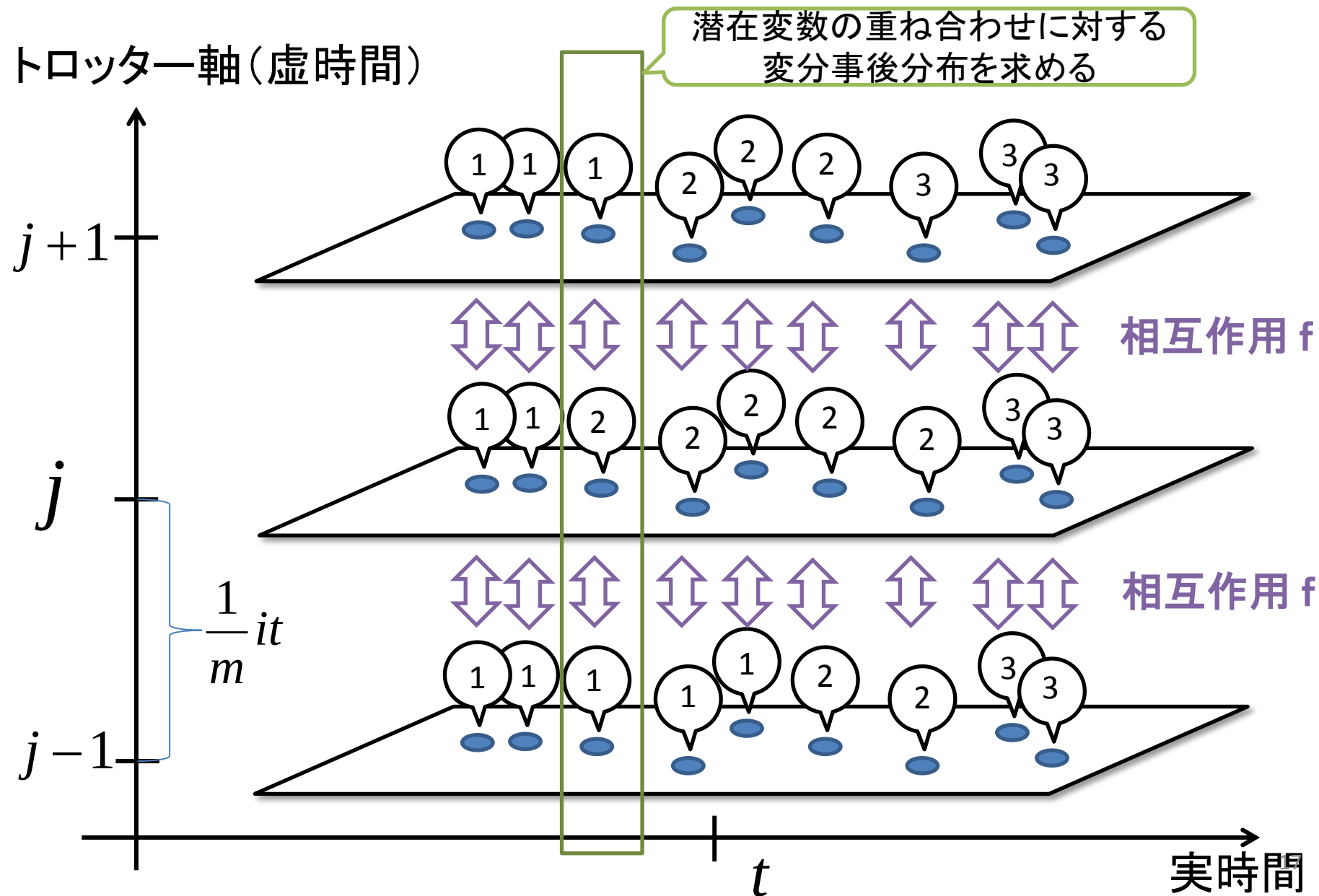
$$i\hbar \frac{\partial}{\partial t} |\sigma(t)\rangle = \mathcal{H} |\sigma(t)\rangle$$

量子状態の時間発展：

$$|\sigma(t + \Delta t)\rangle = e^{-i\mathcal{H}\Delta t/\hbar} |\sigma(t)\rangle$$

計算量的に実行不可能
→変分ベイズ法を導出する

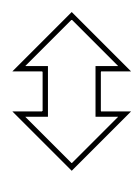
概要：経路積分表示＋変分ベイズ法



変分ベイズ法

[Attias, 1999]

$$q(\sigma, \theta) = \arg \min_{q(\sigma, \theta)} KL[q(\sigma, \theta) \mid p(\sigma, \theta \mid x_{1:N})]$$

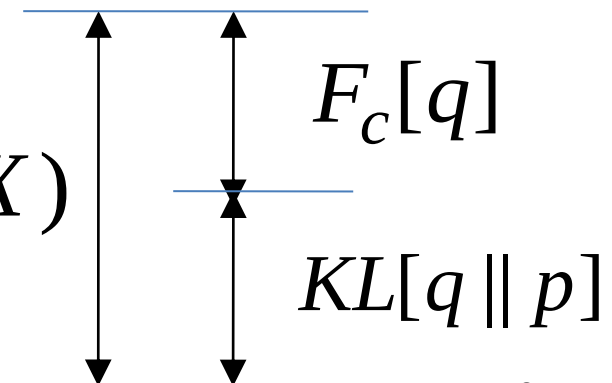


$$q(\sigma, \theta) = \arg \max_{q(\sigma, \theta)} F_c[q(\sigma, \theta)]$$

$$\log p(X) \geq F_c[q]$$

変分下限

$$\log p(X)$$



ポイント

変分ベイズ法をハミルトニアンを用いて再定式化

$$\log p(\boldsymbol{x}) = \log \text{Tr}\{e^{-\mathcal{H}_c}\} \geq F_c[q]$$

$$\Downarrow \mathcal{H} = \mathcal{H}_c + \mathcal{H}_q$$

$$\log p(\boldsymbol{x}; \beta, \Gamma) = \log \text{Tr}\{e^{-\beta \mathcal{H}}\} \geq F[q]$$

変分ベイズ法 +

鈴木・トロッター展開

- トロッター公式
- 経路積分

[Suzuki, 1976]

