Particle Filter による文脈の動的ベイズ推定

持橋大地 [*1,*2]　松本裕治 [*1]
{daiti-m,matsu}@is.naist.jp

[1] NAIST 情報科学研究科 松本研究室
[2] ATR 音声言語コミュニケーション研究所 音声言語処理研究室

IPSJ 2005-NL-165

2005.1.12 (Wed)

Today’s Topic

- 適切な文脈長を自動的に選択することのできるベイズ言語モデル

Today’s Topic

- 適切な文脈長を自動的に選択することのできるベイズ言語モデル
  - 文脈の変化に対する, 確率的生成モデル

Today’s Topic

- 適切な文脈長を自動的に選択することのできるベイズ言語モデル
  - 文脈の変化に対する, 確率的生成モデル
- 文脈推定問題を, フィルタリング問題と捉える

Today's Topic

- 適切な文脈長を自動的に選択することのできるベイズ言語モデル
  - 文脈の変化に対する, 確率的生成モデル
- 文脈推定問題を, フィルタリング問題と捉える
  - カルマンフィルタ等では解けない

Today’s Topic

- 適切な文脈長を自動的に選択することのできるベイズ言語モデル
  - 文脈の変化に対する, 確率的生成モデル
- 文脈推定問題を, フィルタリング問題と捉える
  - カルマンフィルタ等では解けない
    - 自然言語は非ガウシアン・非ベクトル空間

Today’s Topic

- 適切な文脈長を自動的に選択することのできるベイズ言語モデル
  - 文脈の変化に対する, 確率的生成モデル
- 文脈推定問題を, フィルタリング問題と捉える
  - カルマンフィルタ等では解けない
    - 自然言語は非ガウシアン・非ベクトル空間
  - Particle Filter (非線型ベイジアンフィルタ) を使うことで, オンラインで解くことができる

Today's Topic

- 適切な文脈長を自動的に選択することのできるベイズ言語モデル
  - 文脈の変化に対する, 確率的生成モデル
- 文脈推定問題を, フィルタリング問題と捉える
  - カルマンフィルタ等では解けない
    - 自然言語は非ガウシアン・非ベクトル空間
  - Particle Filter (非線型ベイジアンフィルタ) を使うことで, オンラインで解くことができる
    - 各粒子の持つ, 複数の文脈からの予測の混合

Today’s Topic

- 適切な文脈長を自動的に選択することのできるベイズ言語モデル
  - 文脈の変化に対する, 確率的生成モデル
- 文脈推定問題を, フィルタリング問題と捉える
  - カルマンフィルタ等では解けない
    - 自然言語は非ガウシアン・非ベクトル空間
  - Particle Filter (非線型ベイジアンフィルタ) を使うことで, オンラインで解くことができる
    - 各粒子の持つ, 複数の文脈からの予測の混合
- テキストの「意味」の変化点をオンラインで検出し, TextTiling の確率化とも捉えられる

Background

- さまざまな言語的状況に適応するために, n-gram を超える**長距離言語モデル**が重要
  - 例: 翻訳モデル $p(J|E) \propto p(J)p(E|J)$ (Brown 1990)
- しかし, これまでの研究はほぼすべて, **単一の文脈**への適応しか考えてこなかった
  - 話の流れには, サブトピックが存在する
  - 話題の変化?

    応用:
    - 実環境の音声認識, ロボティクス
    - 文脈を意識した翻訳, かな漢字変換, HMI (Human-Machine Intraction)
    - テキスト理解 (理学的目標)

Context models so far

- これまでの長距離言語モデル
  - Cache/trigger models (Old)
  - LSI 言語モデル (Bellegarda 1998)
  - PLSI 言語モデル (Gildea & Hofmann 1999)
  - LDA 言語モデル (三品, 山本 2002)
  - DM 言語モデル (山本 2003)
- テキストモデルの応用 (Bag of Words を仮定)
  - 文脈は, 単語の単純な集合
  - 文脈には, 時間的順序が**ない**
    - Bag of words「文脈」モデル

Context models so far (2)

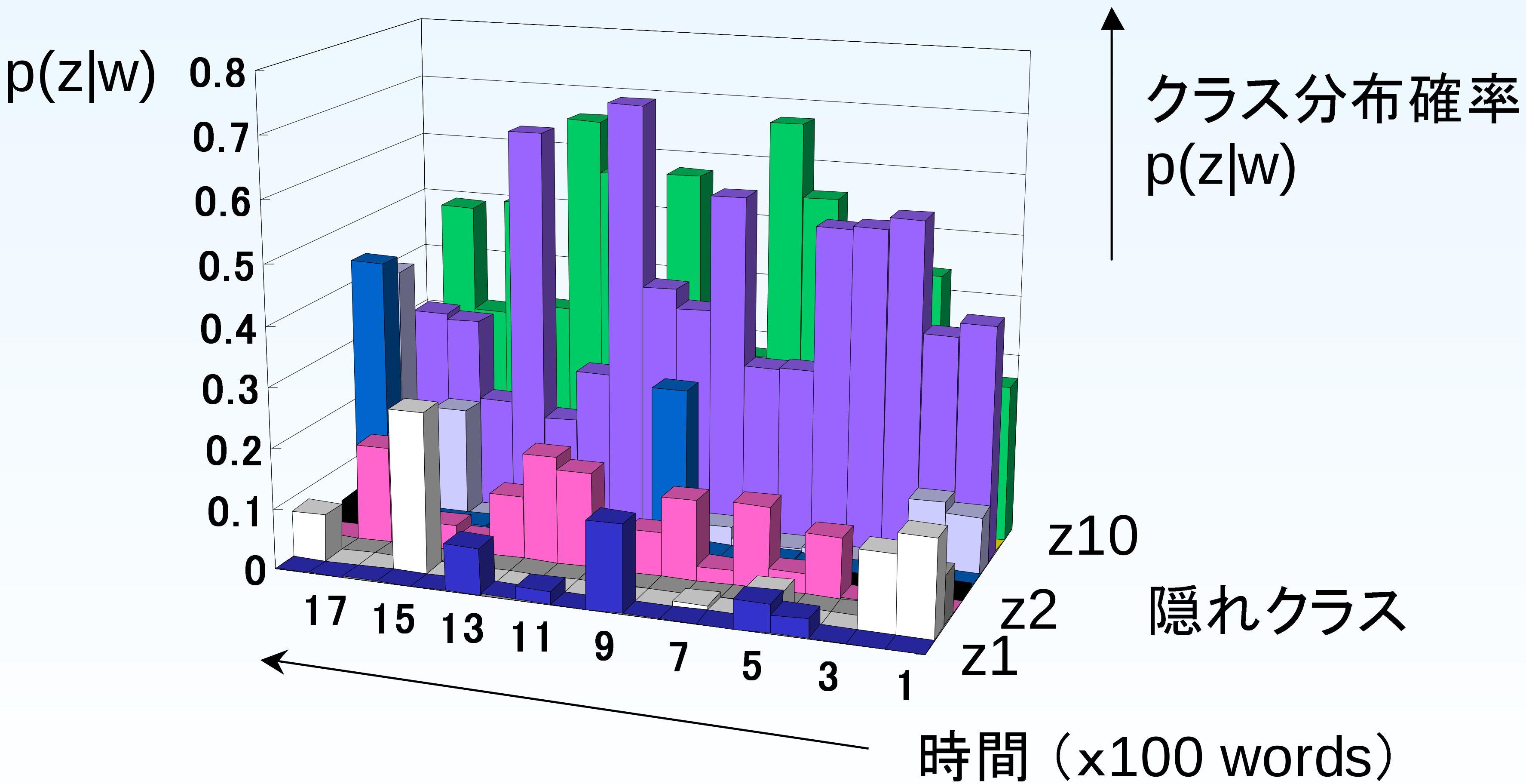
- 文脈として, テキストの先頭から全て用いている
  - or, 1000 語などの単純な閾値 (黒橋, 織 2000)
  - テキストの後ろほど推定値が悪くなる (高橋他 2003)
- テキストが長い場合はどうする?
  - 小説
  - 論説文
- 実環境への適用? (例: ロボティクス)
  - 複数の文脈の切り替え, 長時間の動作
  - 動的な適応
    - eg. かな漢字変換

