# 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，佐藤一誠，中川裕志**

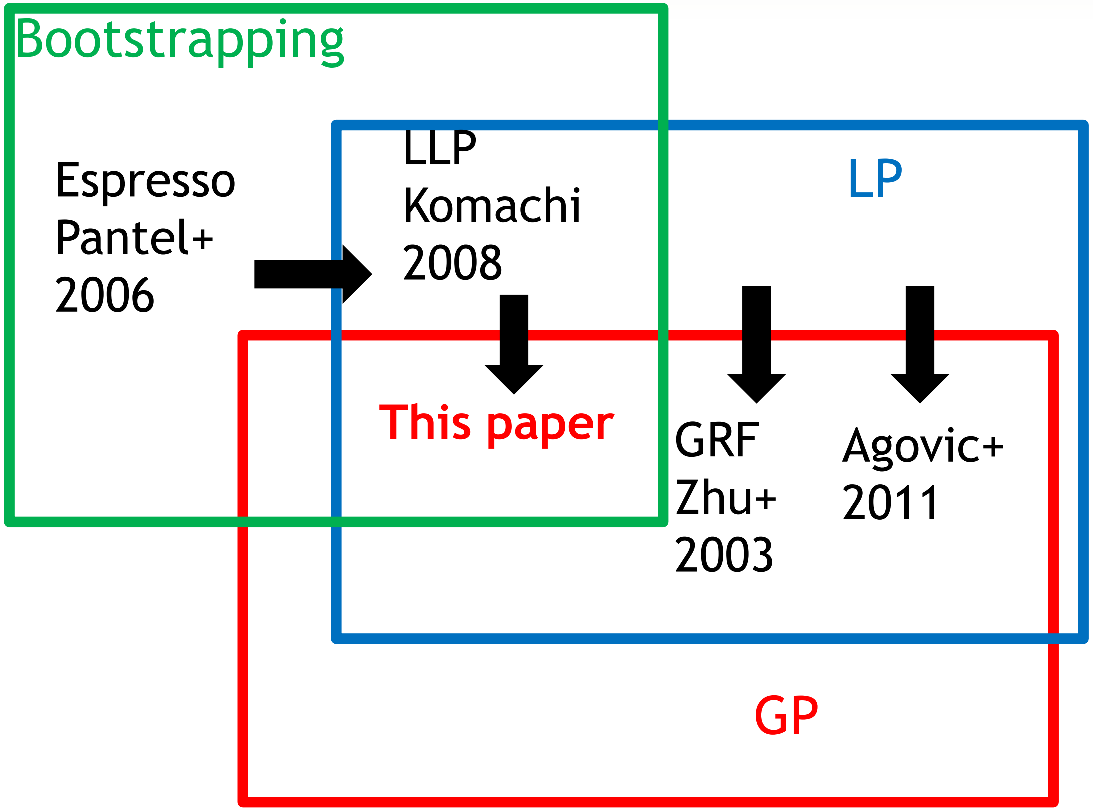
**東京大学**

概要：