変分下限の導出

$$F[q] = \frac{\beta}{m} \sum_{j=1}^m F_c[q(\sigma_j)] + f(\beta, \Gamma) R[q(\sigma_1, \sigma_2 \dots \sigma_m)] + O\left(\frac{1}{m}\right)$$

M個の古典系の和

事後分布の制約項

$$R[q(\sigma_1, \sigma_2 \dots \sigma_m)] = \sum_{j=1}^m \sum_{\sigma_j} \sum_{\sigma_{j+1}} q(\sigma_j) q(\sigma_{j+1}) s(\sigma_j, \sigma_{j+1})$$

$$s(\sigma_j, \sigma_{j+1}) = \sum_{i=1}^N \delta(\sigma_{j,i}, \sigma_{j+1,i})$$

状態間の類似度関数

相互作用関数

$$f(\beta, \Gamma) = \log\left(\frac{a+b}{b}\right)$$

$$a = e^{-\frac{\beta\Gamma}{m}}$$

$$b = \frac{1}{K}e^{-\frac{\beta\Gamma}{m}(1-K)} - \frac{1}{K}e^{-\frac{\beta\Gamma}{m}}$$

更新式

・潜在変数の事後分布

$$q(\sigma_{j,i,k}) \propto \exp\left\{ \int q(\theta_j) \beta_{\text{eff}} \log p(\mathbf{x}, \sigma_j, \theta_j) d\theta_j \right. \\ \left. + f(\beta, \Gamma) \left(\underbrace{q(\sigma_{j-1,i, \rho_{j-1}^{-1}(k)})}_{\text{j-1でのデータiの}} + \underbrace{q(\sigma_{j+1,i, \rho_j(k)})}_{\text{j+1でのデータiの}} \right) \right\}$$

問題点

簡単にはわからない

j-1でのデータiの

潜在変数の事後分布

j+1でのデータiの

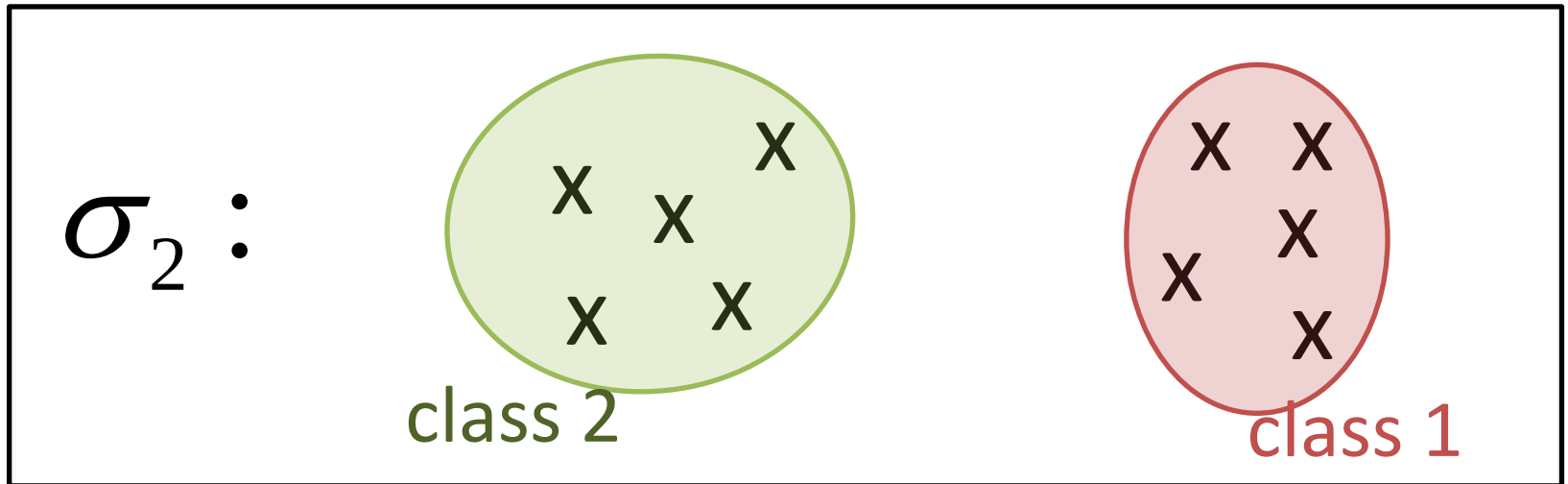
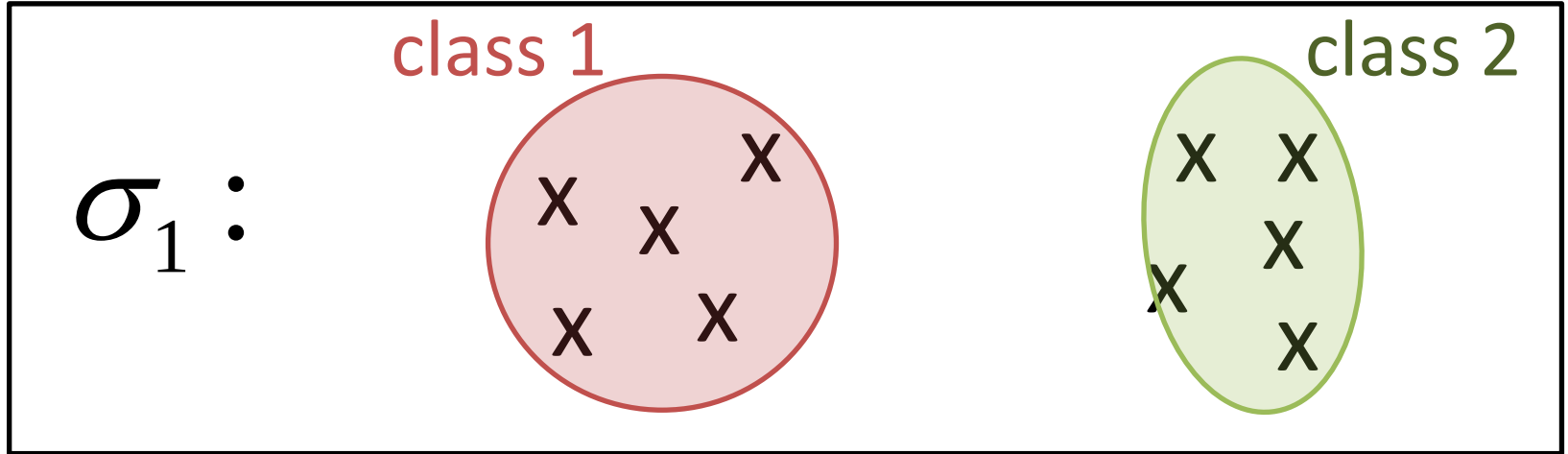
潜在変数の事後分布

