課題番号 jh130050-DA03

確率的潜在変数モデルの大規模学習アルゴリズム開発

佐藤 一誠（東京大学）

概要

潜在変数と呼ばれる確率変数を導入することで、データ中に隠れた情報を抽出する確率的潜在変数モデルは、近年のデータ解析の主要素技術となっている。本研究では、特に研究が盛んな確率的潜在変数モデルの１分野である潜在トピックモデルの学習アルゴリズムの開発を行う。潜在トピックモデルの学習には、決定的アルゴリズムと確率的アルゴリズムが存在し、これまで確率的アルゴリズムの汎化性能の良さが知られていた。しかし、近年、周辺化変分ベイズ法と呼ばれる決定的アルゴリズムが開発され、確率的アルゴリズムと同程度の性能が出ることが知られている。本研究では、より広範囲な応用を考えて、周辺化変分ベイズ法を大規模学習アルゴリズムへと拡張することを目的とする。周辺化変分ベイズ法の問題点は、使用メモリ量が非常に大きいことと原理的に並列計算ができないことであり、この２点の問題を解決する。

1. 研究の目的と意義

潜在変数と呼ばれる確率変数を導入することで、データ中に隠れた情報を抽出する確率的潜在変数モデルの研究がデータ解析において幅広く用いられている。　これまでの確率的潜在変数モデルの研究では、データを解析するためのモデリングの研究が主なテーマとなっていたが、近年のデータの増加にともない，学習の処理時間に関する課題が多く残されているため、大規模学習アルゴリズム開発が１つの重要なテーマとなっている。

確率的潜在変数モデルの学習は、非観測の潜在変数をデータから推定する教師なし学習である。例えば、教師情報を与えずに、文書をいくつかの潜在トピックに分類する場合を考える(図１参照)。潜在トピックモデルは、文書中に内在する隠れたトピック情報を潜在変数として文書の生成過程を学習することで、文書を自動的に分類する。

最先端の確率的潜在変数モデルの研究では、潜在トピックモデル及びその拡張手法としてノンパラメトリックベイズモデルに基づくモデリング手法が注目を集めている。ノンパラメトリックベイズモデルは、潜在トピック数などのモデル構造をデータから自動決定することができる。したがって、データ解析を行う側でのパラメータチューニングに関する専門的な必要とせず、解析を複数回行う必要もないため解析時間の短縮にも役立つ。

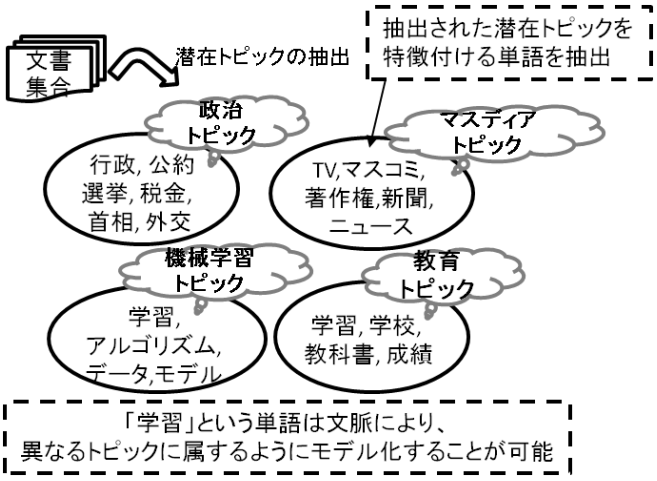

図 1 文書集合からの潜在トピック抽出の例

現在、潜在トピックモデルは文書データを対象とした自然言語処理への応用をきっかけとして、そのモデリングの柔軟性からWebマイニング、評判分析、推薦システム、画像処理、音声・音響処理、バイオインフォマティックス、地理情報解析、ネットワーク構造解析などさまざまな分野で用いられている。例えば、ユーザの購買データに関して潜在トピックの抽出はユーザの隠れた嗜好パターンをトピックとして抽出することに対応する。ネットワーク構造解析では、潜在コミュニティ抽出に対応する。したがって、潜在的トピックモデルの学習効率の向上は周辺領域へと波及するため学際的な意義があると考えられる。

