

量子化による測定結果への影響

城 真範, 赤穂 昭太郎

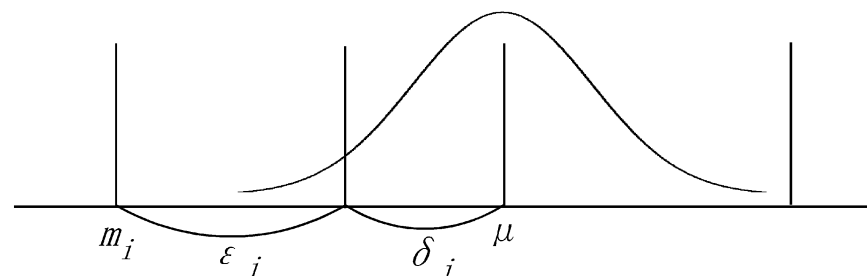
(産総研ヒューマンライフテクノロジー研究部門)

デジタル機器を使って測定する際には、測定に伴うアナログ的な誤差のほかに、データを数値に丸め込む際の量子化の影響も考慮すべきである。もし、量子化窓の大きさが十分に小さければアナログ的な誤差のみが寄与するが、量子化窓の大きさが大きいとアナログ的な誤差と量子化に伴う値のずれは相関を持つ。今回アナログ的な誤差を正規分布と仮定し、量子化窓の大きさと二乗誤差の関係をモンテカルロ法を使って調べた。その結果、アナログ的な誤差と量子化窓の大きさが拮抗するところに、トータル誤差の最小値が見いだされた。これは、アナログ的な誤差が既知であるとき、最適な量子化幅が存在することを示している。