j±1で潜在変数がとっている状態ほど

Jでの確率も高くなる

この項を追加するだけ・・・？

Label Identification Problem



Label projection

$$\rho = \arg \max_{\rho} R = \sum_{j=1}^m \sum_{i=1}^N \sum_{k=1}^K q(\sigma_{j,i,k}) q(\sigma_{j+1,i,\rho(k)})$$

j+1

Label projection
matrix

j

label	1	2	3
1		○	
2			△
3	□		

$$O = \sum_{i=1}^N q(\sigma_{j,i,1}) q(\sigma_{j+1,i,2})$$

max ○+△+□

完全マッチ: $O(K^3)$ アルゴリズム

重複マッチ: $O(K^2)$ アルゴリズム

MAP重複マッチ: $O(K)$ アルゴリズム

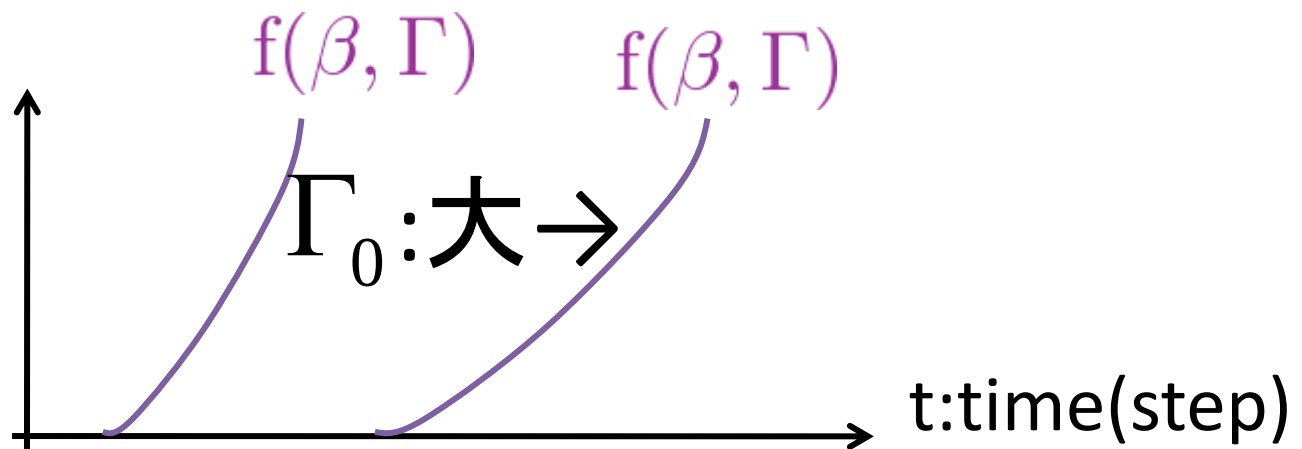
揺らぎの制御

温度

$$T = T_0 \frac{1}{\sqrt{t}}$$

量子効果

$$\Gamma = \Gamma_0 \frac{1}{t^2}$$

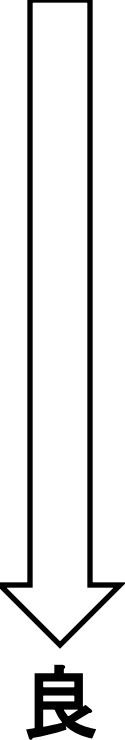


もくじ

- 密度行列
- 量子揺らぎ＋変分ベイズ法
- 実験

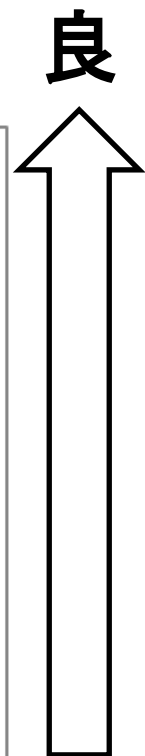
実験 1

- モデル: Unigram mixture (UM) model
- データセット: 文書データ
 - BBCコーパス:
 - 約2500文書, 約9千語彙, 5カテゴリー
 - 20NGコーパス
 - 約4000文書, 約1万語彙, 8カテゴリー
- 評価方法:
 - Free energy
 - F-measure(文書分類)



