

# ガウス過程を用いたラプラシアンラベル伝搬法の拡張

江原 遥<sup>†1</sup> 佐藤 一 誠<sup>†2</sup> 中川 裕 志<sup>†2</sup>

本研究では、自然言語処理における代表的なブートストラップ法であるラプラシアンラベル伝搬法を、ガウス過程を用いて拡張する。近年、自然言語処理において、パターン（文脈）とインスタンス（語）のグラフ構造と少数のシード語から、シード語との意味的な関連度を計算する半教師あり学習法であるブートストラップ法が注目されている。グラフを用いたブートストラップ法の中では、ラプラシアンラベル伝搬法は最も代表的な手法である。しかし、この手法は、関連度を求めるために全てのインスタンスを一辺とする正方行列の逆行列計算を扱わなければならない、空間計算量・時間計算量共に大きかった。本研究では、ラプラシアンラベル伝搬法がガウス過程の MAP (Maximum a posteriori) 推定のラグランジュ双対問題になっていることを示す。ガウス過程の文脈においては、適切な設定の元では、1つの重みベクトルを使用する MAP 推定よりも、重みベクトルを事前分布を考慮しながら積分消去するベイズ推定から導出される予測分布の方が高い判別性能を示すことが知られている。本稿では、ラプラシアンラベル伝搬法と、対応するガウス過程の予測分布を比較し、語義曖昧性解消の課題においては、ラプラシアンラベル伝搬法よりも高い分類性能を示すことを確認した。

## 1. はじめに

近年、自然言語処理においては、入力となるシード語とパターン（文脈）から、シード語と意味的に関連する語を獲得する方法として、ブートストラップ法が注目されている。ブートストラップ法の基本的な仕組みは次のようである。大規模なテキストからシード語と共起するパターンを収集し、次に、そのパターンを用いてシード語を拡張することを繰り返すことにより、シード語と意味的に関連する語を獲得していく。

図 1 に、本研究の位置づけを示す。ブートストラップ法は、主に自然言語処理分野での名称であり、機械学習の分野では、類似の手法が Label Propagation（ラベル伝搬）と

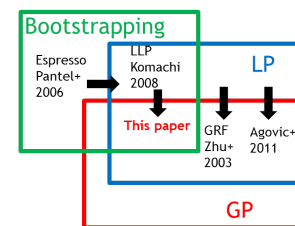


図 1 本研究の位置づけ

いう名称で呼ばれている。近年のブートストラップ法の有名な例としては、Pantel ら<sup>5)</sup> が Espresso を提案した。小町ら<sup>3)</sup> は、この Espresso を単純化した Simplified Espresso を行列を用いて定式化することで Espresso では意味ドリフトの問題が起きやすいことを示し、ラプラシアンラベル伝搬法 (Laplacian label propagation, LLP) では意味ドリフトの問題が軽減されることを示した。

ラプラシアンラベル伝搬法では、インスタンスとパターンをそれぞれノードとする重み付き二部グラフを構成し、そのグラフ上でシードインスタンスから与えたラベルが伝搬する過程を考える。しかし、ラプラシアンラベル伝搬法は、全てのインスタンス数  $N$  を一辺とする正方行列の逆行列の計算を含むため、空間計算量が  $O(N^2)$ 、時間計算量は  $O(N^3)$  と大きく、特に空間計算量によって扱える問題のサイズが実質的に制限されてしまうという問題を抱えている。この問題を解決するため、小町ら<sup>4)</sup> は、逆行列を無限級数展開し有限次元で計算を近似することで、大規模な問題に対してこのラプラシアンラベル伝搬法が近似的に適用できるようにする Quetchup アルゴリズムを提案している。

ラプラシアンラベル伝搬法の本質的な問題点は、テストデータまで含めてスコアを求めたい全てのインスタンスからなる巨大な行列を考えなければならないことにある。そこで、本研究では、ラプラシアンラベル伝搬法の改善を目指した結果、以下のような貢献に値する知見を得た。

- ラプラシアンラベル伝搬法のガウス過程 (GP) の MAP (Maximum a posteriori) 推定に帰着することを示した。
- ガウス過程において、適切なパラメータや事前分布を設定すれば MAP 推定より高い分類性能を示すことが知られているベイズ推定を利用して、本稿の目的でも MAP 推定より高い分類性能を示すことを確認した。