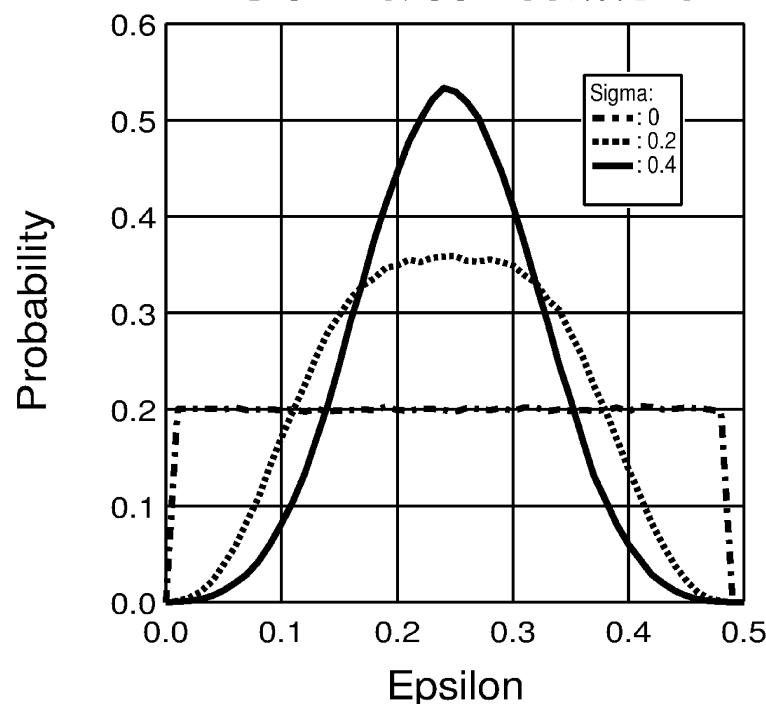
真値： μ
アナログ観測値： ν
デジタル観測値： m

アナログ誤差： δ
量子化誤差： ε

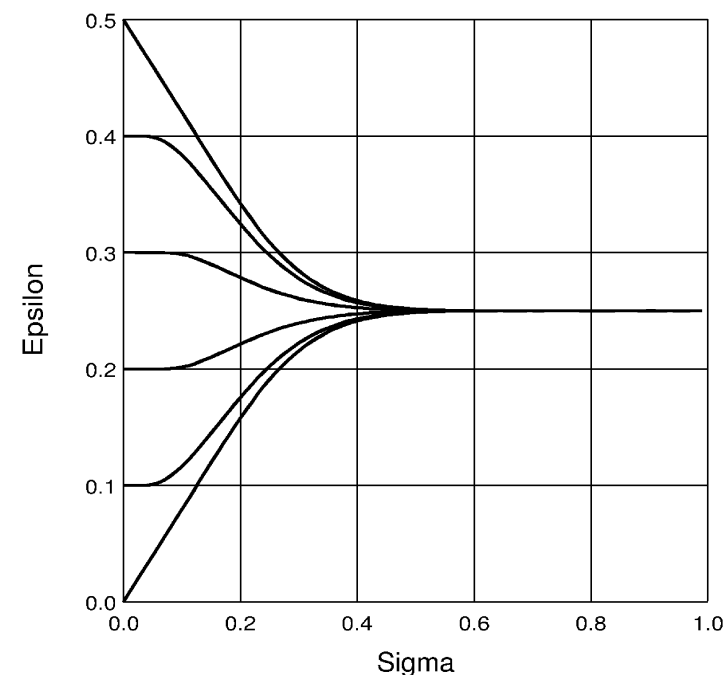
$$\nu = \mu + \delta, \quad m = \nu + \varepsilon$$