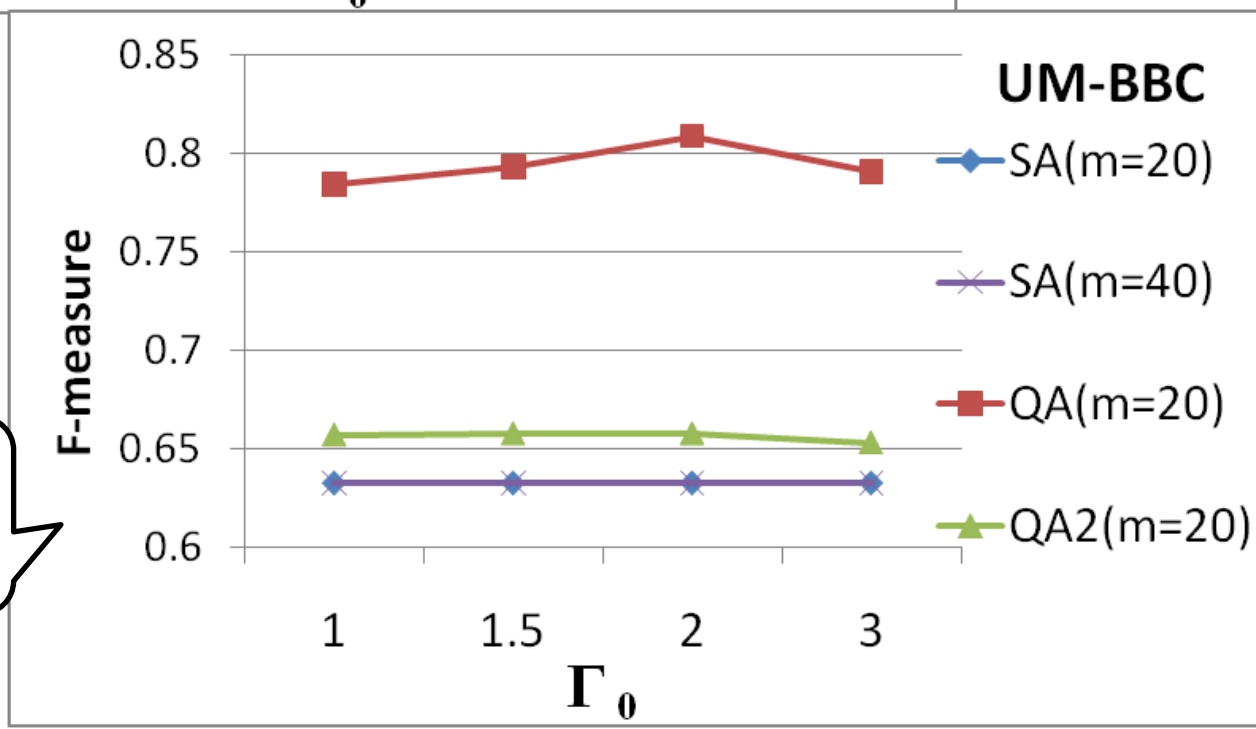
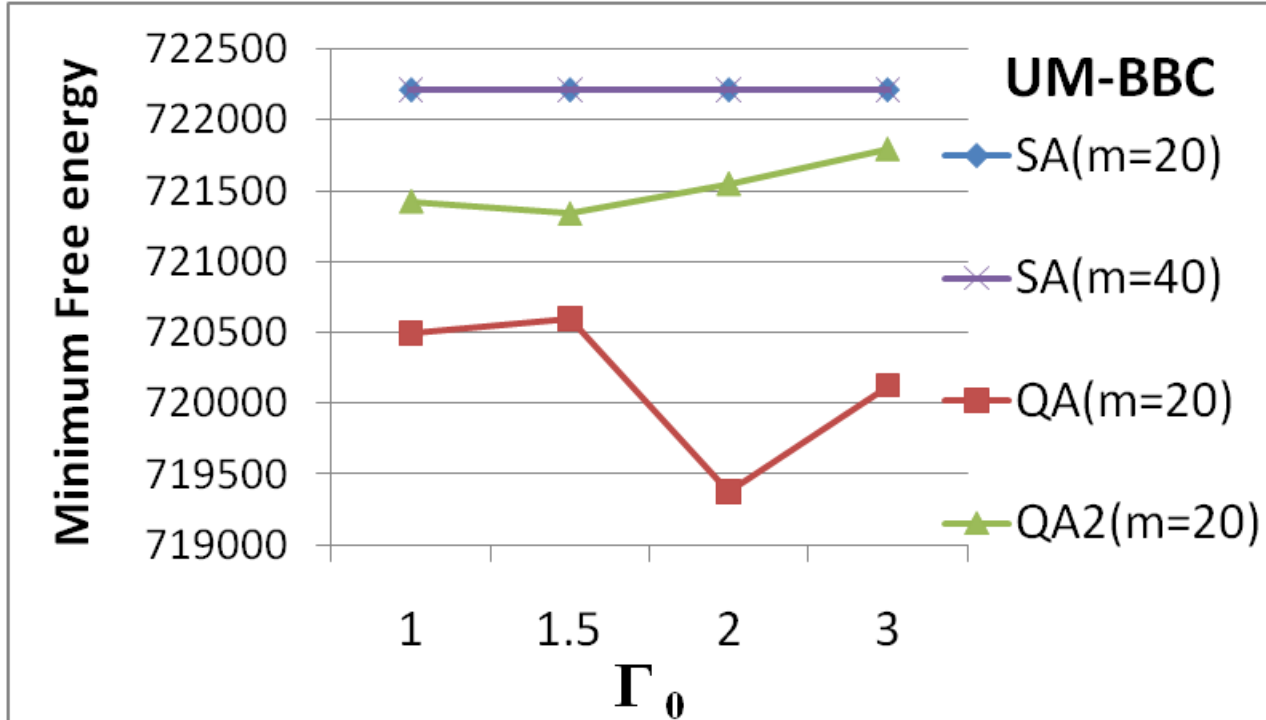
良