より広範囲な応用研究への展開を考えると、シングルプロセスによる学習の処理速

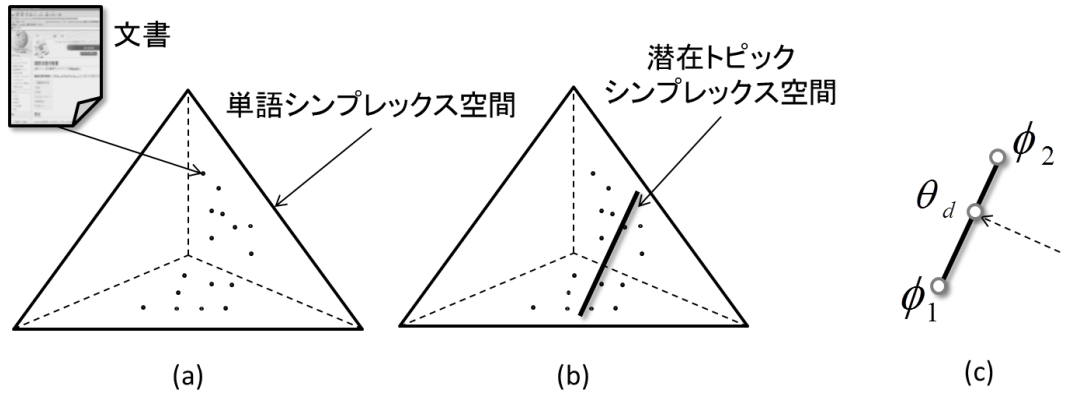

図 2 潜在トピックモデルの幾何学的な意味

度では不十分である。本研究では、学習の処理速度を上げる１つの有力な方法として学習の並列化に注目した。しかし、潜在的トピックモデルの学習に関しては、並列化の容易さと学習性能の良さはトレードオフの関係になっており、近年、１つの大きな研究テーマとなっている。ここで言う学習性能の良さとは、Perplexity と呼ばれる学習の汎化誤差に関する評価値である。潜在的トピックモデルの学習方法は複数提案されており、その並列化が１つのテーマとなっているが、並列化が容易な学習アルゴリズムは、学習性能が悪く、学習性能の良いアルゴリズムの並列化は原理上不可能であることが知られている。

本研究では、学習性能が state-of-the-art である従来手法と同程度で、並列化も容易なアルゴリズム開発を行う。基本的な方針は、state-of-the-art が制御システムにおける同期的ネガティブフィード機構であるとみなすことで、非同期ネガティブフィードバック機構へとアルゴリズムを変更すれば達成できることがわかった。

また、state-of-the-art である従来手法の大きな問題点とし、メモリ使用量が非常に高コストであることが挙げられる。例えば、文書データに対して適用する場合、文書中の全単語数を N、潜在変数の次元を K とすると、O(NK)のメモリ空間を要する。提案するアルゴリズムは、単語の種類数を V とすると O(KV)とすることができる。もちろん、V<<N である。

2. 当拠点公募型共同研究として実施した意義

(1) 共同研究を実施した拠点名および役割分担
東京大学(情報基盤センター)

(2) 共同研究分野
超大規模データ処理系応用分野

(3) 当公募型共同研究ならではという事項など
並列アルゴリズム開発

3. 研究成果の詳細と当初計画の達成状況

(1) 研究成果の詳細について

［潜在トピックモデル］

潜在トピックモデルは、bag と呼ばれる重複を許す集合に対する統計モデルである。例えば、文書データの場合、文書中の単語の出現頻度情報を bag of words と呼ばれる多重集合としてみることで、文書のモデル化ができる。購買履歴の場合は、各ユーザが購入したアイテム集合を bag として捉える。また、画像データでは、近年 bag of visual words と呼ばれる特徴抽出手法によ

り潜在トピックモデルで扱うことが可能になっている。

潜在トピックモデルは、統計モデルであるため様々な視点で説明することができるが、ここでは、最も分かりやすいシンプレックス空間上での幾何学的な解釈について説明する（図２参照)。各文書の単語の出現頻度は、和が１になるように正規化することで確率ベクトルと見なせる。これは各単語が基底ベクトルとなる単語シンプレックス空間上へ文書を射影したことに相当する（図２(a)参照)。