アナログ誤差(正規分布仮定)と
量子化の影響の合成分布



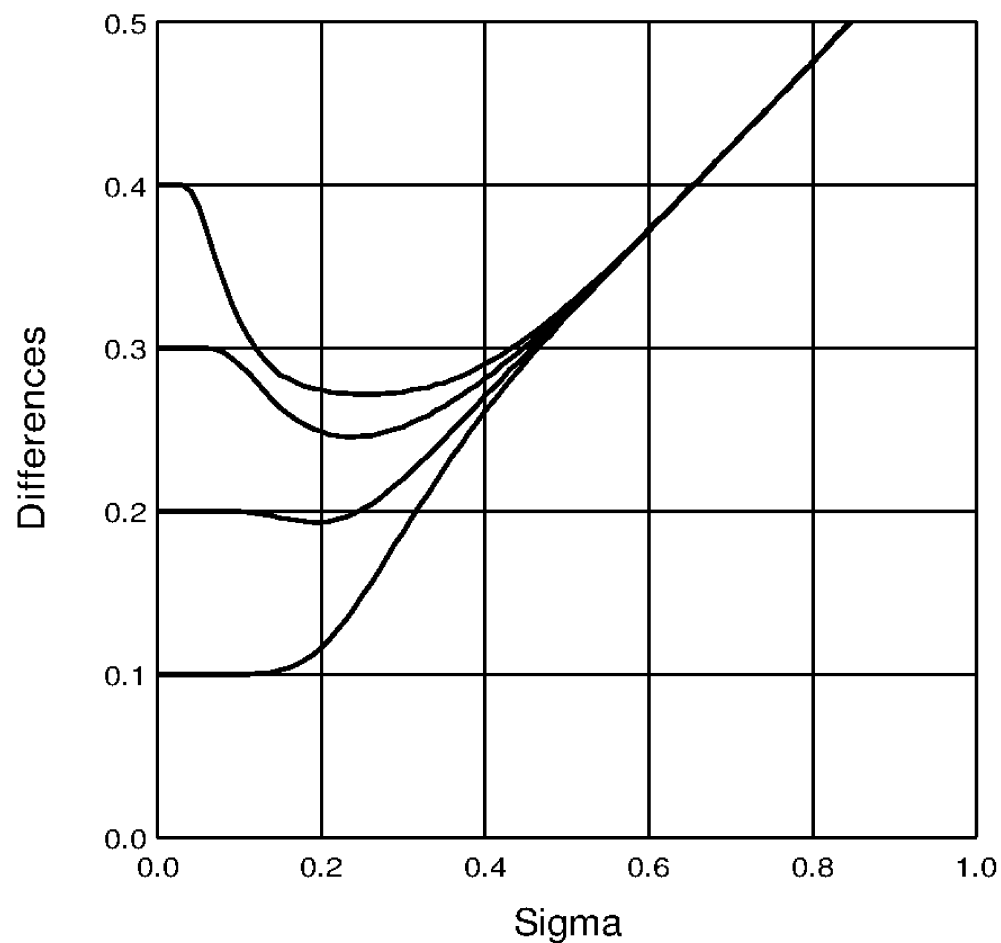
アナログ誤差の大きさと
量子化の大きさの期待値



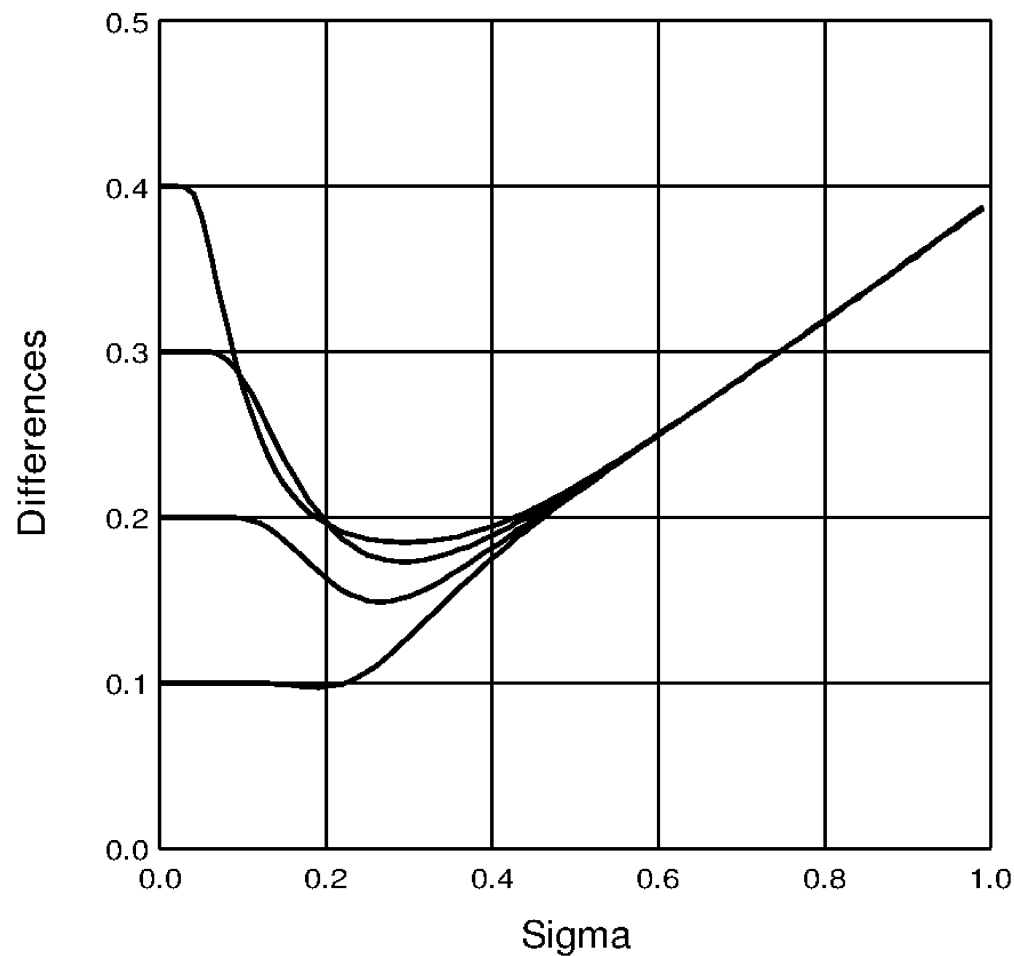
n 個のサンプルの平均と真値の違い

(二つの効果の拮抗による谷の出現)

n=4



n=8



High-Dimensional Feature Selection by Feature-Wise Kernelized Lasso

山田 誠 (NTT CS研), Wittawat Jitkrittum (東工大), Leonid Sigal (Disney Research),
Eric P. Xing (CMU), 杉山 将 (東工大)