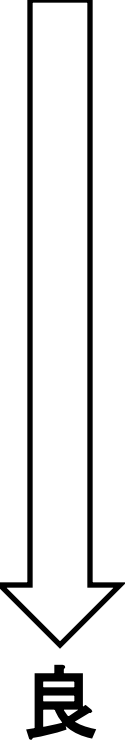
目的



28

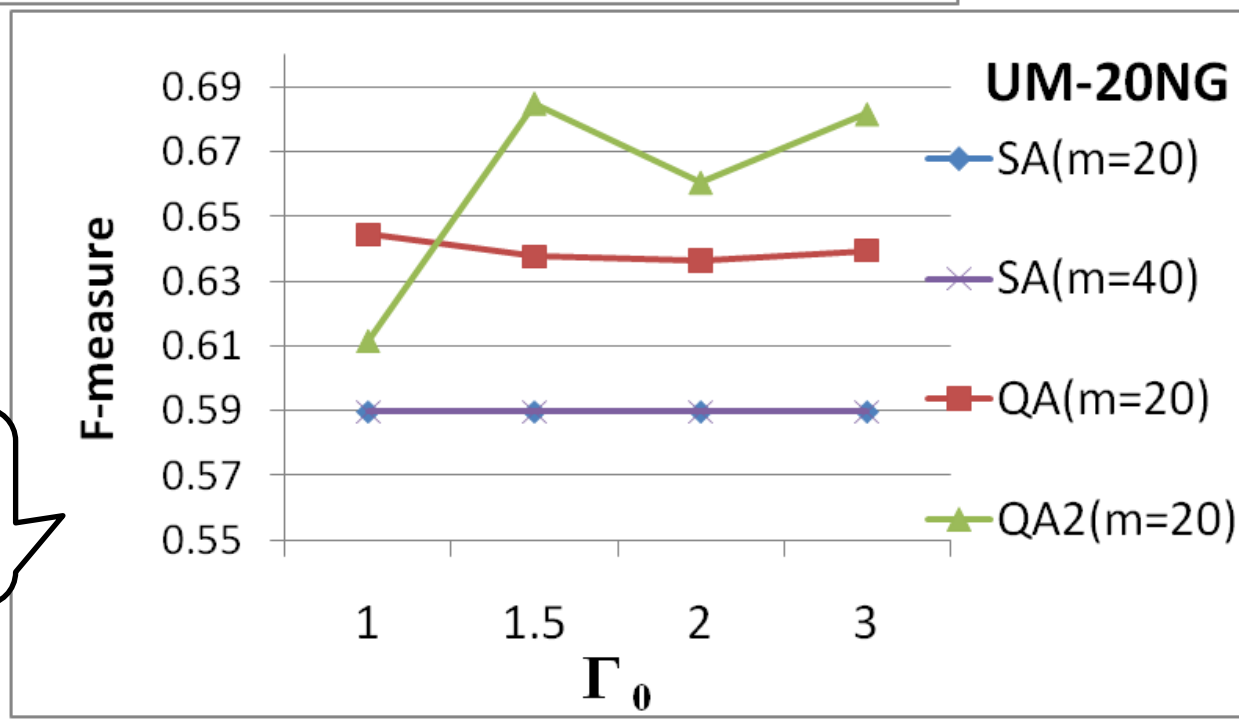
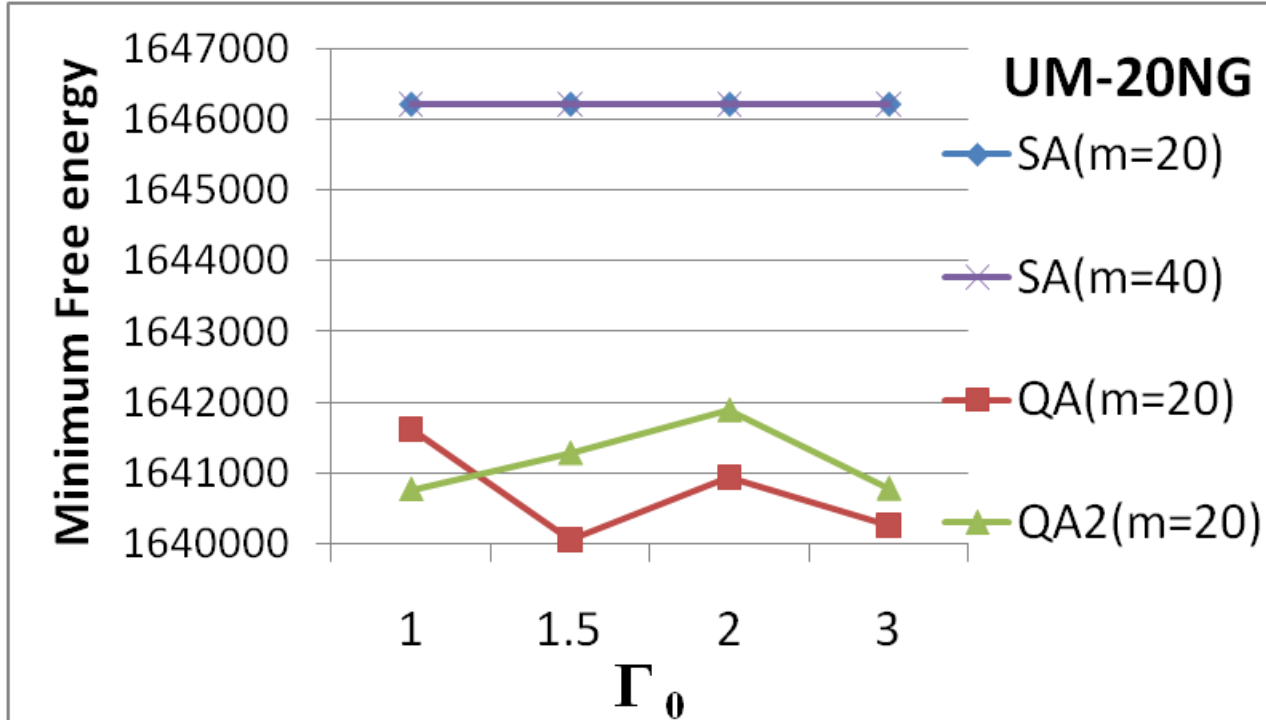
文書分類
性能