図 1 に示す通り、ラベル伝搬と GP との関連を示した研究としては、Zhu ら<sup>8)7)</sup> が先鞭

<sup>†1</sup> 東京大学大学院情報理工学系研究科

<sup>†2</sup> 東京大学情報基盤センター

をつけている．しかし、これらの文献では、彼らの提案する GRF と呼ばれる手法と GP との関連が示されており、ラプラシアンラベル伝搬法については述べられていない．また、これらの文献では GP の予測分布を用いることによる上記の利点は実証されていない．また、近年、Agovic<sup>1)</sup> は、我々の定式化と似た形でラベル伝搬法をまとめ、その GP との関連について記述している．しかし、この文献<sup>1)</sup> においても、ラプラシアンラベル伝搬法は扱われておらず、また、ベイズ推定による分類性能の向上は GP に帰着させることによる上記の利点を実証するものではない．

## 2. 問題設定

データ集合  $\mathcal{D} = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l), x_{l+1}, \dots, x_{l+q}, x_{l+q+1}, \dots, x_{l+q+r}\}$  が与えられたとする．このうち、 $(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)$  は  $l$  個のラベル付きデータであり、 $y_i \in \{-1, 1\}$  は素性ベクトル  $x_i$  に対するラベルである．一方、ラベルなしデータは  $x_{l+1}, \dots, x_{l+q+r}$  の  $q+r$  個であるが、このうち、 $x_{l+1}, \dots, x_{l+q}$  のラベルだけが求めれば良いとする．

以後、特に断らない限り、2 値判別の場合について述べる． $\mathbf{y}_l = (y_1, \dots, y_l)^\top$  と置く．ある行列  $B$  に対して、 $(B)_{i,j}$  を  $B$  の  $i, j$  成分とする．また、全ての要素が 0 からなる  $l$  次元ベクトルを  $\mathbf{0}_l$  と書く．

本稿で述べるブートストラップ法の目的は、 $\mathcal{D}$  が与えられたときに、 $q$  個のラベルなしデータ  $x_{l+1}, \dots, x_{l+q}$  に対するラベル  $y_i$  を求めることである．2 値判別の場合、ラベルは実数値のスコア  $f_q = (f_{l+1}, \dots, f_{l+q})^\top$  を介して求められる．求めたスコアベクトル  $f_q$  から、各インスタンスに対するラベルを求めるには、通常、しきい値を用いる．すなわち、スコアの高いインスタンスほど 1 のラベルを持つ傾向が高いとみなし、しきい値を決めて  $f_q$  の要素がしきい値以上であればラベル 1、しきい値未満であればラベル  $-1$  とみなす． $r$  個の中から  $0 \leq r_c \leq r$  個を選び、 $u = q + r_c$  と置く． $r_c$  個の選び方については、様々な方法が考えられるが、本稿では特に断らない限りランダムに選ぶものとする． $\mathbf{s} = \begin{pmatrix} \mathbf{y}_l \\ \mathbf{0}_u \end{pmatrix}$  と置き、これをシードベクトルと呼ぶ．

本稿では、特に断りのない限り、二値判別について説明するが、多値判別への拡張は、通常、one-versus-rest 法で 2 値判別を用いて行われるため、容易である．one-versus-rest 法では、各ラベルに属するか否かの二値判別をラベル数回行い、 $\mathcal{D}$  中の各インスタンスを、出力されたラベル数個のスコアのうち最も高いクラスに属すると判別する．

個々のインスタンス  $x_i$  に対して  $M$  個のパターン（素性）との関係を、 $N \times M$  行列  $W$  を用いて表す．ただし、 $N = l + u$  である． $(W)_{i,j}$  は、インスタンス  $x_i$  の  $j$  個目の素性値である．インスタンスとパターンの両方  $N + M$  個をノードする重み付き 2 部グラフを  $G$  とし、 $A_0$  をグラフ  $G$  の隣接行列とすると、 $A_0^2$  のインスタンスに関する部分行列が  $WW^\top$  となる．以下、簡単のために、 $A = WW^\top$  とする．