しかし、文書中の単語には偏りがあるため、単語シンプレックス空間上に射影された文書集合にもやはり偏りが見られるはずである。この傾向を利用して単語シンプレックス空間上に部分空間を作成し(図 2(b)参照)、その部分空間上へ文書を射影することで(図 2(c)参照)、文書集合の傾向を解析するモデルが潜在トピックモデルである。ここで、潜在トピックというのは、部分空間を構成する基底ベクトルで、通常単語数よりも非常に小さい次元数となる。この低次元の基底ベクトルを解析することで文書中のトピックの解析などが可能になる。

[潜在トピックモデルの学習]

従来の潜在トピックモデルの学習には、主に３つの手法が提案されている。

[決定的なアルゴリズムとして]

(1)変分ベイズ法

(2)周辺化変分ベイズ法

[確率的なアルゴリズムとして]

(3)周辺化ギブスサンプラー

がある。

(1)変分ベイズ法は、学習の並列化が容易であるが、学習性能が他の２つのアルゴリズムと比べて非常に悪い。

(2)周辺化変分ベイズ法は、変分ベイズ法の改良アルゴリズムで、次に説明する周辺化ギブスサンプラーと遜色ない学習性能を誇り、データやモデルによっては周辺化ギブスサンプラーよりも性能が良いことが実験的にわかってきている。しかし、並列化に関しては、周辺化ギブスサンプラーと同様の問題を含んでいる。確率的潜在変数モデルの学習は局所解を多く含む非線形最適化問題であるため、周辺化変分ベイズ法は変分ベイズ法と比べより良い局所解構造を見つけていると考えることができる。

(3)周辺化ギブスサンプラーは、サンプリングによる確率的アルゴリズムであるため局所解を回避しつつ学習を行えるため変分ベイズ法よりも性能がよい。しかし、学習に要する反復回数は、決定的アルゴリズムに対して 20-50 倍程度必要である。

したがって、現在の state-of-the-art は周辺化変分ベイズ法である。もともと、変分ベイズ法、周辺化ギブスサンプラー、周辺化変分ベイズ法の順に提案された。しかし、周辺化変分ベイズ法は、周辺化ギブスサンプラーと同様に、１つの潜在変数の更新が他の潜在変数の更新に影響を与えるため、並列して同時に複数の潜在変数の更新を行うことが原理的に不可能である。

本研究では、変分ベイズ法と同様に並列化が容易で、周辺化変分ベイズ法と同様の学習性能を達成する学習アルゴリズムを提案する。基本的な方針は、周辺化変分ベイズ法が制御システムにおける同期的ネガティブフィード機構であるとみなすことで、非同期ネガティブフィードバック機構へと

アルゴリズムを変更すれば達成できることがわかった。また、この視点は、周辺化変分ベイズ法が変分ベイズ法に比べて過学習というデータへのオーバーフィットを防ぐことを示唆しており、学習理論として新たな見方を与えるのではないかと考えている。

さらに、周辺化変分ベイズ法の大きな問題点とし、メモリ使用量が非常に高コストであることが挙げられる。例えば、購買データに対して適用する場合、ユーザ-アイテムペアの総数を N、潜在変数の次元を K とすると、O(NK)のメモリ空間を要する。提案するアルゴリズムは、アイテムの種類数を V とすると O(KV)とすることができる。もちろん、V<<N である。

提案するアルゴリズムの詳細について以下説明する。ただし、現在投稿中の内容であるため要点のみ説明する。

まず、潜在変数 $z_{d,i}$ を導入する。これは、文書 d の i 番目の単語がもつ潜在変数である。文書中で、この値が同じ単語集合は同じトピックに属することを意味する。購買データの場合は、ユーザ d が i 番目に購入した商品に対して、この潜在変数を仮定する。この時、潜在変数は、ユーザの嗜好や購入意図を表す隠れた情報を意味する。潜在変数の値が同じということは同じ意図で購入したということを表現する。この意図は商品のジャンルなどを表していると考えられる。この潜在変数の推定は、周辺化変分ベイズ法（実際には 0 次周辺化変分ベイズ法と呼ばれる方法であるが、ここでは周辺化変分ベイズ法と呼ぶことにする）を用いると、以下の式で求めることができる。

$$q(z_{d,i}=t) \propto \frac{\beta + \mathbb{E}[n_{t,w_{d,i}}^{\backslash d,i}]}{V\beta + \mathbb{E}[n_{t,\cdot}^{\backslash d,i}]}(\gamma_t + \mathbb{E}[n_{d,t}^{\backslash d,i}])$$