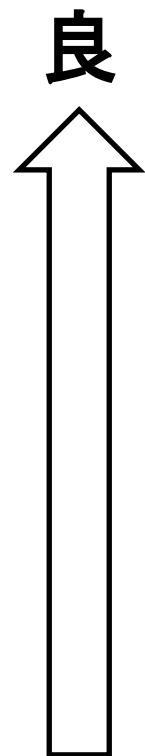


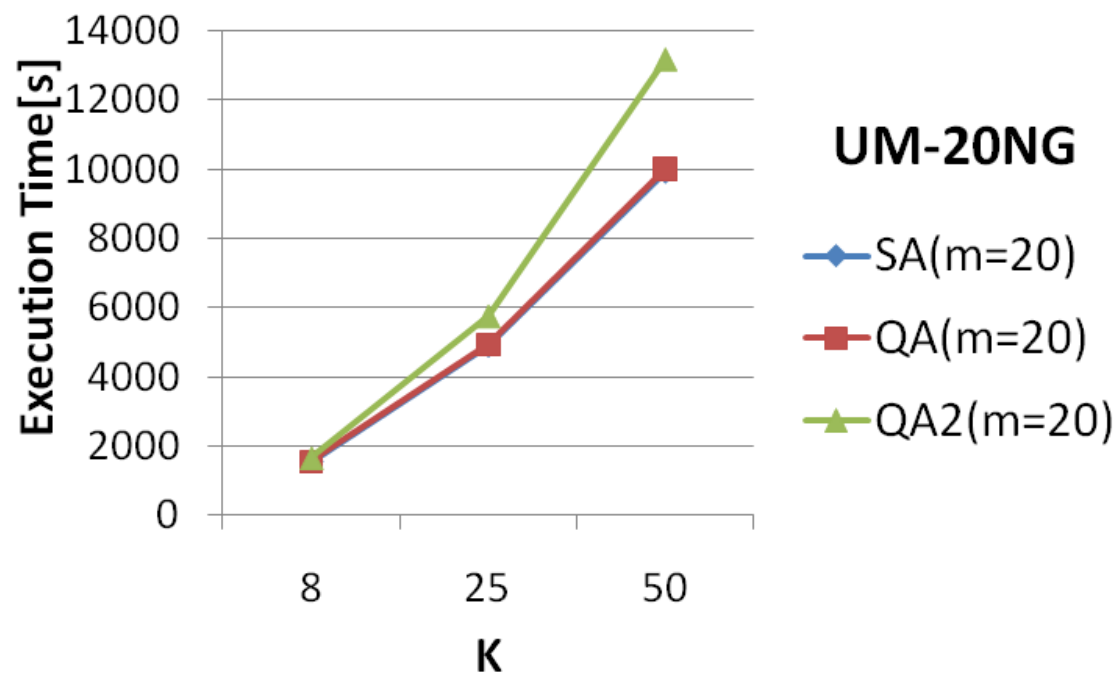
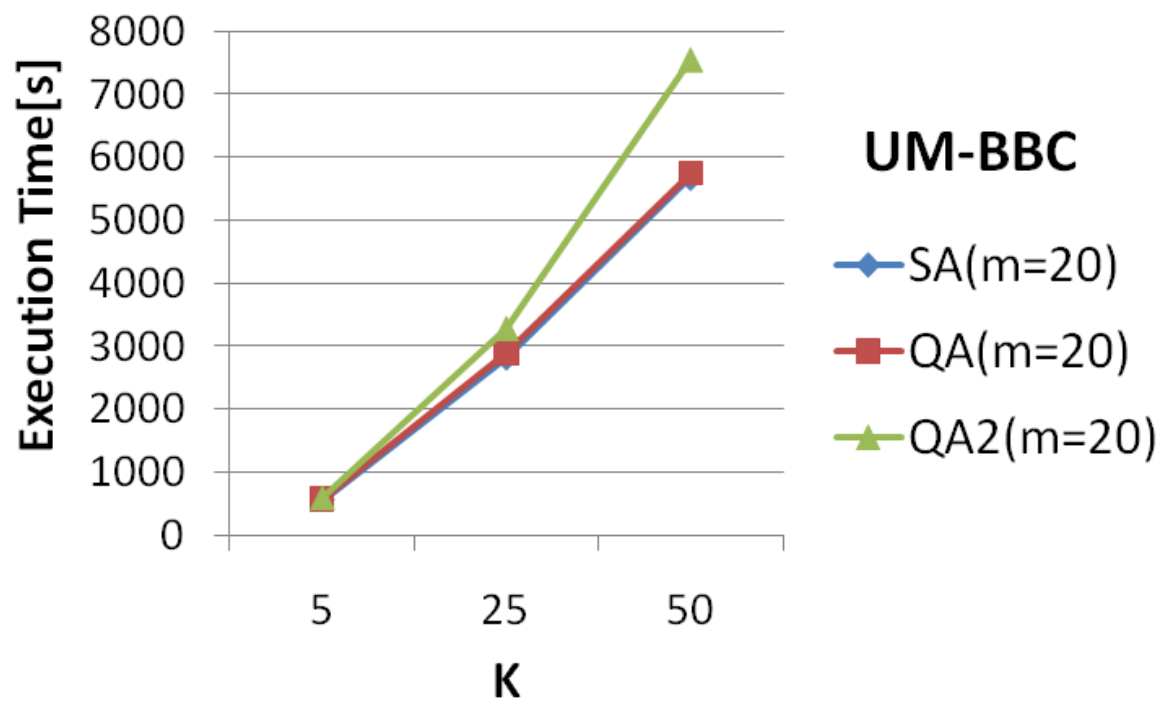
良

目的



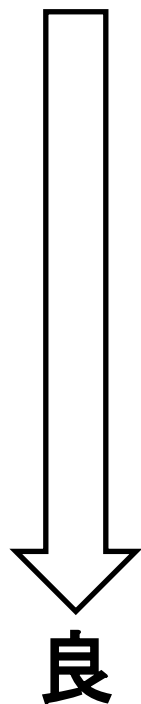
文書分類
性能



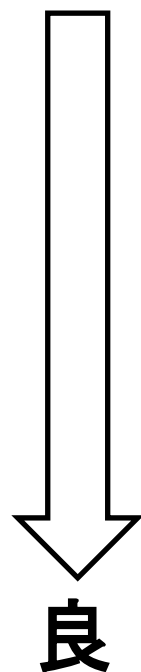
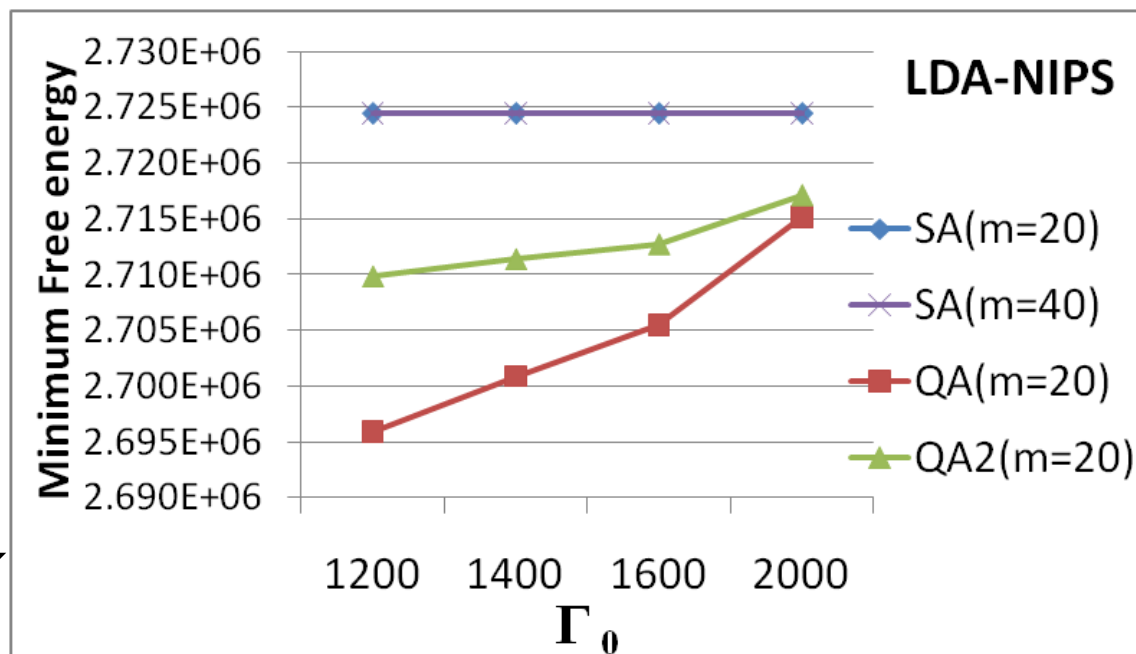