### 2.1 ラプラシアンラベル伝搬法

本節では、ラプラシアンラベル伝搬法<sup>4)</sup> について説明する．ラプラシアンラベル伝搬法は、数式 (1) のように定式化できる． $A$  上で近い位置にあるインスタンス同士は近い値を持つようにペナルティが課される（数式 (1) 第 2 項）．

$$\underset{\mathbf{f}}{\text{minimize}} \quad \|\mathbf{s} - \mathbf{f}\|^2 + \alpha \sum_{i=1}^N a_{i,j} (f_i - f_j)^2. \quad (1)$$

数式 (1) の第 2 項は、ラプラシアン行列を用いて、 $\mathbf{f}^\top \mathcal{L}(WW^\top) \mathbf{f}$  と表せる．ここで、行列  $X$  に対するグラフラプラシアン行列は、 $\mathcal{L}(X) = D_X - X$  と定義される．ただし、 $(D_X)_{i,i} = (\sum_j x_{i,j})$  である．ラプラシアン行列の代わりに、固有値を制限した正規化ラプラシアン行列  $\mathcal{L}_{\text{norm}}(X) = D_X^{-1/2} \mathcal{L}(X) D_X^{-1/2} = I - D_X^{-1/2} X D_X^{-1/2}$  が使われることが多い．

正規化ラプラシアンを使用する場合、 $L = \mathcal{L}_{\text{norm}}(WW^\top)$  を用いて、数式 (1) は、次のように書ける．

$$\underset{\mathbf{f}}{\text{minimize}} \quad \frac{1}{\alpha} \|\mathbf{s} - \mathbf{f}\|^2 + \mathbf{f}^\top L \mathbf{f}, \quad (2)$$

数式 (2) は二次計画問題なので、その解は次のように求められる．

$$\mathbf{f} = \frac{1}{\alpha} \left( L + \frac{1}{\alpha} I \right)^{-1} \mathbf{s}. \quad (3)$$

## 3. ガウス過程による解釈

### 3.1 カーネル線形回帰

本節では、前節で説明したラプラシアンラベル伝搬法をカーネル線形回帰の枠組みで解釈する．カーネル線形回帰はガウス過程に帰着できることが知られている．カーネル線形回帰からガウス過程への帰着方法は、§3.2 で述べる．

カーネル線形回帰の枠組みでは、前節のラプラシアンラベル伝搬法の数式 (4) のように書ける．ただし、 $\epsilon \sim \mathcal{N}(0, \sigma_n^2)$  である．また、 $\phi: \mathcal{X} \rightarrow \mathbb{R}^M$  を、インスタンス  $x_i$  から  $M$  次元のパターンのベクトル（素性ベクトル）に写す写像とする．

$$\forall i \in \{1, \dots, \mathcal{N}\}; s_i = \mathbf{w}^\top \phi(x_i) + \epsilon, \quad (4)$$

数式 (4) の枠組みでは、各インスタンス  $x_i$  と、そのラベル  $s_i$  (ラベルがない場合は  $s_i = 0$ ) が与えられたときに、次の二乗損失関数を 2 ノルム正則化付きで最小化するような  $w_i$  を見つけることが目的となる。この問題は、インスタンス  $x_i$  のスコアを  $\langle w_i, \phi(x_i) \rangle$  とおいたときに、 $s_i$  とこのスコアの差の二乗損失を小さくしつつ (第 1 項)、大きすぎる  $w_i$  を 2 ノルムで正則化 (第 2 項) する問題と解釈できる。

$$\underset{\mathbf{w}}{\text{minimize}} \sum_{i=1}^N \|s_i - \langle w_i, \phi(x_i) \rangle\|^2 + \|\mathbf{w}\|^2. \quad (5)$$

以下では、まず、数式 (5) の解を求め、数式 (3) に示すラブラシアンラベル伝搬法の解と対応していることを示す。次に、この 2 つの最適化問題の解がたまたま一致しているだけでなく、最適化問題としての数式 (5) 自体が、数式 (2) と対応する問題に書き換えることが出来ることを示す。