⇓

これまでのモデルは, テキストに対して**定常情報源を仮定** (!)

Example of Real Text

- 毎日新聞テキスト, about 1700 words
  - PLSI で 10 クラス (図示のため) に分解



Some Exceptions

- Bellegarda (1998)
  - LSI (Latent Semantic Indexing) による文脈モデル
  - 過去の履歴を, $\lambda = 0.975$ (固定) の比率で忘れると予測が最良. ただし,
    1. 文書によって最適な λ には違いがあるのでは?
    2. 過去を一様に忘れるのでよいか?
       - 文脈の変化点からとるべき
       - 「文脈の変化点」を記述するモデル?
- Beeferman, Berger, Lafferty (1997)
  - 言語の時間的な非定常性に着目
  - テキスト中での語の意味的な関係が, 距離に関して指数的に減少
  - 指数分布としてモデル化・EM で全体のパラメータを推定

“Context as a Filtering”

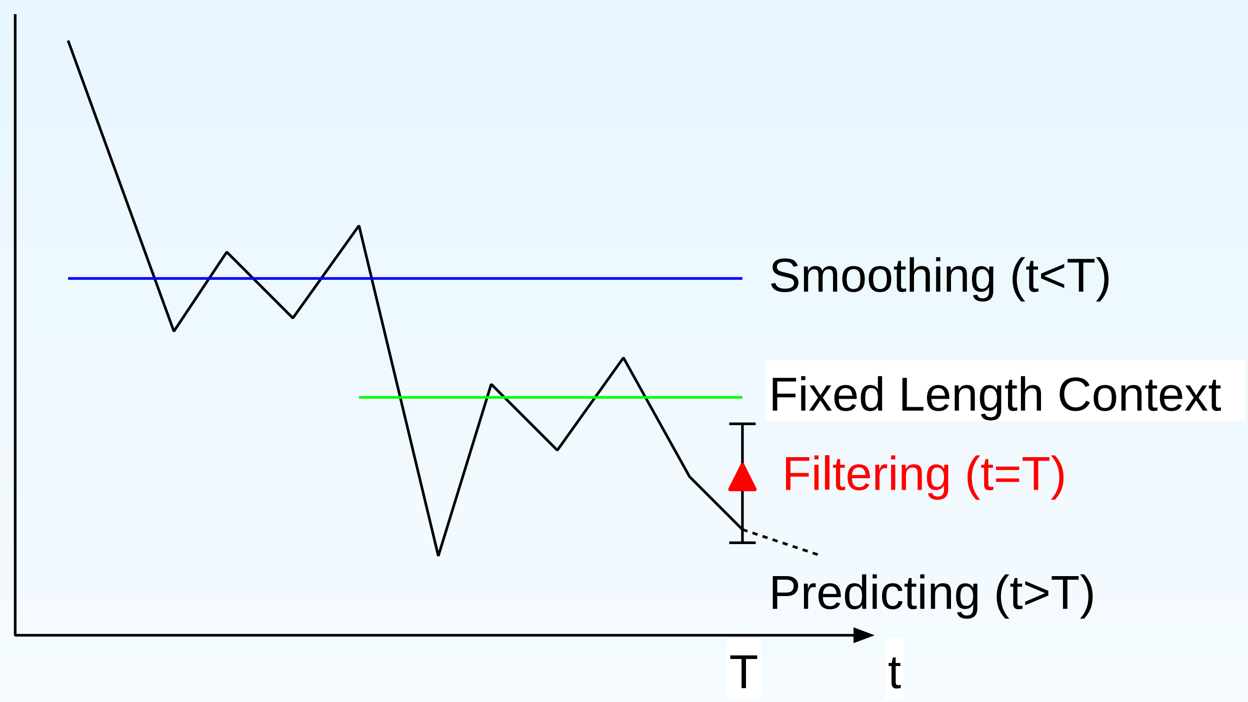
- なぜ, 文脈の遷移がモデル化されてこなかったのか?
  - 長距離文脈の問題は, 統計的には状態を逐次推定する **フィルタリング問題**と考えられる
    - カルマンフィルタ (KF), 拡張カルマンフィルタ (EKF)
    - 制御, 信号処理, ロボティクスでこれまでよく使われている
    - ベクトル空間, Gaussian Noise を仮定

  $\Downarrow$

  しかし, 自然言語は非ベクトル空間, かつ非 Gaussian
  $\cdots$ 多項分布, 多項単体 (Multinomial Simplex)

- 多項分布のフィルタリング問題
  - 難しい!

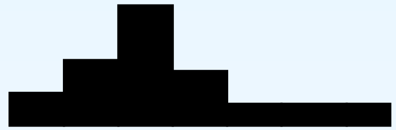
Filtering, Smoothing, Predicting ..