ここで、$\delta(w_{d,i}=v)$ を $w_{d,i}=v$ のとき 1、そうでなとき 0 を出力する関数として

$$\mathbb{E}[n_{d,t}] = \sum_{i=1}^{n_d} q(z_{d,i}=t)$$

$$\mathbb{E}[n_{d,t}^{\backslash d,i}] = \mathbb{E}[n_{d,t}] - q(z_{d,i}=t)$$

$$\mathbb{E}[n_{t,v}] = \sum_{d,i} q(z_{d,i}=t)\mathbb{I}(w_{d,i}=v)$$

$$\mathbb{E}[n_{t,v}^{\backslash d,i}] = \mathbb{E}[n_{t,v}] - q(z_{d,i}=t)\delta(w_{d,i}=v)$$

である。

$\mathbb{E}[n_{d,t}]$ は文書 d 内で $z_{d,i}=t$ と推定される単語の総数の期待値、$\mathbb{E}[n_{t,v}]$ は、全文書において単語 v が $z_{d,i}=t$ と推定された総数の期待値である。また、$\mathbb{E}[n_{t,\cdot}^{\backslash d,i}]$ は、$\mathbb{E}[n_{t,v}^{\backslash d,i}]$ の v に関する和である。つまり、$q(z_{d,i}=t)$ は、同一文書中での t の期待総数と、同一単語中での t の期待総数によって計算される。$\gamma_t + \mathbb{E}[n_{d,t}]$ を和が 1 になるように正規化することで図 2 の θ を得ることができる。ここで、ポイントなのは、この期待総数には、対象とする $z_{d,i}$ の 1 ステップ前に推定された情報（確率値）$q(z_{d,i}=t)$ が引かれている点である。これがネガティブフィードバックとなりデータへのオーバーフィットを防いでいると考えられる。

しかし、周辺化変分ベイズ法では、$q(z_{d,i}=t)$ の推定を文書全体に対して何度も反復する必要があるため、1 ステップ前の $q(z_{d,i}=t)$ を保持しておく必要がある。したがって、t が取りうる値を K 個とすると、$q(z_{d,i}=t)$ は、K 次元の確

率ベクトルとなるので、これを総単語数 N 全てに対して保持しておくと、必要なメモリ使用量は O(NK) となり大規模データには適用できない。

また、周辺化変分ベイズ法では、$q(z_{d,i}=t)$の変更が期待総数$\mathbb{E}[n_{t,w_{d,i}}]$、$\mathbb{E}[n_{t,\cdot}]$、$\mathbb{E}[n_{d,t}]$に影響をあたえるため、そのままの理論では並列に更新することができず、並列化できないアルゴリズムとなっている。

提案手法は、周辺化変分ベイズ法を α 情報量最小化として定式化し、α 情報量最小化問題を確率的最適化アルゴリズムによる解くことで導出したものである。確率的最適化にすることで、文書全体に対して反復を繰り返す必要がないため、$q(z_{d,i}=t)$を保持しておく必要がない。また、この最適化問題は非同期並列で実行することも容易である。以下、提案手法について説明していく。

まず、α 情報量最小化について説明し、次にその確率的最適化アルゴリズムについて説明する。

α 情報量は $\alpha \in (-\infty, \infty)$ をパラメータとして確率分布間の情報量を次のように一般化したものである。

$$D_\alpha[p||q] = \frac{\int \alpha p(\boldsymbol{x}) + (1-\alpha)q(\boldsymbol{x}) - p(\boldsymbol{x})^\alpha q(\boldsymbol{x})^{1-\alpha} d\boldsymbol{x}}{\alpha(1-\alpha)}$$

ここで、重要な点として p(x)、q(x) は正規化されている必要はない。p = q のとき α 情報量は 0 となる。

学習アルゴリズムは、ある計算困難な確率分布 $p(\boldsymbol{x})$ $(\boldsymbol{x} = \{x_1, x_2, \cdots, x_n\})$ を近似分布 $q(\boldsymbol{x}) = \prod_{i=1}^n q(x_i)$ で近似することが目的となっている。そこで、α 情報量を目的関数とした最適化問題として学習アルゴリズムの一般化を考える。この時、更新式は以下のように与えられる。

$$q(x_i) \propto \mathbb{E}\left[\left(\frac{p(\boldsymbol{x})}{q(\boldsymbol{x}^{\setminus i})}\right)^\alpha\right]^{\frac{1}{\alpha}}_{q(\boldsymbol{x}^{\setminus i})}$$