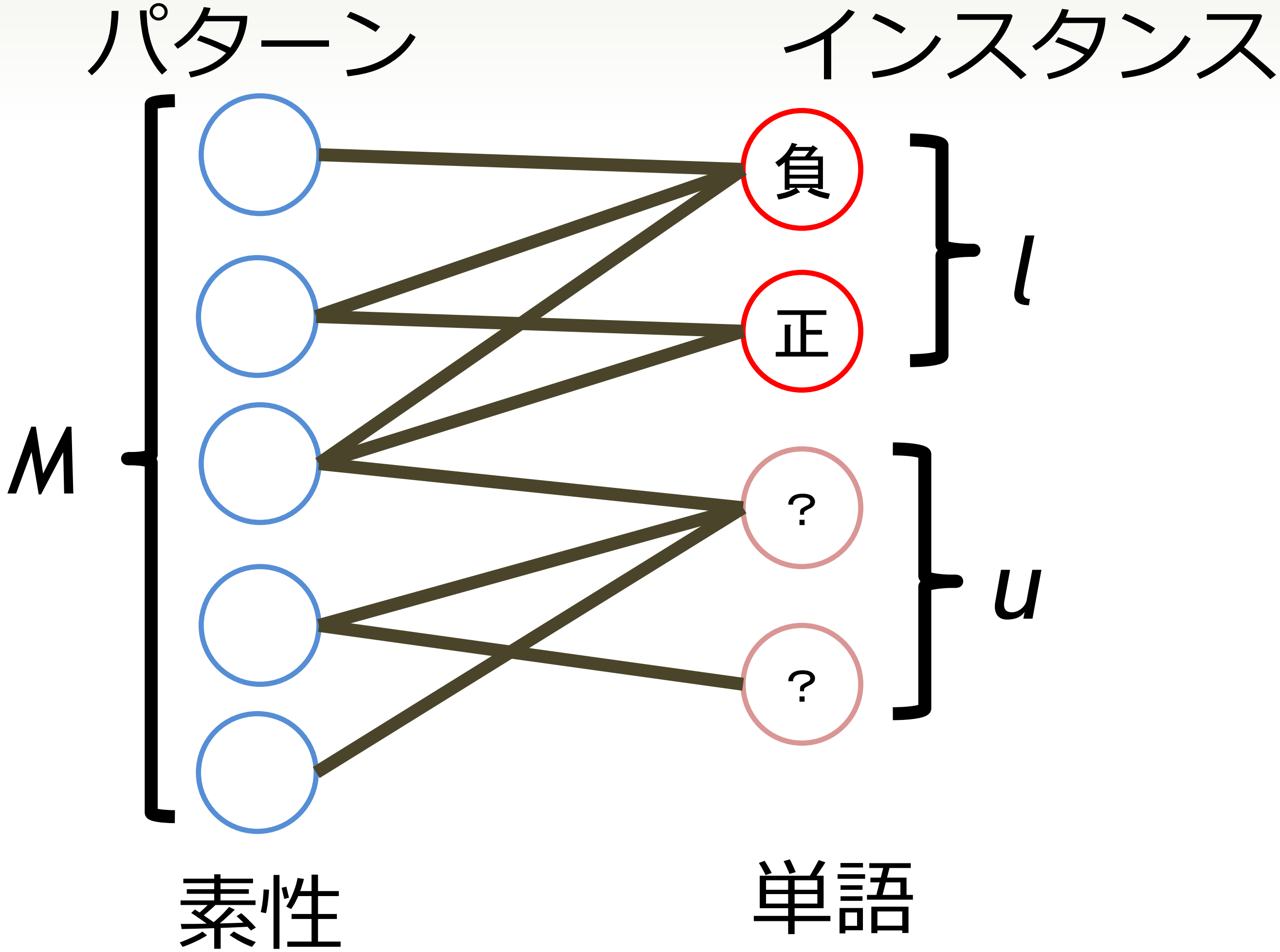
ブートストラップ法の１つである
ラプラシアンラベル伝搬法を,
ガウス過程を用いて,
- **予測分布を用いた自然な予測**
- **ベイズによる分類性能の向上**

ように拡張しました.