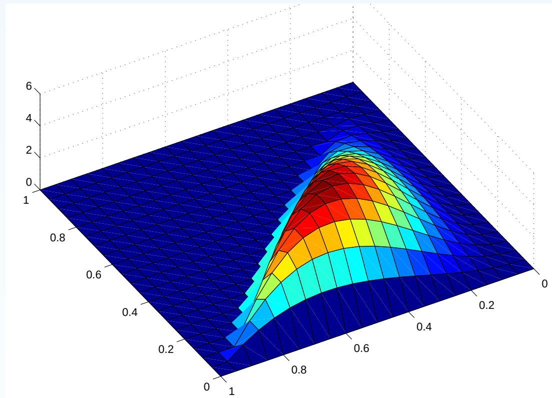
- 従来の文脈モデルは, Smoothing をしているものと考えられる
- 正確な Filtering のためには, 状態空間モデルが必要
  - 信号処理, 計量経済学 (Econometrics) などでよく使われている

Multinomial and Dirichlet distribution

- 多項分布 $\mathrm{Mult}(\boldsymbol{\theta})$ $\cdots$ 離散シンボルの生起確率
  - $\boldsymbol{\theta} = (\theta_1 \theta_2 \ldots \theta_K)$
 $\theta_1 + \theta_2 + \cdots + \theta_K = 1.$
  - 例: n-gram 確率, 混合モデルの混合比



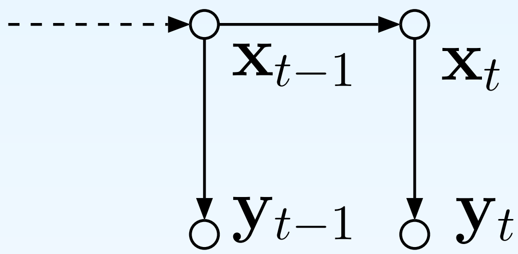
- ディリクレ分布 $\mathrm{Dir}(\boldsymbol{\alpha})$ $\cdots$ 多項分布の分布
  - $\mathrm{Dir}(\boldsymbol{\theta}|\boldsymbol{\alpha}) = \dfrac{\Gamma(\sum_k \alpha_k)}{\prod_{k=1}^K \Gamma(\alpha_k)} \displaystyle\prod_{k=1}^K \theta_k^{\alpha_k - 1}$
  - Simplex (単体) 上の確率分布
  - 多項分布を推定するには,
 多項分布の分布であるディリクレ分布が
 必要.



State Space Models

- 状態空間モデル (State space models) $=$ 一般化 HMM

$$
\begin{cases}
\mathbf{x}_t \sim p(\mathbf{x}_t|\mathbf{x}_{t-1}) & (\text{状態遷移方程式}) \\
\mathbf{y}_t \sim p(\mathbf{y}_t|\mathbf{x}_t) & (\text{観測方程式})
\end{cases}
$$



- 状態空間がベクトル空間&正規分布の場合
  - 最小二乗の厳密解が存在 (Kalman Filter) (Kalman 1960)
- 状態空間が離散変数のどれか一つである場合
  - 離散隠れマルコフモデル
  - Baum-Welch 法 (転送行列法) が厳密解
- それ以外…解が簡単には求められない
  - 言語モデル, 混合モデル…真の状態が一つでなく, 多項分布
  - データが少ない場合 (ex. 言語の文脈)

Multinomial State Space Model

- 文脈に, 隠れた多項分布 θ_t が存在するとする. θ_t とは,
  - 単語の出現確率 (ユニグラム)
  - 隠れたトピックの混合比
    - 例. “スポーツ” = 0.2, “政治” = 0.4, $\ldots$ etc.
    - 「トピック」はコーパスから (VB)-EM で求まる
- このとき, θ に次の状態空間モデルを仮定する (Mean shift model: MSM)

$$\begin{cases} \theta_t \sim \begin{cases} \mathrm{Dir}(\boldsymbol{\alpha}) & \text{with probability } \rho \\ \theta_{t-1} & \text{with probability } (1-\rho) \end{cases} \\ y_t \sim \mathrm{Mult}(\theta_t) \end{cases} \tag{1}$$

  - 確率 ρ で隠れた多項分布 θ_t が変化
  - θ_t は未知, いつ変化したのかも未知
  - 観測されるのは, 出力 $\mathbf{y} = (y_1 y_2 \ldots y_T)$ のみ.

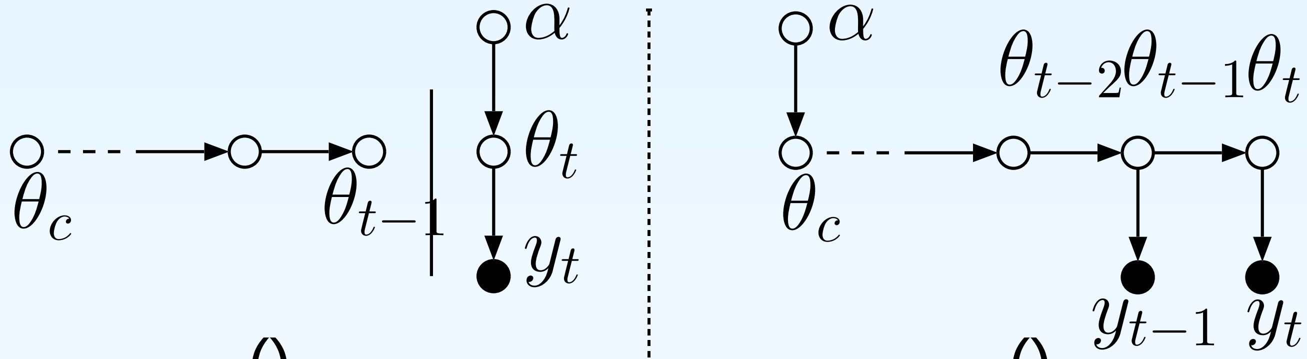
Multinomial Mean Shift Model (1)

次のようなアルファベット a,b,c の系列が観測されたとする.

aaaaaabaacbaabaaaaabbbbabababaabababbab\
bbbabcacccbcacacccccccccccccacccacccc\
ccccccccacaaaacbbbbb

- 次のアルファベットは何?
  - b の可能性が高い (しかし, オーバーフィットの危険)
  - 全部平均してしまう (今までの文脈モデル!)
- 最近変化した場所に依存
  - 変化点検出問題 (Change-point problem) として知られた問題
  - もし最近の変化点がわかれば, 推定値は正確に求まる. (次スライド)

Multinomial Mean Shift Model (2)



() ()

最近の変化点の位置によって, 推定値は

(a) の場合:

$$p(y_t|y_c \cdots y_{t-1}) = \int p(y_t|\theta_t)p(\theta_t|\alpha)d\theta_t = \alpha_y / \textstyle\sum_y \alpha_y \qquad (2)$$

(b) の場合:

$$p(y_t|y_c \cdots y_{t-1}) = \int p(y_t|\theta_t)p(\theta_t|y_c \cdots y_{t-1})d\theta_t \qquad (3)$$

$$= \frac{\alpha_y + \sum_{t=c}^{t-1} \delta(y)}{\alpha + \sum_{t=c}^{t-1} \delta(y_t)} \qquad (4)$$

のように求まる.