■ 特徴選択

不要で冗長な特徴量をデータから除去する

■ 高次元データ

次元数 d がサンプル数 n よりもはるかに大きい (例: マイクロアレイデータ)

特徴選択手法	非線形性	最適化	高次元データ
Lasso Tibshirani. (JRSSB 2005)	×	凸	○
FVM Li et al. (NIPS 2005)	○	非凸	×
SpAM Ravikumar et al. (NIPS 2008)	△	凸	○
提案法	○	凸	○

高次元かつ非線形関係のある特徴量を高速に選択する手法を提案！

提案手法 (HSIC Lasso):

高次元データを扱うためスパース性を利用

$$\min_{\alpha \in \mathbb{R}^d} \quad \frac{1}{2} \left\| \bar{\mathbf{L}} - \sum_{k=1}^d \alpha_k \bar{\mathbf{K}}^{(k)} \right\|_{\text{Frob}}^2 + \lambda \|\alpha\|_1$$

$\bar{\mathbf{K}}^{(k)}$: k番目の特徴量のデータのグラム行列
 $\bar{\mathbf{L}}$: 出力データのグラム行列

s.t. $\alpha_1, \dots, \alpha_d \geq 0$

統計的解釈 (第一項): $\text{HSIC}(\mathbf{u}_k, \mathbf{y}) = \text{tr}(\bar{\mathbf{K}}^{(k)} \bar{\mathbf{L}})$: 独立 $\Leftrightarrow$ HSIC = 0

$$\frac{1}{2} \text{HSIC}(\mathbf{y}, \mathbf{y}) - \sum_{k=1}^d \alpha_k \text{HSIC}(\mathbf{u}_k, \mathbf{y}) + \frac{1}{2} \sum_{k,l=1}^d \alpha_k \alpha_l \text{HSIC}(\mathbf{u}_k, \mathbf{u}_l)$$

出力と特徴kが**従属**だと大 特徴kと特徴lが**独立**だと小

出力に従属かつ非冗長な特徴が選択される!

最適化 (非負制約付きLasso):

$$\min_{\alpha \in \mathbb{R}^d} \quad \frac{1}{2} \left\| \text{vec}(\bar{\mathbf{L}}) - [\text{vec}(\bar{\mathbf{K}}^{(1)}), \dots, \text{vec}(\bar{\mathbf{K}}^{(d)})] \alpha \right\|_2^2 + \lambda \|\alpha\|_1$$

s.t. $\alpha_1, \dots, \alpha_d \geq 0$ 本研究ではDAL(<http://www.ibis.t.u-tokyo.ac.jp/ryotat/dal/>)を使用

Matlab code: <http://www.kecl.ntt.co.jp/icl/lis/members/myamada/software/HSICLasso.zip>