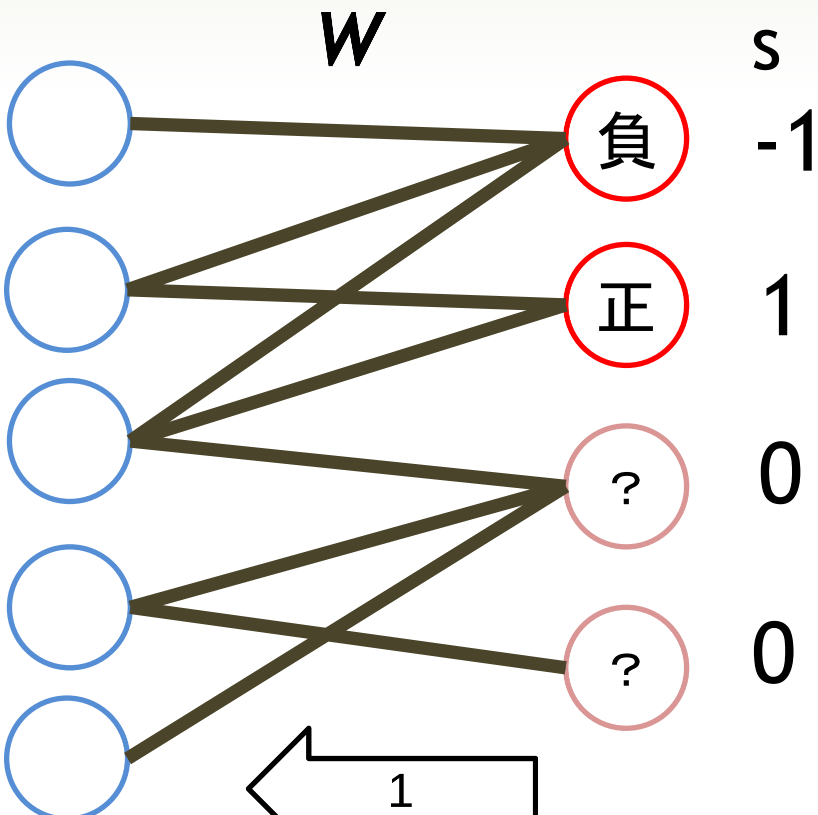
# 本研究の位置づけ


Bootstrapping
Espresso
Pantel+
2006
LLP
Komachi
2008
LP
This paper
GRF
Zhu+
2003
Agovic+
2011
GP