How to detect a change point?

- ベイズの公式により, 変化確率は計算できる (Chen and Lai, 2003; Yao, 1984)

$$p(change|\mathbf{observed}) \tag{5}$$

$$\propto p(\mathbf{observed}|change)\,p(change) \tag{6}$$

$$= \begin{cases} p(\mathbf{observed}|change=1)p(change=1) \\ p(\mathbf{observed}|change=0)p(change=0) \end{cases} \tag{7}$$

- $p(\mathbf{observed}|change)$ は, 今求めたばかり
- $p(change) = \rho$ ととりあえずおく [文脈変化の事前確率]
  - 後で動的に推定
- $p(change|\mathbf{observed})$ が計算できる!

How to detect a change point? (2)

- しかし..
  - 変化点は, 計算上確定されなければならない
    - 「履歴を 0.214 だけ用いる」ようなことはできない.
  - ベルヌーイ試行でサンプル ··· 推定が不安定 (一種の弱学習器)

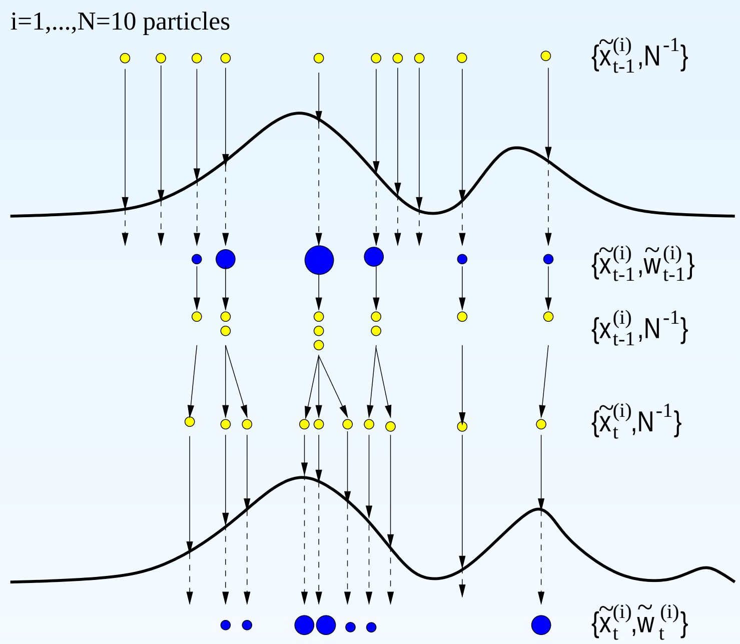
↓

- Particle Filter (逐次モンテカルロ法) を使う
  - 複数の Particle (粒子, モンテカルロサンプル) による予測を同時に動かす
  - 混合分布として予測
  - 文脈変化の事前確率も動的に推定できる.
    - 話題の移り変わりの速さ, 遅さ

Particle Filter (1)

- a.k.a. 逐次 (時系列) モンテカルロ法
  - (Kitagawa 1987; Doucet, de Freitas, Gordon 2001)
- 複数のモンテカルロサンプル (Particle) による重み付き推定
  - 信号処理, ロボティクスなどの分野で最近使われ始めている
    - 計算資源の急速な増大
- 任意の非線型ダイナミクスを追跡できる (!)
  - 非ガウシアンダイナミクス, 外れ値推定
  - 自然言語のような離散データにも, 原理的に適用可能
- **複数の変化点履歴を持つ予測を, 同時に動かす**

Particle Filter (2)



- 時間 t での N 個の particle $p_t^{(1)} \cdots p_t^{(N)}$ →**重み** $w_t^{(1)} \cdots w_t^{(N)}$
- 各 particle は, 異なった変化点履歴を確率的にサンプル
- 重み付き期待値: $E[f(\mathbf{x}_t)] \simeq \sum_{i=1}^{N} w_t^{(i)} f(p_t^{(i)}(\mathbf{x}_t))$

Particle Filter (3)

- 粒子 i の重み $w_t^{(i)}$ のアップデート (和が 1 になるように正規化)

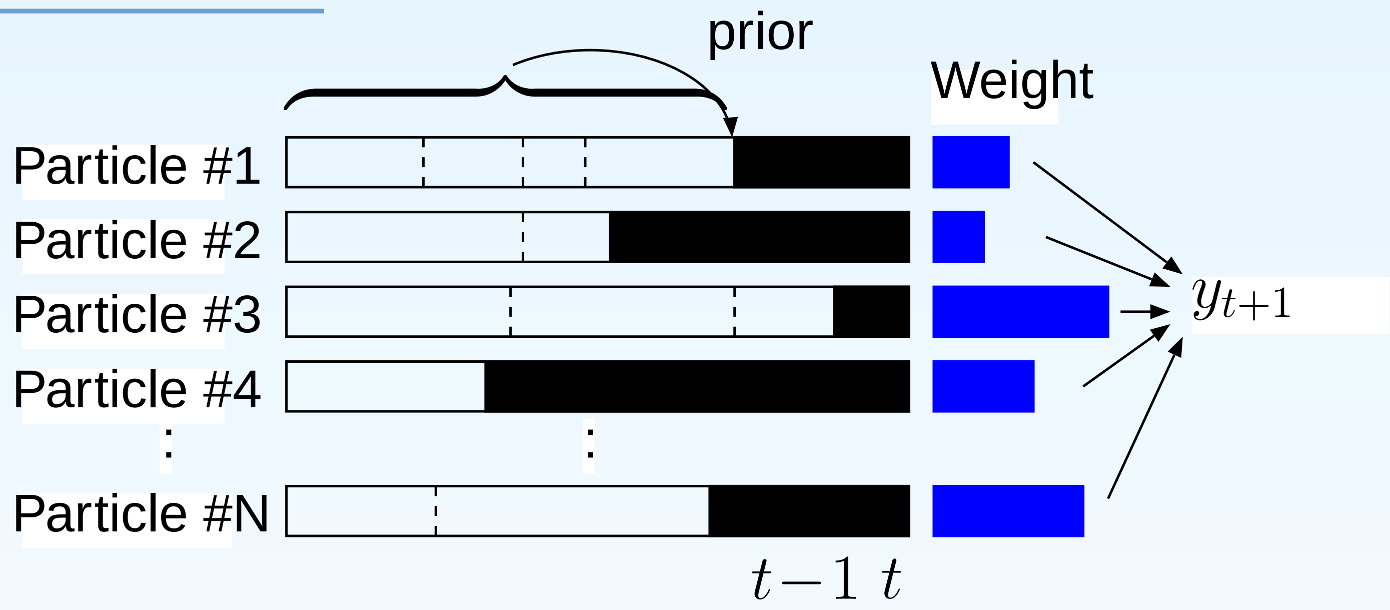
$$w_t^{(i)} \propto w_{t-1}^{(i)} \cdot \frac{p(y_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{x}_{t-1})}{q(\mathbf{x}_t|X_{t-1}, Y_t)} \tag{8}$$