数式 (5) 中の第 1 項の損失関数を  $\xi_i = \|s_i - \langle w_i, \phi(x_i) \rangle\|$  と置くと、数式 (5) を数式 (6) に示す等価であるが制約が導入された問題に置き換えることが出来る<sup>2)</sup>。

$$\begin{aligned} \underset{\mathbf{w}}{\text{min.}} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + \frac{1}{2\lambda} \sum_{i=1}^N \xi_i^2 \\ \text{subject to} \quad & \forall i \in \{1, \dots, N\}; s_i = \mathbf{w}^\top \phi(x_i) + \xi_i. \end{aligned} \quad (6)$$

数式 (6) は等式制約しか持たないので、ラグランジュ未定乗数法を使うことにより、数式 (6) のように書き換えることが出来る。ただし、 $\xi = (\xi_1, \dots, \xi_N)$  であり、 $\mathbf{f} = (f_1, \dots, f_N)$  である。

$$g(\mathbf{w}, \xi, \mathbf{f}) = \frac{1}{2} \|\mathbf{w}\|^2 + \frac{1}{2\lambda} \sum_{i=1}^N \xi_i^2 - \sum_{i=1}^N (\langle \mathbf{w}, \phi(x_i) \rangle + \xi_i - s_i) f_i, \quad (6)$$

数式 (6) を  $\mathbf{w}$  と  $x_i$  で微分することによって、数式 (7) と数式 (8) を得る。

$$\frac{\partial g}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^N f_i \phi(x_i). \quad (7)$$

$$\frac{\partial g}{\partial \xi_i} = 0 \Rightarrow f_i = \frac{\xi_i}{\lambda}. \quad (8)$$

得られた数式 (7) と数式 (8) を使って数式 (6) から  $\mathbf{w}$  と  $x_i$  を消去することによって、数式 (9) が導出できる。

$$\sum_{j=1}^N f_j \langle \phi(x_i), \phi(x_j) \rangle + f_i \lambda = s_i. \quad (9)$$

数式 (9) において、内積を  $K_{i,j} = \langle \phi(x_i), \phi(x_j) \rangle$  であるようなカーネル行列  $K$  によって書き換えることにより、数式 (10) を導出できる。

$$\mathbf{f} = (K + \lambda I)^{-1} \mathbf{s} = ((K + \tau \lambda I) + ((1 - \tau) \lambda I))^{-1} \mathbf{s}. \quad (10)$$

数式 (10) と数式 (3) を比較すると、 $K = L + \tau \lambda I$ 、 $(1 - \tau) \lambda = \frac{1}{\alpha}$  とすれば、両式が一致することが分かる。ここで、 $\tau$  は、 $0 < \tau < 1$  を満たす実数であり、0 固有値を必ず持つ半正定値行列である  $L$  に  $\tau I$  を加えることによって、 $L + \tau I$  を正定値とし、正定値行列である  $K$  とみなせるように調整している。

これで、数式 (11) と数式 (2) の解が一致することが示された。

次に、数式 (5) の問題自体を書き換えて、数式 (2) に対応する形に変形できることを示す。まず、数式 (5) が、等価な制約付き最適化問題数式 (6) に問題として変形できることは既に述べた。数式 (7) と数式 (8) を直接数式 (6) に代入して、前述の  $K_{i,j} = \langle \phi(x_i), \phi(x_j) \rangle$  であるようなカーネル行列  $K$  を導入することによって、 $\mathbf{w}$  に関する最小化問題であった数式 (6) を、次の  $\mathbf{f}$  に関する最小化問題に書き換えることが出来る。

$$\underset{\mathbf{f}}{\text{minimize}} \|\mathbf{s} - K\mathbf{f}\|^2 + \lambda \mathbf{f}^\top K \mathbf{f}. \quad (11)$$

これで、数式 (5) が数式 (11) に書き換えられた。数式 (2) と数式 (11) の最大の違いは、第 1 項の  $K$  である。数式 (11) では、第 1 項の  $\mathbf{f}$  に  $K$  がかかっているが、数式 (2) では、この  $K$  はない。この違いは、次の補題によって解消される。

**Lemma 1.** 数式 (11) と数式 (2) を比較すると、この 2 式には定数倍の違いしかないので、 $L$  と  $\lambda$  を適切に設定すれば、次を成り立たせることが可能である。

$$\underset{\mathbf{f}}{\text{minimize}} \|\mathbf{s} - K\mathbf{f}\|^2 + \lambda \mathbf{f}^\top K \mathbf{f} = \underset{\mathbf{f}}{\text{minimize}} \|\mathbf{s} - \mathbf{f}\|^2 + \alpha \mathbf{f}^\top L \mathbf{f}. \quad (12)$$