多くの問題では、この期待値計算は計算困難となる場合が多いが、

$$q(x_i) \propto \mathbb{E}\left[\left(p(x_i|\boldsymbol{x}^{\setminus i})\frac{p(\boldsymbol{x}^{\setminus i})}{q(\boldsymbol{x}^{\setminus i})}\right)^\alpha\right]^{\frac{1}{\alpha}}_{q(\boldsymbol{x}^{\setminus i})}$$

より、$p(\boldsymbol{x}^{\setminus i})$を$q(\boldsymbol{x}^{\setminus i})$と置き換えると

$$q(x_i) \propto \mathbb{E}\left[p(x_i|\boldsymbol{x}^{\setminus i})^\alpha\right]^{\frac{1}{\alpha}}_{q(\boldsymbol{x}^{\setminus i})}$$

を得る。この更新式は $\alpha = 1$ の場合、Belief Propagation (BP) または Expectation Propagation (EP) と呼ばれる手法になっている。したがって、これらの手法を一般化したものとみることができる。また、この更新式は、次のように局所的に α 情報量を用いた射影と見ることができる。

$$q^*(x_i) = \underset{q(x_i)}{\operatorname{argmin}} D_\alpha[p(x_i|\boldsymbol{x}^{\setminus i})q(\boldsymbol{x}^{\setminus i})||q(\boldsymbol{x})]$$

この射影アルゴリズムの観点から周辺化変分ベイズ法を導出する。

まず、正規化していない近似事後分布分布 $\tilde{q}(z_{d,i})$

をパラメータ $a_{d,t}^{\setminus d,i}$、$b_{t,v}^{\setminus d,i}$、$c_t^{\setminus d,i}$ を用いて以下のように定義する。

$$\tilde{q}(z_{d,i}=t) = \frac{b_{t,v}^{\setminus d,i}}{c_t^{\setminus d,i}} a_{d,t}^{\setminus d,i}$$

これにより、周辺化変分ベイズ法は、上記３つのパラメータ推定として考えることができる。ここで、$\tilde{q}(z_{d,i})$が $\boldsymbol{b}_v^{\setminus d,i} = \{b_{t,v}^{\setminus d,i}\}_{t=1}^T$ の関数であることを

強調する必要がある場合、表記 $\tilde{q}(z_{d,i}|\boldsymbol{b}_v^{\setminus d,i})$ を用いることにする．他のパラメータについても同様である。更に以下、幾つかの定義を行う。まず、

$$\tilde{q}^{\setminus b}(z_{d,i}=t) = \tilde{q}(z_{d,i}=t)/b_{t,v}^{\setminus d,i} = \frac{a_{d,t}^{\setminus d,i}}{c_t^{\setminus d,i}}$$

とし、$b_{t,v}^{\backslash d,i}$ と$n_{t,v}^{\backslash d,i}+\beta$ を交換した式を

$$\tilde{q}^{b\to n}(z_{d,i}=t)=(n_{t,v}^{\backslash d,i}+\beta)\tilde{q}^{\backslash b}(z_{d,i}=t)$$

とする。

ここで、以下の α 情報量最小化問題を考える。α = 1 として

$$\min_{b_{t,v}^{\backslash d,i}} \mathcal{D}_1[\tilde{q}^{b\to n}(z_{d,i})q(\boldsymbol{z}^{\backslash d,i})||\tilde{q}(z_{d,i}|\boldsymbol{b}_v^{\backslash d,i})q(\boldsymbol{z}^{\backslash d,i})]$$

は、解析的に解くことができて

$$b_{t,v}^{\backslash d,i}=\mathbb{E}[n_{t,v}^{\backslash d,i}]+\beta$$

を得る。

同様に、以下の最適化問題を解くことで

$$\min_{a_{d,t}^{\backslash d,i}} \mathcal{D}_1[\tilde{q}^{a\to n}(z_{d,i})q(\boldsymbol{z}^{\backslash d,i},\alpha,\boldsymbol{\pi})||\tilde{q}(z_{d,i}|\boldsymbol{a}_d^{\backslash d,i})q(\boldsymbol{z}^{\backslash d,i},\alpha,\boldsymbol{\pi})]$$

$$\min_{c_t^{\backslash d,i}} \mathcal{D}_{-1}[\tilde{q}^{c\to n}(z_{d,i})q(\boldsymbol{z}^{\backslash d,i})||\tilde{q}(z_{d,i}|\boldsymbol{c}^{\backslash d,i})q(\boldsymbol{z}^{\backslash d,i})]$$