- われわれの問題の場合, 実際には

$$\begin{aligned} w_t^{(i)} \propto w_{t-1}^{(i)} \cdot (p(y_t|\mathit{change} = 0)p(\mathit{change} = 0) + \\ p(y_t|\mathit{change} = 1)p(\mathit{change} = 1)) \end{aligned} \tag{9}$$

  で更新できる [予稿集参照]
  - ただし, 時間がたつと, 重み $w_t^{(1)} \cdots w_t^{(N)}$ に大きな偏りが生じる (「Particle がすり減る」) → **リサンプリング**
    - 重みの大きい粒子の「子供」を作る
    - 重みの少ない粒子を殺す.

Particle Filter (4)



- 各粒子は, 異なった変化点履歴 $\boldsymbol{I}_t^{(i)}$ をもつ
- 複数の粒子のもつ, 様々な文脈長からの予測の重みつき推定 (混合分布)
- 文脈の変化速度を「正しく」推定したサンプルが生き残る
  - 間違ってサンプルした粒子は, リサンプリングで消える
  - 文脈の変化速度を, 時間に従って正しく推定する

Multinomial Particle Filtering

```
aaaaaabaacbaabaaaaabbbbbababababaabababbab\
bbbabcacccbcacacccccccccccccccacccccacccc\
cccccccccacaaaacbbbbb..
```

a = blue
b = green
c = red

- 水平線 = 隠れた正解 (多項分布)
- # of Particles = 50, $(\alpha, \beta) = (1, 10)$

Mean shift model of Natural Language

- しかし…
  - 自然言語では, アルファベット (単語) の数が膨大
    - 「病院」→「看護婦」では, 隠れた多項分布は不変
    - 「病院」→「大学」では, 文脈が変化している

  ↓

  単語の間の意味的な**相関**を捉える必要.
- Dirichlet Mixture (DM) (山本他 2003),
  Latent Dirichlet Allocation (LDA) (Blei et al. 2001) を用いて,
  MSM を自然言語に拡張した.

MSM-DM (1)

- Dirichlet Mixture (DM) (山本他 2003, 2004) @SLP-48,53
  - 文脈推定のための確率的テキストモデル
    - これまで, 文脈モデルとして性能が最高
  - 単語の意味的な相関を, ハイパーパラメータのレベルで捉える

$$p(\mathbf{w}|\boldsymbol{\lambda},\boldsymbol{\alpha}) = \int_{\triangle} p(\mathbf{w}|\boldsymbol{p})p(\boldsymbol{p}|\boldsymbol{\lambda},\boldsymbol{\alpha})d\boldsymbol{p} \tag{10}$$

$$= \sum_{m=1}^{M} \lambda_m \frac{\Gamma(\sum_v \alpha_{mv})}{\Gamma(\sum_v \alpha_{mv}+n)} \prod_{v\in\mathbf{w}} \frac{\Gamma(\alpha_{mv}+n(v))}{\Gamma(\alpha_{mv})} \tag{11}$$

($\boldsymbol{p}$: ユニグラム確率; $\boldsymbol{\lambda},\boldsymbol{\alpha}$: DM パラメータ)

- パラメータ $\boldsymbol{\lambda}$, $\boldsymbol{\alpha}$ は EM 法と Newton 法を組み合わせてコーパスから求める
  - http://cl.naist.jp/~daiti-m/dist/dm/ でパッケージを公開中
  - 56,000 文書, 100 クラスへの分解で 10 時間程度 (Xeon 2.8GHz)

MSM-DM (2)

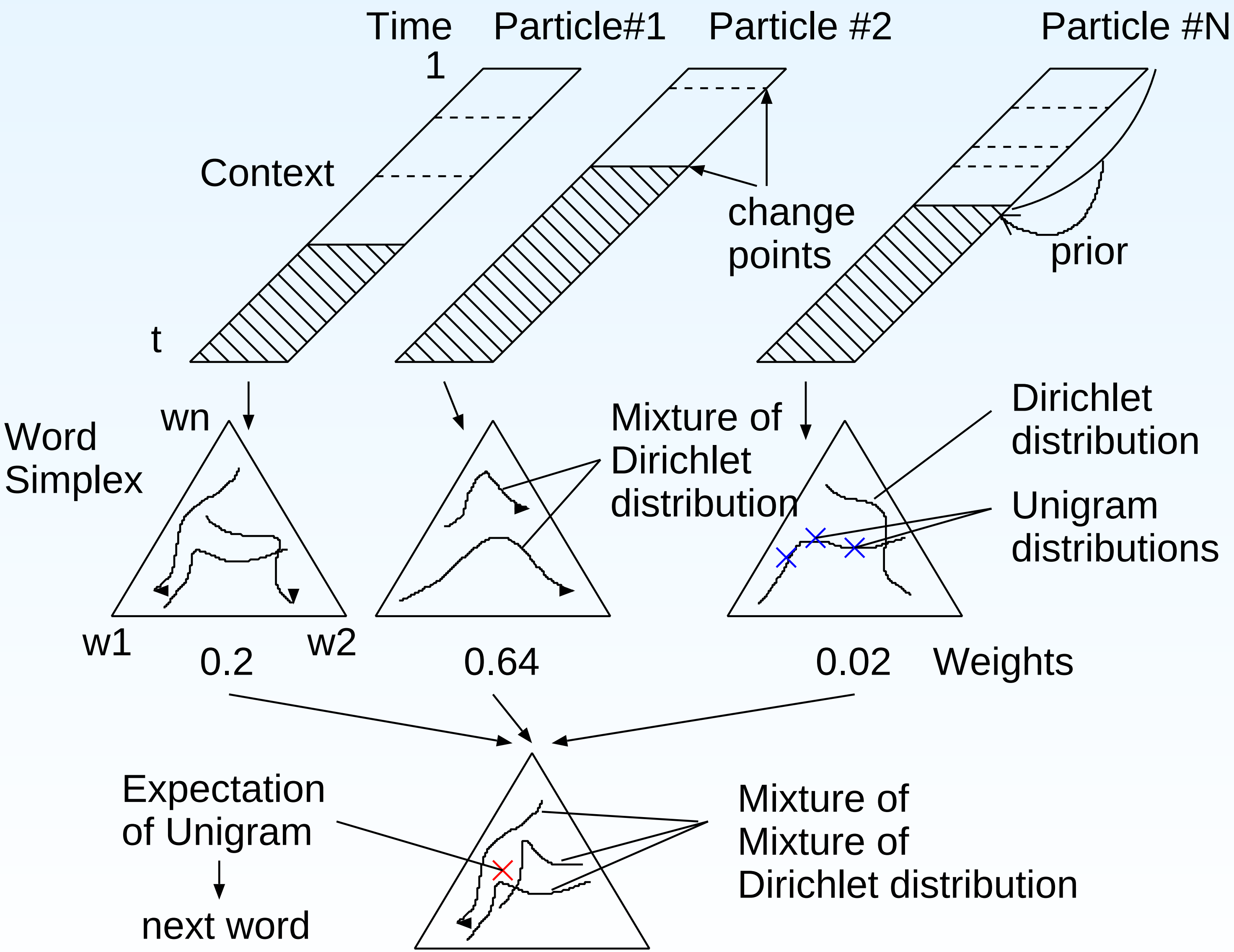
- 文脈 $\boldsymbol{h} = w_1 w_2 \dots w_h$ のもとでの予測確率:

$$p(y|\boldsymbol{h}, \boldsymbol{\alpha}, \boldsymbol{\lambda}) \propto \sum_{m=1}^{M} C_m(\boldsymbol{h}) \frac{\alpha_{my} + n(y)}{\alpha_m + h} \quad (12)$$

($n(y)$: $\boldsymbol{h}$ 中の y の生起回数)

  - 複数のハイパーパラメータ $\boldsymbol{\alpha}_1 \dots \boldsymbol{\alpha}_M$ からの予測の混合
  - ハイパーパラメータ自体を適応的に選択
    - MacKay (1994) の Dirichlet smoothing の拡張
- $\boldsymbol{h}$ を最近の変化点からの履歴とし, この予測確率を多項分布 MSM で用いる
  - もともとの MSM は, 事前分布にディリクレ分布を仮定
  - DM では, 混合ディリクレ分布を仮定
    - 多項分布 MSM の自然な拡張になっている.

Graphical image of MSM-DM



MSM-LDA (1)

- Latent Dirichlet Allocation (LDA) (Blei et al. 2001, 2003)
  - 確率的トピックのテキストモデル
  - 単語の間の相関を, 隠れた確率的**トピック**のレベルで捉える
  - PLSI (Hofmann 1999) のベイズ拡張

$$p(\mathbf{w}|\boldsymbol{\alpha},\boldsymbol{\beta}) = \frac{\Gamma(\sum_{k=1}^{K}\alpha_k)}{\prod_{k=1}^{K}\Gamma(\alpha_k)} \int \left(\prod_{k=1}^{K}\theta_k^{\alpha_k-1}\right)\left(\prod_{n=1}^{N}\sum_{k=1}^{K}\prod_{v=1}^{V}(\theta_k\beta_{kv})^{w_n^v}\right) d\boldsymbol{\theta} \tag{13}$$

- パラメータ:
  - $\boldsymbol{\alpha} = \mathrm{Dir}(\boldsymbol{\lambda}|\boldsymbol{\alpha})$ 隠れトピックへの事前分布パラメータ
  - $\boldsymbol{\beta} = \{p(v|m)\}\ (v = 1..V, m = 1..M)$ トピック別ユニグラム確率行列
  - http://cl.naist.jp/~daiti-m/dist/lda/ でパラメータ推定を行うパッケージを公開している

MSM-LDA (2)

- 文脈 $\boldsymbol{h} = w_1 w_2 \ldots w_h$ の持つ, 潜在的な M 次元の **トピック分布** $q(\boldsymbol{\lambda}|\boldsymbol{h})$ は, 次の変分ベイズ EM 法で求まる

$$\begin{cases} q(z_i^t = 1|\boldsymbol{h}) \propto p(w_i|t)\exp(\Psi(\alpha + n_t)) & \textbf{(VB-E step)} \\ q(\boldsymbol{\lambda}|\boldsymbol{h}) \propto \prod_{t=1}^{M} \lambda_t^{\alpha+n_t-1} & \textbf{(VB-M step)} \end{cases} \tag{14}$$

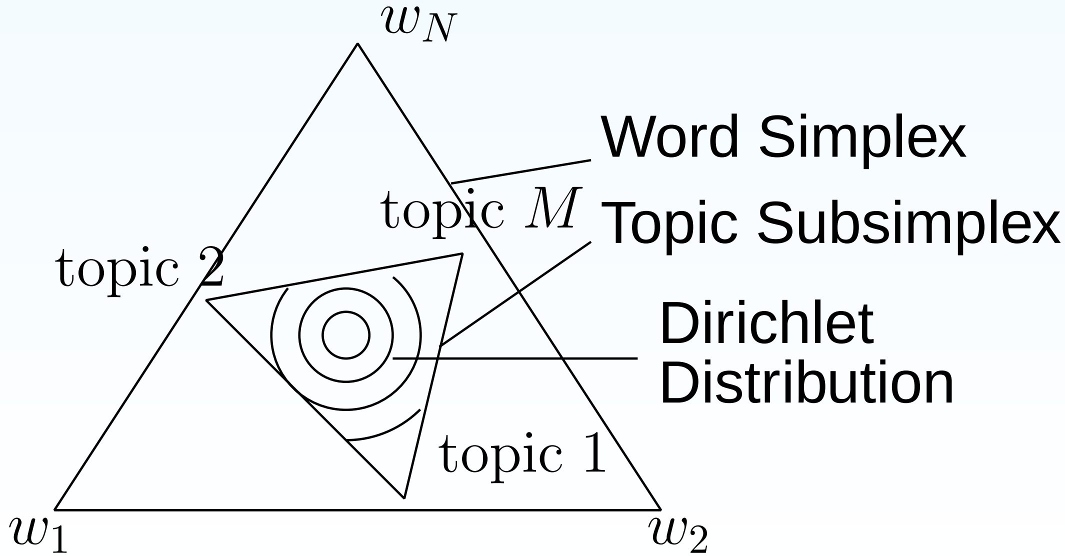
$$\text{where } n_t = \sum_{i=1}^{h} q(z_i^t = 1|\boldsymbol{h})$$