太陽光発電量の予測

楠本 英子(1) 樋口 知之(2) 伊庭幸人(2)
(1)総合研究大学院大学 (2)統計数理研究所

前日の天気予報
+
当日9時までの日照量

予測

線形回帰

発電量

天気予報を13個の
カテゴリーに分類

平均の総発電量
の多い順に並べる

4つに分ける

A~Gの8種類に
グルーピングを行う

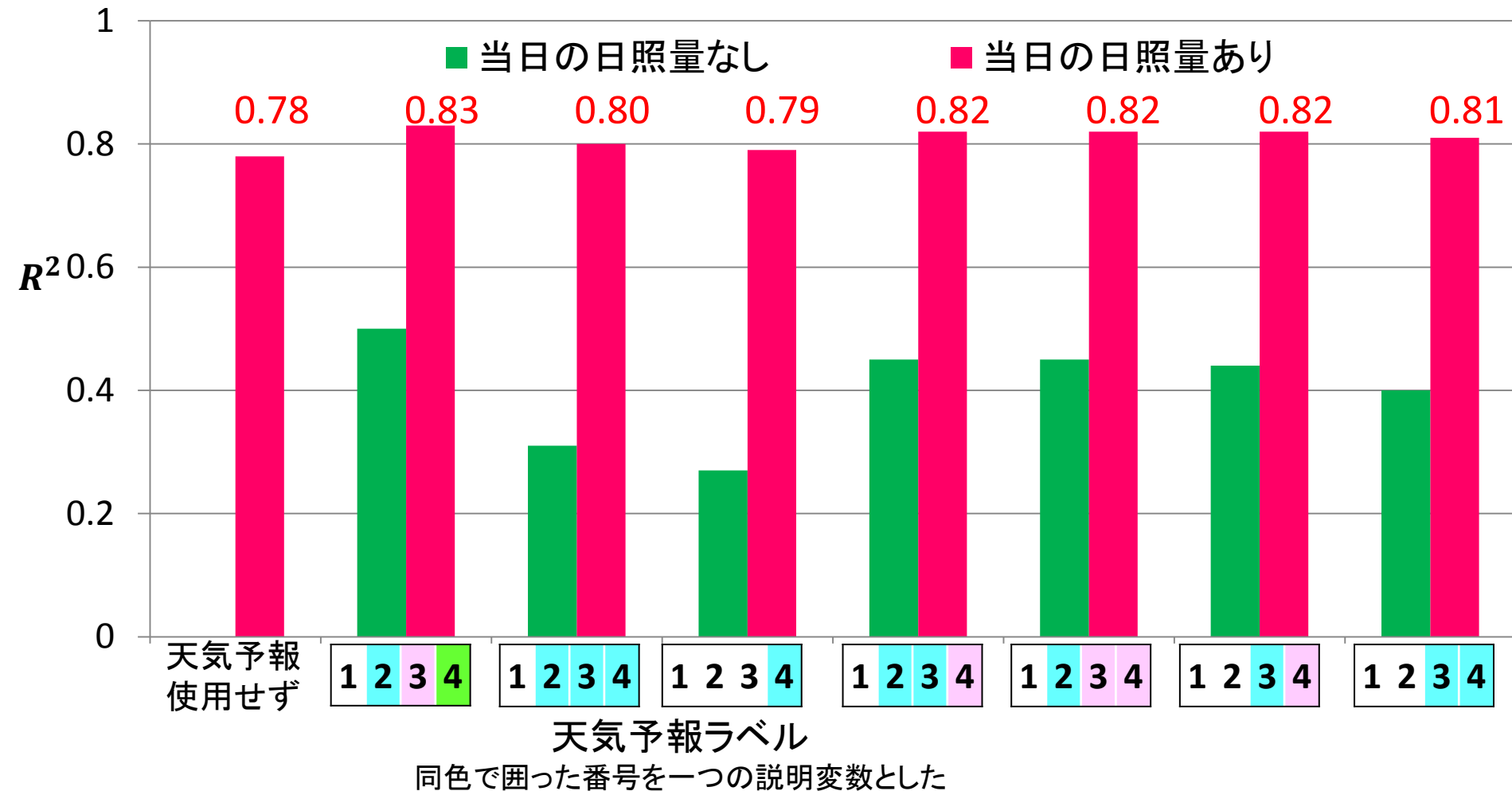
説明変数として、
訓練データに対して
回帰式を作る

実証データに
回帰式をあて
はめる

奇数番目
訓練データ
偶数番目
実証データ

- ①天気予報のみ
- ②当日の日照量も含めた場合

結果 実証データに対する R^2 の比較

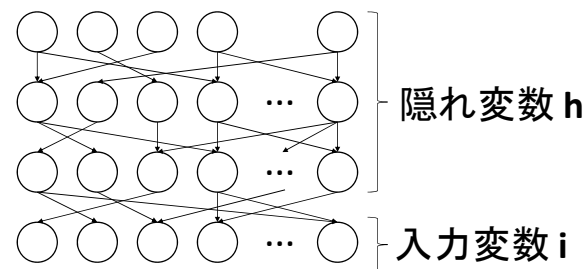


朝の日照量を説明変数に加えると R^2 は大きくなり予測精度は良くなる。また、朝の日照量を使う場合も、天気予報をいれることにより R^2 は0.78から0.83へと改良している。