# 問題設定



# ラベル伝搬の行列表示



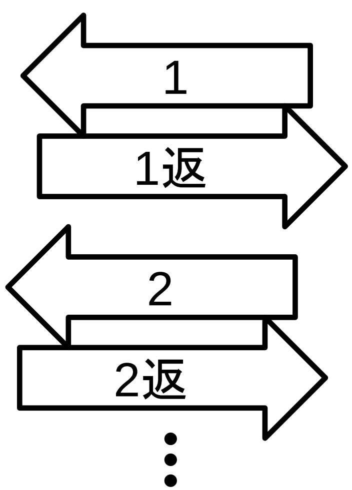

高い順に並べて
高いものほど
正例に近いと解釈

スコアベクトル（出力）

$$\mathbf{f} = \sum_{n=1}^{\infty} \alpha^n \left(WW^T\right)^n \mathbf{s}$$

$$= \left(I + \alpha\left(WW^T\right)\right)^{-1} \mathbf{s}$$

シードベクトル
（入力）




$$aW^TWs$$

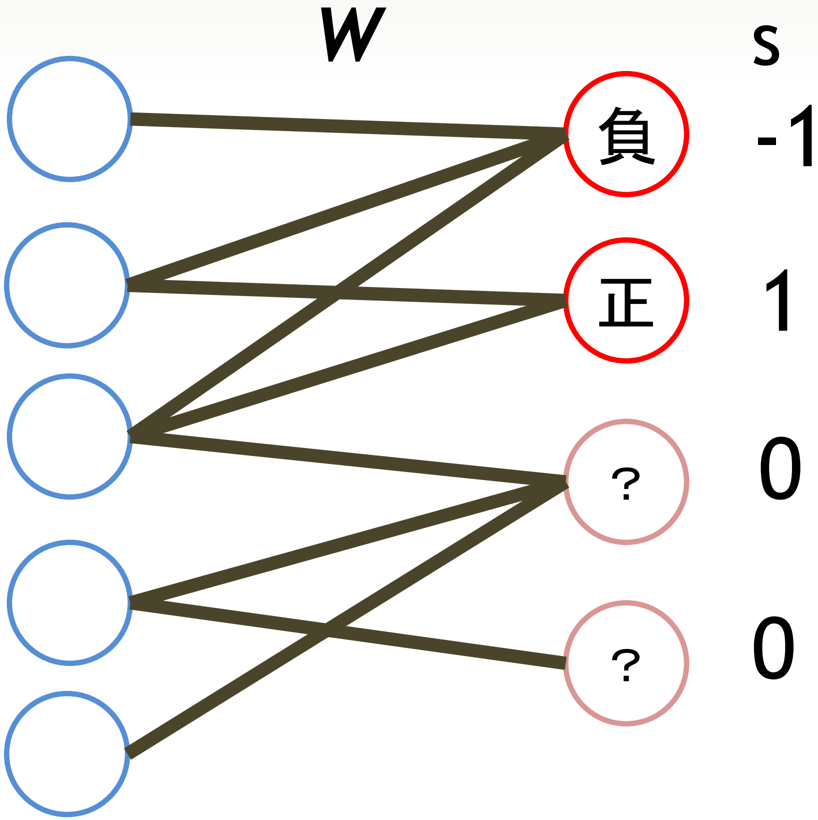
$$+$$

$$a^2(W^TW)^2\, s$$

$$\vdots$$

全部
足すと…

# ラプラシアンラベル伝搬法 (LLP)




$$\mathbf{f} = \sum_{n=1}^{\infty} \alpha^n \left( WW^T \right)^n \mathbf{s}$$

$$= \left( I + \alpha \left( WW^T \right) \right)^{-1} \mathbf{s}$$

WW$^T$の代わりに
<u>正則化グラフラプラシアン</u>
を使うと良い

$$L = I - D^{-1/2} WW^T D^{-1/2}$$

(Komachi+, EMNLP2008)

理由：　次の最適化問題の解になっている

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

(Zhu+, 2003)
(Agovic, 2011)

# 提案手法

従来：ラプラシアン
　　　ラベル伝搬法

$$\mathbf{f} = \boxed{\left(L+\alpha I\right)^{-1}}\mathbf{y}$$

(l+q)² 以上のサイズの逆行列．大きい&重い！

⇕ ラグランジュ双対

ガウス過程
MAP推定

$$\hat{\mathbf{w}} = \arg\max_{\mathbf{w}} p\left(\mathbf{w} \mid X_{l+q}, \mathbf{y}_{l+q}\right)$$

$$X_{l+q} = \{x_1, \ldots, x_l, x_{l+1}, \ldots, x_{l+q}\}$$

$$\mathbf{y}_{l+q} = \{y_1, \ldots, y_l, 0, \ldots, 0\}$$

提案：ベイズ推定

$$p\left(\mathbf{y}_q \mid X_q, X_{l+t}, \mathbf{y}_{l+t}\right) = \int p\left(\mathbf{y}_q \mid X_q, \mathbf{w}\right) p\left(\mathbf{w} \mid X_{l+t}, \mathbf{y}_{l+t}\right) d\mathbf{w}$$

LLP→ガウス過程のMAP推定に等価．適切な設定のもとでは
分類性能はMAP推定＜ベイズ推定（予測分布）

⇒**LLPの代わりに予測分布を使ってみよう**

# 分類性能はMAP推定＜ベイズ推定（予測分布）←なんで？

MAP推定（LLP）

$$\hat{\mathbf{w}} = \arg\max_{\mathbf{w}} p\left(\mathbf{w} \mid X_{l+q}, \mathbf{y}_{l+q}\right)$$

$\hat{\mathbf{w}}$という1つのパラメータだけを使用

ベイズ推定

$$p\left(\mathbf{y}_q \mid X_q, X_{l+t}, \mathbf{y}_{l+t}\right) = \int p\left(\mathbf{y}_q \mid X_q, \mathbf{w}\right) p\left(\mathbf{w} \mid X_{l+t}, \mathbf{y}_{l+t}\right) d\mathbf{w}$$