- 文脈 $\boldsymbol{h}$ のもとでの予測確率 = 各トピックからの予測の混合:

$$p(y|\boldsymbol{h}) = \sum_{m=1}^{M} p(y|m) E_q[\lambda_m|\boldsymbol{h}] \tag{15}$$

MSM-LDA (3)

- ユニグラム分布を直接ではなく, M 次元トピックの**混合比**の変化を追跡する
  - 直感的な解釈が容易
  - トピック空間へ圧縮して追跡するため, やや柔軟性に欠ける
    - Topic simplex は, Word simplex の subsimplex
  - 一般の混合モデルの混合比追跡に応用可能.
    - eg. Gaussian Mixtures



Experiments on BNC

- British National Corpus (BNC)
  - 約 1 億語のバランスドコーパス
  - そのうち 56,939 文書, 11,032,233 語を LDA/DM のパラメータ推定に使用
  - 仮想的にファイルを分割して文書とした
    - LDA などのテキストモデルは, 長い文書には無力 (!)
    - 本研究の手法の, 文書集合への拡張 (Future work)
- サンプリングにより, 4 種類の文脈変化速度の評価セットを作成した
  - 1 セット $\equiv$ 100 文書
  - 1 文書はサンプリングされた 100 文からなる (約 2,000 語)
    - Raw : 連続した 100 文
    - Slow : RAW より速い変化
    - Fast : SLOW より速い変化
    - VeryFast : FAST より速い変化

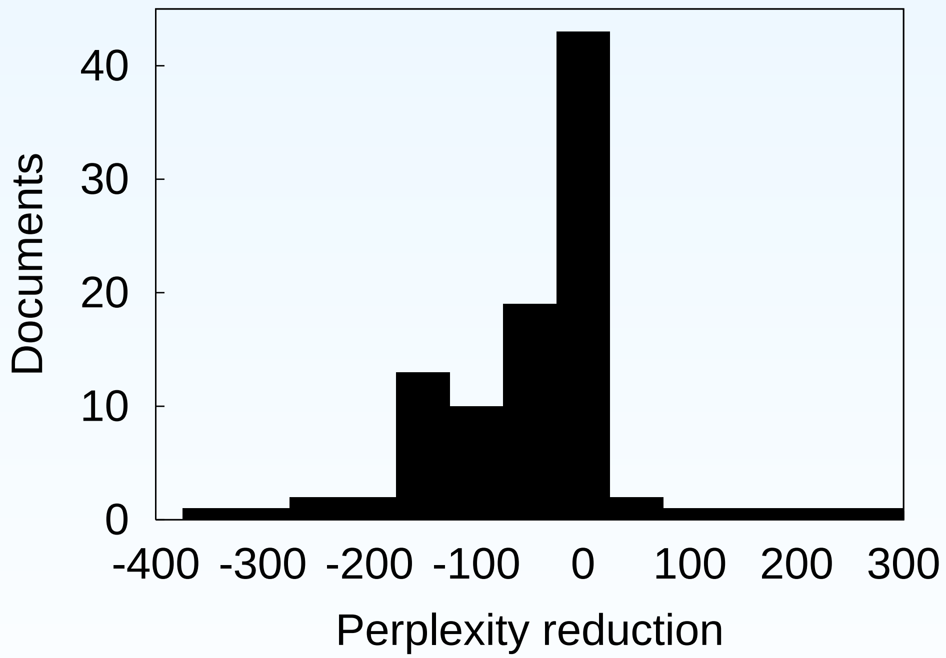
Experiments on BNC (2)

Text	MSM-DM	DM	MSM-LDA	LDA
Raw	870.06 (-6.02%)	925.83	1028.04	1037.42
Slow	893.06 (-8.31%)	974.04	1047.08	1060.56
Fast	898.34 (-9.10%)	988.26	1044.56	1061.01
VFast	960.26 (-7.57%)	1038.89	1065.15	1050.83

- MSM-DM の方が直接 Word Simplex 上で動くため, 追跡が柔軟
- MSM-LDA も文脈を自動選択する効果はあるが, トピック空間に圧縮するため, やや柔軟性に欠ける

Experiments on BNC (3)

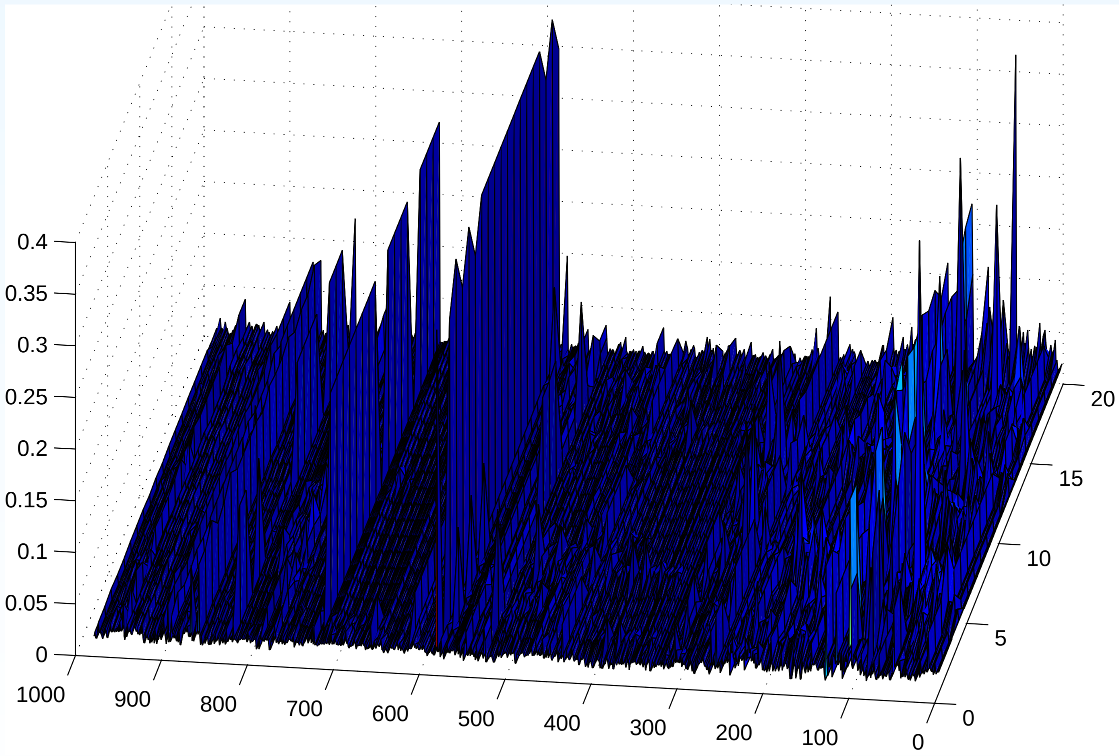
- Dirichlet Mixture に対する, 各文書のパープレキシティ減少
  - “Raw” データの場合 (= 生テキスト)
  - $\mathrm{PPL}_{MSM} - \mathrm{PPL}_{DM}$