実験 2

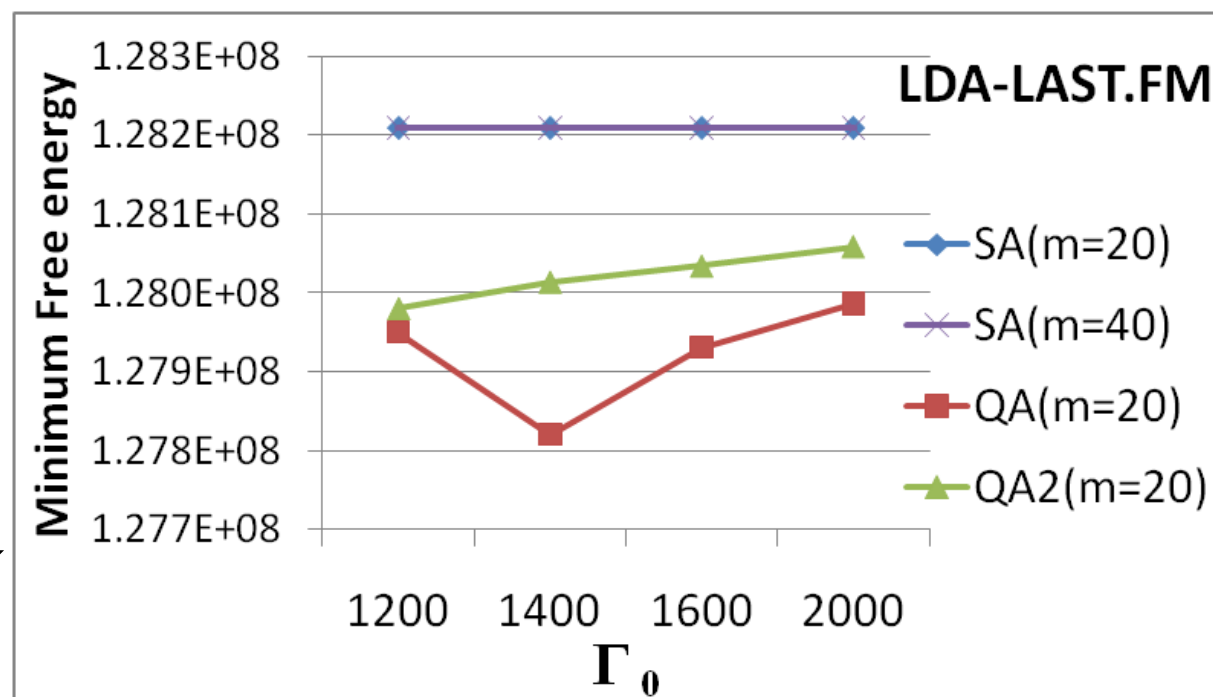
- モデル: Latent Dirichlet Allocation(LDA) $K=20$
- データセット:
 - NIPSコーパス:
 - 約1500文書, 約1万語彙,
 - LAST.FM コーパス
 - 約2400ユーザー, 1万アーティスト
- 評価方法:
 - Free energy
 - Perplexity(予測, train 70%, test 30%)

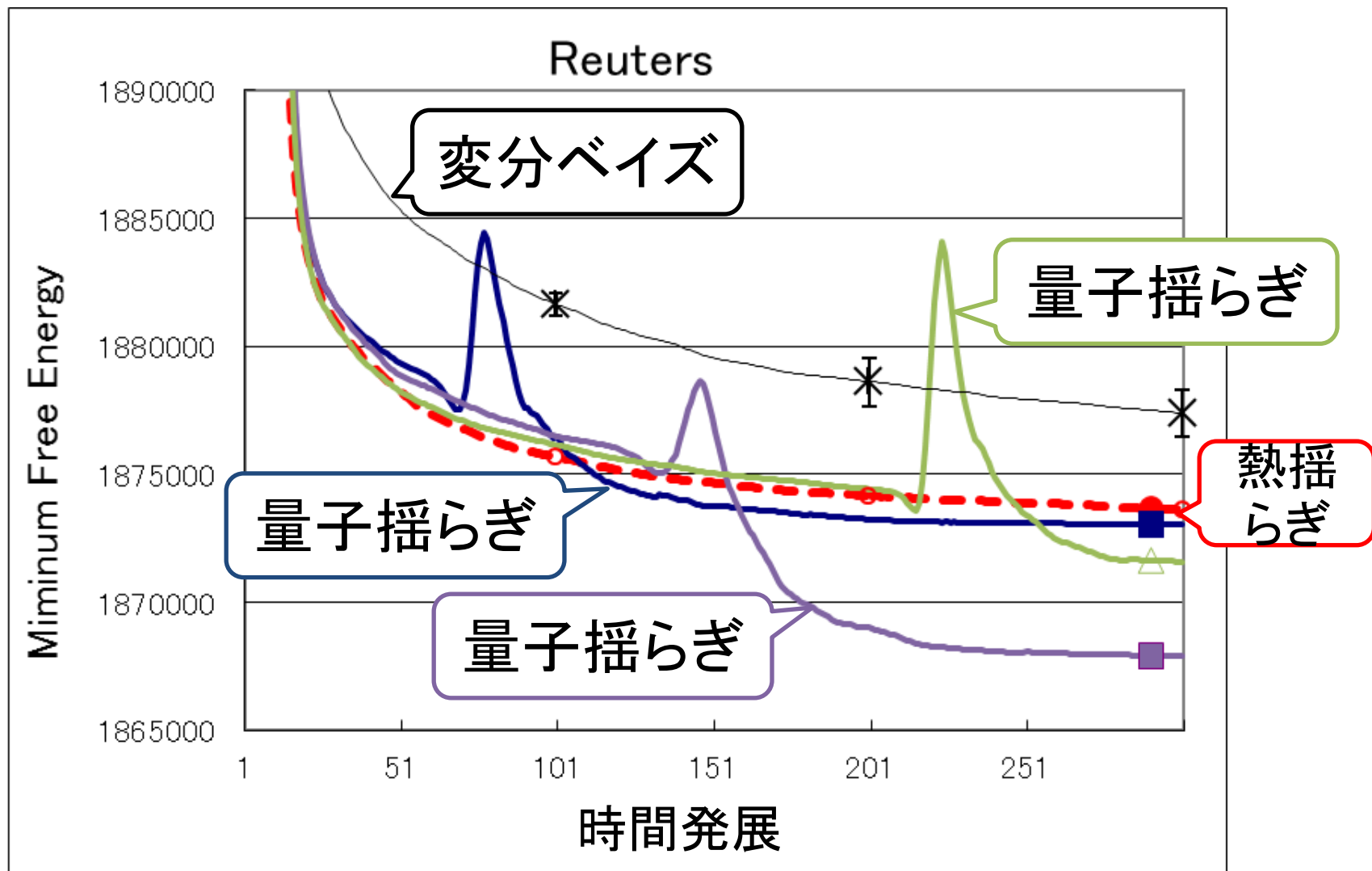
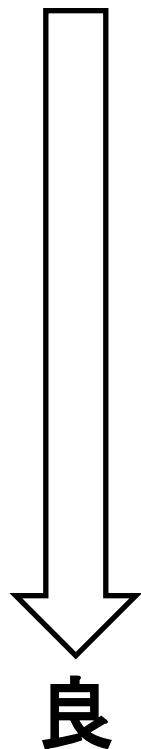