*Proof.* 数式 (2) を  $\mathbf{f}$  に関して微分して 0 とおくことにより、次を得る。

$$\underset{\mathbf{f}}{\text{minimize}} \frac{1}{\alpha} \|\mathbf{s} - \mathbf{f}\|^2 + \alpha \mathbf{f}^\top L \mathbf{f} = (L + \frac{1}{\alpha} I)^{-1} \mathbf{s}. \quad (13)$$

一方、数式 (11) では、 $\mathbf{f}$  に関して微分して 0 と置くことにより次を得る。

$$\|\mathbf{s} - K\mathbf{f}\|^2 + \lambda \mathbf{f}^\top K \mathbf{f} = (K^\top K + \lambda K)^{-1} K^\top \mathbf{s}. \quad (14)$$

仮定より  $K$  は対称正定値であるので、 $K^\top = K$  であり、 $K^{-1}$  が存在する。従って、数

式 (14) は、次の形で書ける．

$$\text{数式 (14)} = (K^\top + \lambda I) K^{-1} K^\top \mathbf{s} = (K^\top + \lambda I) \mathbf{s}. \quad (15)$$

数式 (14) と数式 (15) を比べることにより、 $K$  と  $\lambda$  を適切に設定すれば、両式は一致する．即ち、適切な設定のもとでは、数式 (2) と数式 (11) は解空間が一致するので、同じ最適化問題とみなしてよい．

□

### 3.2 カーネル線形回帰とガウス過程

前節で、数式 (2) がカーネル線形回帰として解釈できることを示した．本節では、カーネル線形回帰の数式 (4) がガウス過程の MAP 推定であることを説明する．先にガウス過程を導入し、次にその MAP 推定が数式 (4) に帰着することを示す．

<sup>6)</sup> の weight-space view に対応してガウス過程を説明する． $X = \{x_1, \dots, x_N\}$ 、重み  $\mathbf{w}$  が与えられたときに、 $x_i \in X = \{x_1, \dots, x_N\}$  中の各  $x_i$  に対して、 $s_i \in \mathcal{S}$  が出力される． $\Phi_X^\top$  がデザイン行列となるような行列  $\Phi_X = (\phi(x_1) \ \phi(x_2) \ \dots \ \phi(x_N))$  を定義し、 $\sigma^2 = \lambda$  とする．さらに、期待値  $\boldsymbol{\mu}$ 、分散共分散行列が  $\Sigma$  であるような多変量ガウス分布の確率密度関数を  $h_{\mathcal{N}}(\boldsymbol{\mu}, \Sigma) = (2\pi)^{-\frac{k}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x - \boldsymbol{\mu})^\top \Sigma (x - \boldsymbol{\mu})\right)$  とする．

$$p(\mathbf{s} | X, \mathbf{w}) = h_{\mathcal{N}}(\Phi_X^\top \mathbf{w}, \sigma^2 I) \quad (16)$$

$$= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{s} - \Phi_X^\top \mathbf{w}\|^2\right) \quad (17)$$

$$= \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (s_i - \phi(x_i)^\top \mathbf{w})^2\right). \quad (18)$$

$\mathbf{w}$  の事前確率を  $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, I)$  とすると、事後確率はベイズの定理を用いて、 $p(\mathbf{w} | X, \mathbf{s}) = \frac{p(\mathbf{s} | X, \mathbf{w})p(\mathbf{w})}{p(\mathbf{s} | X)}$  となる．分母は  $\mathbf{w}$  に関しては定数であるので、事後確率  $p(\mathbf{s} | X, \mathbf{w})$  を最小化する問題は、以下のように数式 (5) に帰着する．

$$\operatorname{argmax}_{\mathbf{w}} p(\mathbf{w} | X, \mathbf{s}) = \operatorname{argmax}_{\mathbf{w}} p(\mathbf{s} | X, \mathbf{w})p(\mathbf{w}) \quad (19)$$

$$= \operatorname{argmin}_{\mathbf{w}} (-\log p(\mathbf{s} | X, \mathbf{w})p(\mathbf{w})) = \arg \quad (\text{数式 (5)}). \quad (20)$$