積分で全てのパラメータを足しこんでいる
たくさん寄れば文殊の知恵

# スコア不要のラベルなしデータ

既存の2手法では
全てのラベルなし
データに対する
スコアを求めようと
してしまうが…

| 手法 | 数式 |
| --- | --- |
| ラプラシアンラベル伝搬法 | $\begin{pmatrix} \boldsymbol{f}_l \\ \boldsymbol{f}_u \end{pmatrix} = \frac{1}{\alpha}\left(L_{l+u,l+u} + \alpha I\right)^{-1} \begin{pmatrix} \boldsymbol{y}_l \\ \boldsymbol{0}_u \end{pmatrix}$ |
| GRF | $\boldsymbol{f}_u = \left(D_{uu} - A_{uu}\right)^{-1} A_{ul}\boldsymbol{y}_l$ |
| 提案手法（GP 予測分布） | $\boldsymbol{f}_q = L_{q,l+t}\left(L_{l+t,l+t} + \alpha I\right)^{-1} \begin{pmatrix} \boldsymbol{y}_l \\ \boldsymbol{0}_t \end{pmatrix}$ |

*u*=*q*+*t*
提案手法では
全ラベルなしデータ
*u*のうち，スコアを
求めたい部分*q*
とスコアは不要で
単に半教師として
入れたい*t*を明示的に
分けて扱える

| 時間計算量 | 空間計算量 | スコア信頼度 |
| --- | --- | --- |
| $O((l+u)^3)$ | $O((l+u)^2)$ | なし |
| $O(u^3)$ | $O(u^2+ul)$ | なし |
| $O((l+t)^3)$ | $O((l+t)^2 + q(l+t))$ | 数式（4） |

# 実験１：Word Sense Disambiguation

- which you step on to <head>**activate**</head> it . Used
- parts of the sensory system are <head>**activated**</head> : stimulation of the retinal
- the number of neurones <head>**activated**</head> ( the phenomenon of

下線の単語を特徴として，太字の単語の意味が，その単語の最頻出語義か否かを当てる（２値分類）

AP=average precision（正解だった順位でのprecisionの平均）

| | |
|---|---|
| activate | AP xx.xx% |
| bank | AP xx.xx% |
| … | |

平均した精度
(Mean average precision)
を表示

単語間の類似度はコサイン類似度

# 実験１：結果

LLP: Laplacian label propagation
t=0, t=qが提案.

| ALPHA | 手法 | Mean Average Precision | 時間 (ms) |
|---|---|---|---|
| | LLP | 0.5570 | 11.9 |
| 0.001 | GRF | 0.5952 | 1.3 |
| | $t = 0$ | 0.5570 | 5.8 |
| | $t = q$ | 0.6145 | 14.7 |
| | LLP | 0.5569 | 12.7 |
| 0.01 | GRF | 0.5947 | 1.3 |
| | $t = 0$ | 0.5569 | 6.3 |
| | $t = q$ | 0.6146 | 15.8 |
| | LLP | 0.5568 | 13.0 |
| 0.1 | GRF | 0.5947 | 1.4 |
| | $t = 0$ | 0.5567 | 6.1 |
| | $t = q$ | 0.6147 | 16.1 |

t=0の時，LLPと
同程度の性能で計算時間
半分.

t=qの時に，LLPと
同程度の時間が掛かるが
最高精度.

# まとめ

- ガウス過程を用いてラプラシアンラベル伝搬法を拡張した
  1. 予測分布を用いた予測手法の提案
  2. ベイズによる分類性能向上効果
     - 語義曖昧性解消のタスクで確認

- 今後の課題：
  分散を信頼度として活用する

# 今後の課題２：
# いい実験設定を探しています

- 多値判別
  - WSDで評価するなら，多値判別にすべき
    - 基本的にはone-versus-the-restで.
  - 実験データセットは何が適切？
    - S3LS datasetは1語あたりのデータが少ないので，line / interest datasetにして評価すべきか
  - (小嵜+, NL196)を読んで実験設定を合わせようとしている最中
- ２値判別
  - Espresso (Pantel+, 2006)がbaseline
  - 要素が0-1の行列Wに対してEspressoを行うとPMI=p(a,b)/p(a)p(b)のp(a,b)が定数になってしまいfairではないので，違う実験設定でやるべき？

# 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