D-27

大脳皮質モデルと deep learning を統合した認識・学習の線形時間アルゴリズム 一杉裕志(産総研)

- ◎ 計算論的神経科学における「大脳皮質ベイジアンネット仮説」に基づいた脳の認識・学習アルゴリズムのモデル。
- ◎ 1入力を処理する計算量は(脳のように)ノード数にほぼ比例。(スパース性を仮定。)
- ◎ 「deep learning」と同じ構造。



認識ステップ:

入力 $\mathbf{i}(t)$ との同時確率が最大となる隠れ変数の値 $\hat{\mathbf{h}}(t)$ を推定

$$\hat{\mathbf{h}}(t) = \arg \max_{\mathbf{h}} P(\mathbf{h}, \mathbf{i}(t) | \theta(t))$$

近似 belief revision algorithm を使用

学習ステップ:

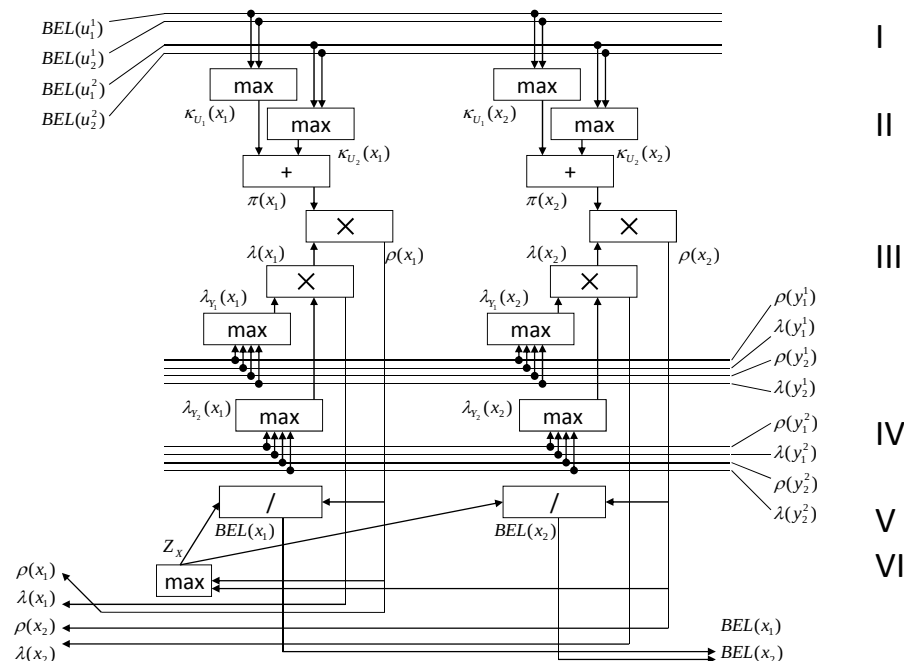
隠れ変数の推定値を真の値と見なし、パラメタ $\theta(t+1)$ をMAP推定

$$\theta(t+1) \approx \arg \max_{\theta} \left[\prod_{i=1}^t P(\hat{\mathbf{h}}(i), \mathbf{i}(i) | \theta) \right] P(\theta)$$

counting を使用

認識ステップ:

- ◎ベイジアンネットとして動作。
- ◎近似 belief revision を使用。
- ◎エッジの数が定数ならば、1ステップあたりの計算量はノード数にほぼ比例。