従って、数式 (4) を解くことは、ガウス過程の MAP 推定と一致する．表 1 中の MAP 推定におけるテストデータを除いたバージョンは、 $\mathbf{w}$  を数式 (16) にそのまま代入することによって得られる<sup>6)</sup>．

## 4. 提案手法

数式 (20) では、事後分布を最大化する 1 つの  $\mathbf{w}$  を用いてスコアを求めている．それに対して、提案手法では、ベイズの枠組みを用いて、全ての  $\mathbf{w}$  を考慮し、 $\mathbf{w}$  パラメータで周辺化したスコアを求める． $l$  個のラベル付きデータと  $t$  個のラベルなしデータ  $x_1, \dots, x_l, x_{l+1}, \dots, x_{l+t}$  が与えられたとする． $x_1, \dots, x_l$  に対応するラベルを  $y_1, \dots, y_l$  とする．また、 $x_{l+1}, \dots, x_{l+t}$  に対しては、便宜的に  $t$  次元のゼロベクトル  $\mathbf{0}_t$  が対応するものとする．この時、ある 1 つの新規のデータ  $x_*$  に対するスコア  $f_*$  の予測分布は、数式 (21) で書ける．

$$\begin{aligned} & p(f_* | x_*, \{x_1, \dots, x_l, x_{l+1}, \dots, x_{l+t}\}, \{y_1, \dots, y_l, \mathbf{0}_t\}) \\ &= \int p(f_* | x_*, \mathbf{w}) p(\mathbf{w} | \{x_1, \dots, x_l, x_{l+1}, \dots, x_{l+t}\}, \{y_1, \dots, y_l, \mathbf{0}_t\}) d\mathbf{w} \end{aligned} \quad (21)$$

数式 (21) は、1 つの新規データに対するスコアであり、1 次元ガウス分布になる．複数のスコアに対する予測分布は多次元ガウス分布になり、その平均ベクトルは表 1 の提案手法の数式のように書け、分散共分散行列は  $(L_{qq} - L_{q,l+t} (L_{l+t,l+t} + \alpha I)^{-1} L_{l+t,q})_{i,i}$  のように書ける．

表 1 に各手法をまとめた．提案手法の特徴は、時間計算量、空間計算量にラベルなしデータ数  $u$  が出現せず、代わりに  $t$  で支配されることである．ここで、 $t$  は、 $\{x_1, \dots, x_l, x_{l+1}, \dots, x_{l+1}, \dots, x_{l+q+r}\}$  の中から自由に選択することが出来る．提案手法では、既存のラプラシアンラベル伝搬法や GRF にはない、この  $t$  というパラメータを用いて、ラベルなしデータの数を調節することが出来る．表中の添字は、添字なしの大きな行列中の対応する部分の部分行列である．例えば、 $A_{uu}$  は、 $l+u \times l+u$  の正方行列である  $A$  のラベルなしデータに対応する  $u \times u$  の部分行列を表す．

## 5. 評価

本研究では語義曖昧性解消 (word sense disambiguation, WSD) のタスクで提案手法を評価した．以下、語義曖昧性解消による評価がブートストラップ法の評価として適切である理由を説明する．ラプラシアンラベル伝搬法などの自然言語処理におけるブートストラップ法は、「文脈 (パターン) が似た語 (インスタンス) は似た意味を持つ」という分布仮説を利用して、シードインスタンスと意味的に近いインスタンスを獲得しようとする手法である．従って、ブートストラップ法という手法の評価においては、分布仮説を利用して文脈からどれだけシードインスタンスと近い意味のインスタンスを予測できるかが主要な性能を

表 1 各手法のまとめ． $l \ll u$ ． $u = q + r_c$ ． $t$  は  $0 \leq t \leq u$  の範囲で自由に調節可能．ラベルなしデータ  $u$  中で興味のある  $q$  個のデータに対するスコア  $f_q$  が出力されればよい．