解析解として

$$a_{d,t}^{\backslash d,i}=\mathbb{E}[n_{d,t}^{\backslash d,i}]+\gamma_k,\ c_t^{\backslash d,i}=\mathbb{E}[n_{t,\cdot}^{\backslash d,i}]+V\beta$$

を得ることができる。

したがって、これらを$\tilde{q}(z_{d,i}=t)$ へ代入することで周辺化変分ベイズ法の更新式が得られた。

周辺化変分ベイズ法の式が最適化問題として定式化されたことにより、確率的最適化アルゴリズムを適用することができる。

α 情報量により最小化では、$a_{d,t}^{\backslash d,i}$ は、文書固有のパラメータであり $c_t^{\backslash d,i}$ は、$c_t^{\backslash d,i}=\sum_v b_{t,v}^{\backslash d,i}$ として求めることができるので、$b_{t,v}^{\backslash d,i}$ の最適化問題を確率的最適化アルゴリズムとして求めればよい。

まず、目的の α 情報量最小化問題の解析解を考えると、$w_{d,i}=v$ のとき

$$b_{t,v}^{\backslash d,i}=\mathbb{E}[n_{t,v}^{\backslash d,i}]+\beta$$

であった。これを以下のように、ステップサイズ ρ を用いた固定点反復法として書き換える。

$$b_{t,v}^{\backslash d,i}\leftarrow(1-\rho)b_{t,v}^{\backslash d,i}+\rho[\mathbb{E}[n_{t,v}^{\backslash d,i}]+\beta]$$

$$=b_{t,v}^{\backslash d,i}+\rho[\mathbb{E}[n_{t,v}^{\backslash d,i}]+\beta-b_{t,v}^{\backslash d,i}]$$

ここで、

$$\mathbb{E}[n_{t,v}^{\backslash d,i}]=(n_v-1)\sum_{d',i'\neq d,i}\frac{1}{n_v-1}q(z_{d',i'}=t)\delta(w_{d',i'}=v)$$

に対して次のように確率的近似を行う。

$(d',i')\neq(d,i)$ を$w_{d',i'}=v$ であるような単語のランダムサンプルとして

$$\tilde{n}_{t,v}^{\backslash d,i}=n_v^{\backslash d,i}q(z_{d',i'}=t)=(n_v-1)q(z_{d',i'}=t)$$

と近似する。したがって、更新式は

$$b_{t,v}^{\backslash d,i}\leftarrow b_{t,v}^{\backslash d,i}+\rho[n_v^{\backslash d,i}q(z_{d',i'}=t)+\beta-b_{t,v}^{\backslash d,i}]$$

となる。

実際にはランダムサンプルではなく、過去に処理したデータを用いることができる。

マルチンゲールの収束定理により、ρ に対して特定の条件を仮定して、上記の近似更新式を繰り返し行うことで、α 情報量最小化問題の近似解を得ることを示すことができた。この手法によりネガティブフィードバックを考慮したアルゴリズムを作ることができる。提案アルゴリズムは、この理論を基に文書全体に対して反復を行うのではなく、各文書毎に反復を行うアルゴリズムを作ることができる。１度処理した文書で学習する必要がないため$q(z_{d,i}=t)$を保持しておく必要がなくメモリ効率も優れている。また、$b_{t,v}^{\backslash d,i}$ の最適化問題は文書に対して並列に解くことができるため、並列化が容易なアルゴリズムになっている。理論の証明などの詳細は投稿中の論文に示しめしているので、採録後確認していただきたい。

次に具体的な実験結果について説明する。ここでは論文中で示した実験のうち Pubmed データセット 1 と呼ばれる学術論文のアブストラクト集合に対する実験結果を説明する。
文書数は 1,000,000(100 万)、語彙数 V は 50,161(およそ 5 万)、総単語数 N は 92,380,343(およそ 9000 万)である。冠詞や機能語のようなストップワードと低頻度語は、学習に悪影響を与えるため削除してある。トピック数は K=1000 とした。ただし、このトピック数は上限であり、提案アルゴリズムはトピック数の自動決定機構も備えている(詳しくは採録後の論文を参照のこと)。