良



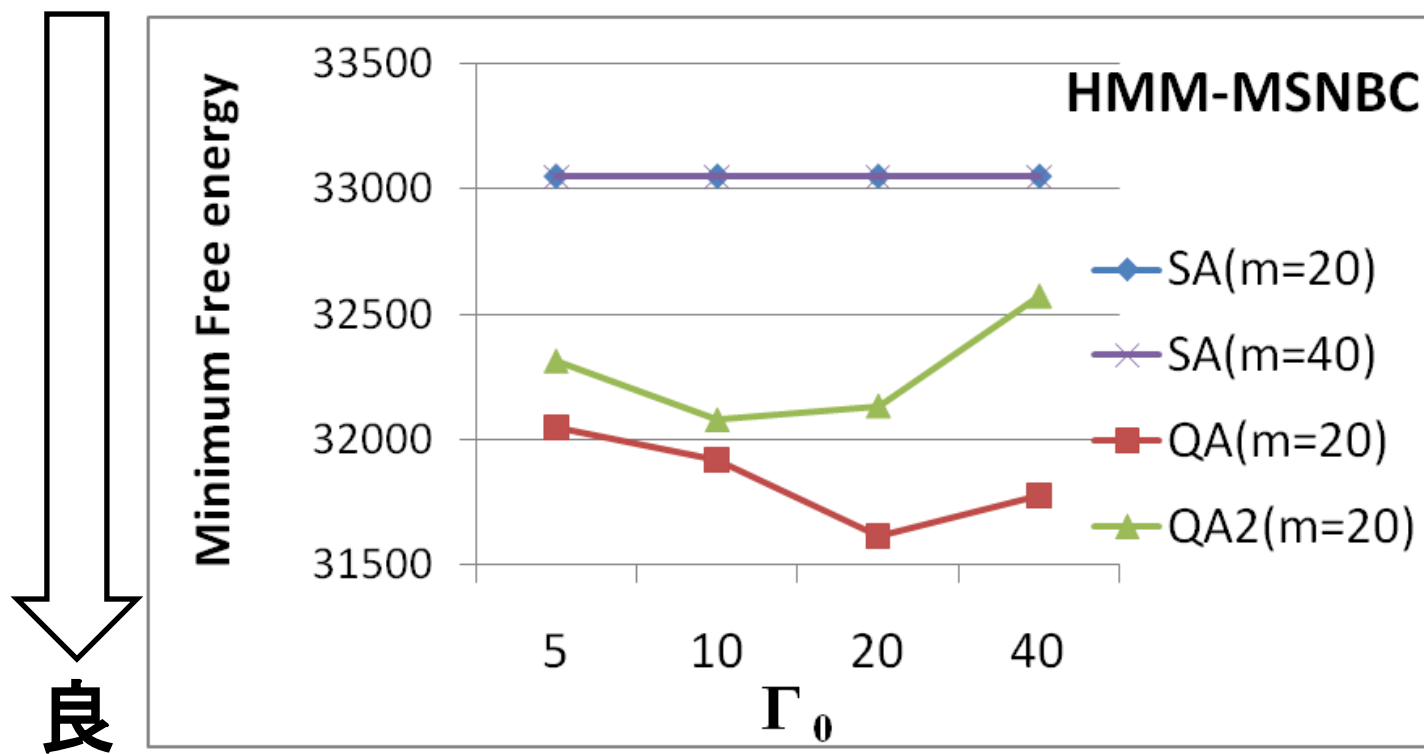
良





実験 3

- モデル: Hidden Markov model (HMM) $K=20$
- データセット: Webの閲覧履歴
 - MSNBC コーパス:
 - 2000 系列
- 評価方法:
 - Free energy
 - Test set log likelihood(予測, train 50%, test 50%)



まとめ

- 潜在変数モデル＋量子揺らぎ＋変分ベイズ法
- 変分ベイズ法で学習可能なモデルならば適用可能
- 並列計算可能

—今後の課題—

- 他の量子効果の導入方法
- 潜在状態間の重ね合わせをどのように表現するか
(=Label Identity Problem)
- データから密度行列の学習
- Quantum Latent Dynamicsとか言ってみたらどうなるか