手法	数式	時間計算量	空間計算量
ラプラシアン ラベル伝搬法	$\begin{pmatrix} f_l \\ f_u \end{pmatrix} = \frac{1}{\alpha} (L_{l+u, l+u} + \alpha I)^{-1} \begin{pmatrix} \mathbf{y}_l \\ \mathbf{0}_u \end{pmatrix}$	$O((l+u)^3)$	$O((l+u)^2)$
GRF 提案手法	$\mathbf{f}_u = (D_{uu} - A_{uu})^{-1} A_{ul} \mathbf{y}_l$	$O(u^3)$	$O(u^2 + ul)$
(GP 予測分布)	$\mathbf{f}_q = L_{q, l+t} (L_{l+t, l+t} + \alpha I)^{-1} \begin{pmatrix} \mathbf{y}_l \\ \mathbf{0}_t \end{pmatrix}$	$O((l+t)^3)$	$O((l+t)^2 + q(l+t))$

表す指標となる．一方，語義曖昧性解消は，この分布仮説を直接の興味とし，文脈からどれだけ語の意味を予測できるか，を計測するタスクであり，しかも基本的な実験用データセットも整備されている．また，既存のラプラシアンラベル伝搬法の評価は，文献<sup>3)</sup>においては語義曖昧性解消のタスクで行われているため，提案手法を語義曖昧性解消のタスクで評価することは妥当である．

実験は，文献<sup>3)</sup>と類似した実験設定で行った．Senseval-3 の lexical sample data (S3LS) というデータセットを用いた．S3LS データセットでは，コーパス中の 1 パラグラフ程度のテキスト中に 1 語種，曖昧な語が含まれている．その語種の各出現に対して，語の意味（語義）のラベルが人手で振られており，そのラベルを自動的に予測ことがタスクの目的である．訓練データでは一回も現れず，テストデータだけに現れるようなラベルは予測することが出来ないため，S3LS データセットでは，訓練データとテストデータにそのようなラベルが存在しないように設定されている．

語義曖昧性解消のタスクでは，語によってクラス数が異なる多値分類問題になっている．しかし，特定の語を見る方法では全体としての性能を評価しにくい．そこで，本実験では，ある語の出現について，最頻出語義かそうでないかを予測する二値分類問題に帰着させて，提案手法を評価した．S3LS データセットでは，多くの語彙について，語義ラベルは多数あっても最もよく使われる語義がその語彙の出現の半数以上を占めている傾向が見られる．従って，この問題は，本来は多値分類であるが，ラベルに関して不均衡 (imbalanced) な設定になっている．パターン (素性) としては，曖昧な語の前後の 3 単語と，単語からの相対位置のあり/なしの組み合わせを用いた．

一つの語種について，1 つの Average Precision が求まるので，表 2 に S3LS データセット中の 56 語種に対する Average Precision の和を語種に対して平均した値 Mean average precision を示した．以後，簡単のため，この値を精度と呼ぶ．

この実験では，以下の方法を比較し，表 2 に示した．表 2 中の ALPHA は，GRF を除く手法のハイパーパラメータになっている  $\alpha$  の値である．

LLP ベースラインであり，既存のラプラシアンラベル伝搬法そのままである．ラプラシアンラベル伝搬法はガウス過程の MAP 推定に相当する (§3)．

GRF Zhu ら<sup>8)</sup>による Gaussian Random Field である．式は表 1 に示した．

$t = 0$  提案手法であり，表 1 中で  $t = 0$  の場合である．

$t = u$  提案手法であり，表 1 中で  $t = q$  の場合である．

表 2 より，提案手法である  $t = q$  は，全ての ALPHA を通じて，最も計算時間が長いものの最も精度が高いことが分かる．また， $t = 0$  は，ラプラシアンラベル伝搬法 LLP とほぼ同じ精度ながら，実行時間を LLP に対して半分以上に削減できていることが分かる． $t = 0$  と  $t = q$  を比較することにより，ラベルなしデータの数を増やすことによって，計算時間を犠牲にしつつも精度を向上させることが可能であることが分かる．より多くのラベルなしデータを用いることによって，より高い精度が期待できる．

実行時間は GRF が最も短いですが，これは，今回の実験設定では  $u$  の値が  $l$  の値と比較してそれほど大きくない実験設定であったためであると考えられる． $u$  の値が大きくなると，GRF は  $O(u^3)$  の計算時間がかかるのに対して，表 1 において提案手法では計算時間は  $O((l+t)^3)$  で抑えられているため，GRF の計算時間は提案手法より長くなるのが予想される．