このような大規模のデータに対して、従来の周辺化変分ベイズ法は、O(KN) のメモリ空間を必要とするため適用することはできない。提案手法では、O(KV) のメモリ空間で済むため適用可能である。

ここでは、ネガティブフィードバック機構の有無で、汎化能力を表す Perplexity を評価する。Perplexity が低いほど、汎化能力の面で学習効果が高いことを示す。
図 1 に実験結果を示す。横軸が学習した文書数で、縦軸が Perplexity である。実験結果からも分かる通り、ネガティブフィードバック機構を備えるアルゴリズムは文書数の変化に関わらず、ネガティブフィードバックを備えていないアルゴリズムよりも汎化能力が高いことがわかる。また、ネガティブフィードバック機構を備えることで、30 万文書程度からの学習結果でも、ネガティブフィードバック機構を備えないアルゴリズムが 100 万文書から学習した結果に匹敵することがわかる。

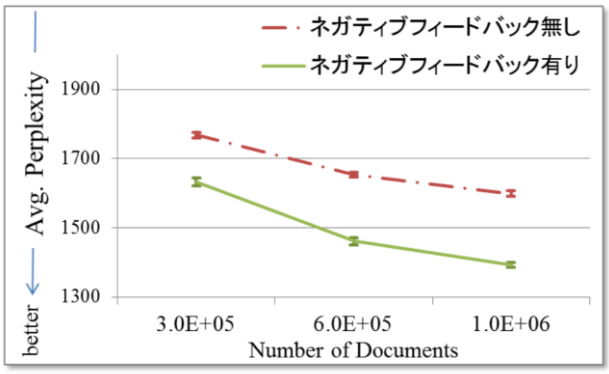

図 1: 実験結果

(2) 当初計画の達成状況について

大規模データへの適用を考える際に並列化と同程度に重要な課題として周辺化変分ベイズ法のメモリコストの問題があるため、まずこの問題の解決を行った。しかし、研究を進めていくにつれ、メモリコストを抑えつつ汎化性能が高い学習アルゴリズムを作ることが非常に難しいことがわかった。

当初の予定では、確率的近似法により、周辺化変分ベイズ法を近似することでメモリコストを抑えることを行っていた。しかし、メモリコストを抑えることには成功したものの高い汎化性能を達成することはできなかった。ただし、この方法は従来の変分ベイズ法よりも汎化能力が高いアルゴリズムではあった。

周辺化変分ベイズ法の性能の良さを分析するにあたり、周辺化変分ベイズ法の性能の良さは、制御システムにおけるネガティブフィードバック機構を備えているからではないかという着想に至った。周辺化変分ベイズ法では、文書全体に対して繰り返し学習を行い、１回の反復辺り１回のネガティブフィードバックを行うアルゴリズムと見なせる。これは同期的な分散ネガティブフィードバックシステムであると見なせるため、非同期分散ネガティブフィードバック機構のアルゴリズムへと変更することを試

みて、予備実験を行ったところメモリコストを抑えつつ汎化性能が高いアルゴリズムとなることが予備実験の結果わかった。さらに、このアルゴリズムは、非同期なネガティブフィードバックであるため並列実行も容易であるという特性があることもわかった。

この着想を基に、当初行っていた確率的近似法の理論を再構築し、ネガティブフィードバック機構を備えるアルゴリズムへと拡張することで、当初行っていた確率的近似アルゴリズム(図１のネガティブフィードバック無し)の性能アルゴリズムを向上させることができた。図１からも分かる通り、最終的に得られたアルゴリズム(図１のネガティブフィードバックあり)は、30 万文書程度からの学習結果でも、ネガティブフィードバック機構を備えないアルゴリズムが 100 万文書から学習した結果に匹敵することがわかる。これは、少ない学習データでも効率的に学習が行えていることを示している。これにより当初の目標を達成できたと言える。

4. 今後の展望

今回得られたアルゴリズムと可視化技術を組合せて大規模学術情報解析を行う予定である。また、このアルゴリズムは他の潜在変数モデルにも適用できるため、幅広い応用が期待できる。

5. 研究成果リスト

(1) 学術論文（投稿中のものは「投稿中」と明記）

(2) 国際会議プロシーディングス

データマイニングの国際会議に投稿中である

(3) 国際会議発表

(4) 国内会議発表

(5) その他（特許，プレス発表，著書等）