- ほとんどの文書で効果がある
  - 75% の文書でパープレキシティ減少 (最大の減少 $= -365$)

Change probabilities of real text

- 本研究は, 単語ごとに文脈変化確率 $p(I_t = 1)$ を計算できるため, TextTiling (Hearst 1994) の確率化ともとらえられる
- 実際のテキストの変化確率:



Conclusion

- 文脈の変化確率を計算し, 適切な文脈長を自動的に選択するベイズ長距離言語モデル

Conclusion

- 文脈の変化確率を計算し, 適切な文脈長を自動的に選択するベイズ長距離言語モデル
  - 複数の文脈長からの予測の重みつき混合

Conclusion

- 文脈の変化確率を計算し, 適切な文脈長を自動的に選択するベイズ長距離言語モデル
  - 複数の文脈長からの予測の重みつき混合
  - 文脈の変化に対する, 確率的生成モデル

Conclusion

- 文脈の変化確率を計算し, 適切な文脈長を自動的に選択するベイズ長距離言語モデル
  - 複数の文脈長からの予測の重みつき混合
  - 文脈の変化に対する, 確率的生成モデル
- Mean shift model の DM, LDA による自然言語への拡張

Conclusion

- 文脈の変化確率を計算し, 適切な文脈長を自動的に選択するベイズ長距離言語モデル
    - 複数の文脈長からの予測の重みつき混合
    - 文脈の変化に対する, 確率的生成モデル
- Mean shift model の DM, LDA による自然言語への拡張
- 非線型な HMM · · · 非線型モンテカルロフィルタである Particle Filter を用いた

Conclusion

- 文脈の変化確率を計算し, 適切な文脈長を自動的に選択するベイズ長距離言語モデル
  - 複数の文脈長からの予測の重みつき混合
  - 文脈の変化に対する, 確率的生成モデル
- Mean shift model の DM, LDA による自然言語への拡張
- 非線型な HMM $\cdots$ 非線型モンテカルロフィルタである Particle Filter を用いた
  - 文脈の変化速度もオンラインで推定できる

Conclusion

- 文脈の変化確率を計算し, 適切な文脈長を自動的に選択するベイズ長距離言語モデル
  - 複数の文脈長からの予測の重みつき混合
  - 文脈の変化に対する, 確率的生成モデル
- Mean shift model の DM, LDA による自然言語への拡張
- 非線型な HMM · · · 非線型モンテカルロフィルタである Particle Filter を用いた
  - 文脈の変化速度もオンラインで推定できる
  - 補助的に, TextTiling の確率化

Conclusion

- 文脈の変化確率を計算し, 適切な文脈長を自動的に選択するベイズ長距離言語モデル
  - 複数の文脈長からの予測の重みつき混合
  - 文脈の変化に対する, 確率的生成モデル
- Mean shift model の DM, LDA による自然言語への拡張
- 非線型な HMM · · · 非線型モンテカルロフィルタである Particle Filter を用いた
  - 文脈の変化速度もオンラインで推定できる
  - 補助的に, TextTiling の確率化
  - 現在, 最もパープレキシティの低い長距離言語モデル

Future Work

- 本研究は言語モデルのため, Forward モデル (予測モデル)

Future Work

- 本研究は言語モデルのため, Forward モデル (予測モデル)
  - Forward-Backward への拡張

Future Work

- 本研究は言語モデルのため, Forward モデル (予測モデル)
  - Forward-Backward への拡張
    - “Forward Filtering, Backward Sampling”
    - Expectation-Propagation (Minka 2001)?

Future Work

- 本研究は言語モデルのため, Forward モデル (予測モデル)
  - Forward-Backward への拡張
    - “Forward Filtering, Backward Sampling”
    - Expectation-Propagation (Minka 2001)?
- 文・パラグラフ単位の変化点

Future Work

- 本研究は言語モデルのため, Forward モデル (予測モデル)
  - Forward-Backward への拡張
    - “Forward Filtering, Backward Sampling”
    - Expectation-Propagation (Minka 2001)?
- 文・パラグラフ単位の変化点
- 文書集合への拡張

Future Work

- 本研究は言語モデルのため, Forward モデル (予測モデル)
  - Forward-Backward への拡張
    - “Forward Filtering, Backward Sampling”
    - Expectation-Propagation (Minka 2001)?
- 文・パラグラフ単位の変化点
- 文書集合への拡張
  - “Document” というナイーブな単位の克服

Future Work

- 本研究は言語モデルのため, Forward モデル (予測モデル)
  - Forward-Backward への拡張
    - “Forward Filtering, Backward Sampling”
    - Expectation-Propagation (Minka 2001)?
- 文・パラグラフ単位の変化点
- 文書集合への拡張
  - “Document” というナイーブな単位の克服
  - HM-DM (Hidden Markov DM)
  - HM-LDA (Hidden Markov LDA)

Thank you

Thank you very much for listening.

Roadmap to MSM-DM

- Unigram

$\Downarrow$

- Dirichlet Mixture of Unigrams
  - ベイズ推定 (MacKay 1994)
  - ディリクレ分布に従った, 無数の Unigram の混合

$\Downarrow$

- Mixture of “Dirichlet Mixture of Unigrams”
  - Dirichlet Mixture (山本他 2003)
  - 複数のディリクレ分布を用いた, 上の Unigram 混合の混合

$\Downarrow$

- Mixture of “Mixture of Dirichlet Mixture of Unigrams”
  - MSM-DM (本発表)
  - 複数の文脈長を自動的に用いた, Dirichlet Mixture の混合