## 6. 結論と今後の課題

本研究では，自然言語処理における代表的なブートストラップ法であるラプラシアンラベル伝搬法を，ガウス過程を用いて拡張する．近年，自然言語処理において，パターン (文脈) とインスタンス (語) のグラフ構造と少数のシード語から，シード語との意味的な関連

表 2 各種法の Mean Average Precision の比較

ALPHA	手法	Mean Average Precision	時間 (ms)
0.001	LLP	0.5570	11.9
	GRF	0.5952	1.3
	$t = 0$	0.5570	5.8
	$t = q$	0.6145	14.7
0.01	LLP	0.5569	12.7
	GRF	0.5947	1.3
	$t = 0$	0.5569	6.3
	$t = q$	0.6146	15.8
0.1	LLP	0.5568	13.0
	GRF	0.5947	1.4
	$t = 0$	0.5567	6.1
	$t = q$	0.6147	16.1
0.5	LLP	0.5562	13.2
	GRF	0.5947	1.5
	$t = 0$	0.5567	7.3
	$t = q$	0.6158	16.4
0.9	LLP	0.5568	12.1
	GRF	0.5947	1.3
	$t = 0$	0.5570	6.1
	$t = q$	0.6164	14.8
0.99	LLP	0.5568	12.6
	GRF	0.5947	1.3
	$t = 0$	0.5572	6.2
	$t = q$	0.6170	15.5
0.999	LLP	0.5568	12.6
	GRF	0.5947	1.4
	$t = 0$	0.5572	6.4
	$t = q$	0.6170	15.7

度を計算する半教師あり学習法であるブートストラップ法が注目されている。グラフを用いたブートストラップ法の中では、ラブラシアンラベル伝搬法は最も代表的な手法である。しかし、この手法は、関連度を求めるために全てのインスタンスを一辺とする正方行列の逆行列計算を扱わなければならない、空間計算量・時間計算量共に大きかった。本研究では、ラブラシアンラベル伝搬法がガウス過程の MAP (Maximum a posteriori) 推定の双対問題になっていることを示した。ガウス過程では、適切な設定の元では MAP 推定より判別性能が向上すると言われているベイズ推定による予測分布が知られており、インスタンスのスコアの予測問題をより自然に扱うことが出来る。本稿では、ラブラシアンラベル伝搬法と、

対応するガウス過程のベイズ推定を比較し、語義曖昧性解消のタスクにおいて判別性能の向上を確認した。

また、ガウス過程に帰着させたことにより、予測分布では、スコアの平均ベクトルだけではなく、スコアの分散共分散行列も計算できることが分かった。今後の課題としては、このスコアの分散共分散行列を、スコアの信頼度として活用することが挙げられる。

## 参 考 文 献

- 1) Amrudin Agovic. *Predictive Modeling using Dimensionality Reduction and Dependency Structures*. PhD thesis, University of Minesota, 2011.
- 2) G.Cawley, N.Talbot, and O.Chapelle. Estimating predictive variances with kernel ridge regression. *Machine Learning Challenges*, pp. 56–77, 2006.
- 3) Mamoru Komachi, Taku Kudo, Masashi Shimbo, and Yuji Matsumoto. Graph-based analysis of semantic drift in Espresso-like bootstrapping algorithms. In *EMNLP*, pp. 1011–1020, Honolulu, Hawaii, October 2008. Association for Computational Linguistics.
- 4) Mamoru Komachi, Shimpei Makimoto, Kei Uchiumi, and Manabu Sassano. Learning semantic categories from clickthrough logs. In *ACL-IJCNLP*, p. 189, 2009.
- 5) Patrick Pantel and Marco Pennacchiotti. Espresso: Leveraging generic patterns for automatically harvesting semantic relations. In *ACL*, pp. 113–120. Association for Computational Linguistics, 2006.
- 6) CarlEdward Rasmussen and Christopher K.I. Williams. *Gaussian processes for machine learning*. The MIT Press, April 2006.
- 7) Xiaojin Zhu. *Semi-Supervised Learning with Graphs*. PhD thesis, Carnegie Mellon University, 2005.
- 8) Xiaojin Zhu, John Lafferty, Lafferty Cs, and C.M.U. Edu. Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions. In *ICML*, 2003.