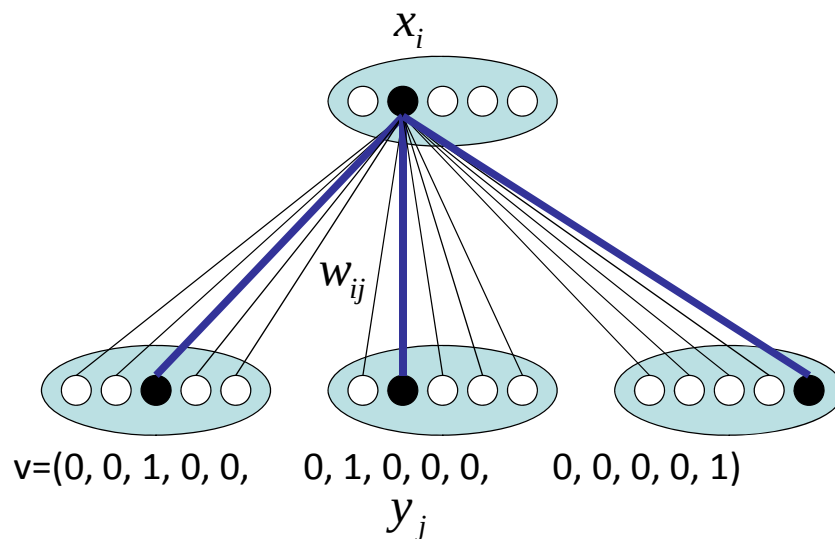
認識アルゴリズムを実行する神経科学的に妥当な回路

学習ステップ:

- ◎各ノードは自己組織化マップとして動作。
- ◎学習結果は条件付確率に収束。

$$w_{ij} \leftarrow w_{ij} + \alpha_i (v_j - w_{ij})$$

$$w_{ij} = P(y_j | x_i)$$



各ノードは子ノードからの入力を学習

タイピング学習システム

有賀功(中央大学)

- 概要

タイピングを効率的に学習するために、被験者が文字をミスタイプする確率を推定し、出題単語を選択する手法の提案

- 提案手法

1. アライメントを用いたミスタイプ箇所の推定
 - ・ ミスタイプした箇所を決定する必要がある
2. 最尤推定法を用いたミスタイプ確率の推定
3. 出題単語の選択
 - ・ 多腕バンディット問題で用いられる手法
UCB1, UCB1-tuned, ϵ -greedy法など

タイピング学習システム

有賀功(中央大学)

• アライメント

出題単語 “quickly” に対して回答文字列が “qucehkly” の時

quickly
qucehkly

ミス
6回

quic__kly
qu_cehkly

ミス
3回

ミス回数最小のアライメントを選択する

- 多腕バンディット問題 (N台のスロットマシンがある時に、どのマシンにコインを投資するか?)

スロットマシン → アルファベット

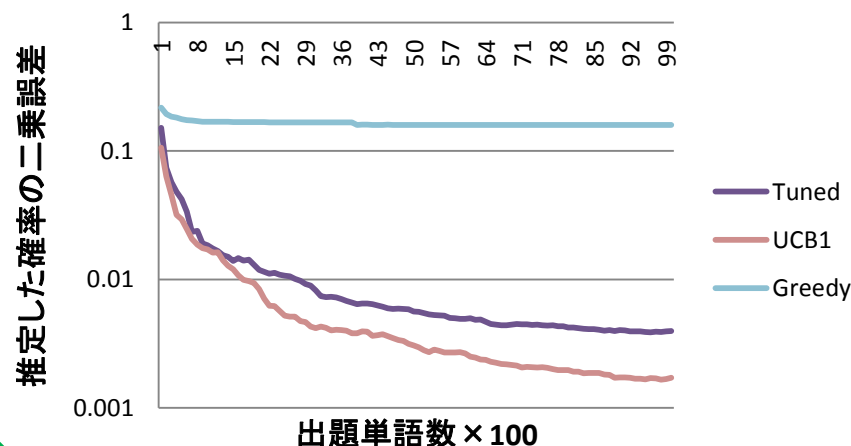
コイン数 → 打鍵数

当選確率 → ミスタイプ確率

と置き換えると出題単語選択問題は
多腕バンディット問題に置き換えられる

解戦略: Random, Greedy, UCB1,

重みづけ手法による二乗誤差



実験結果

耐ノイズ性を有する高速多変量オンライン ノンパラメトリック密度推定法

中村圭宏, 高橋大志, 長谷川 修(東京工業大学)

密度推定において

- ・ 高速オンライン学習
- ・ モデルを仮定しない
- ・ 耐ノイズ性



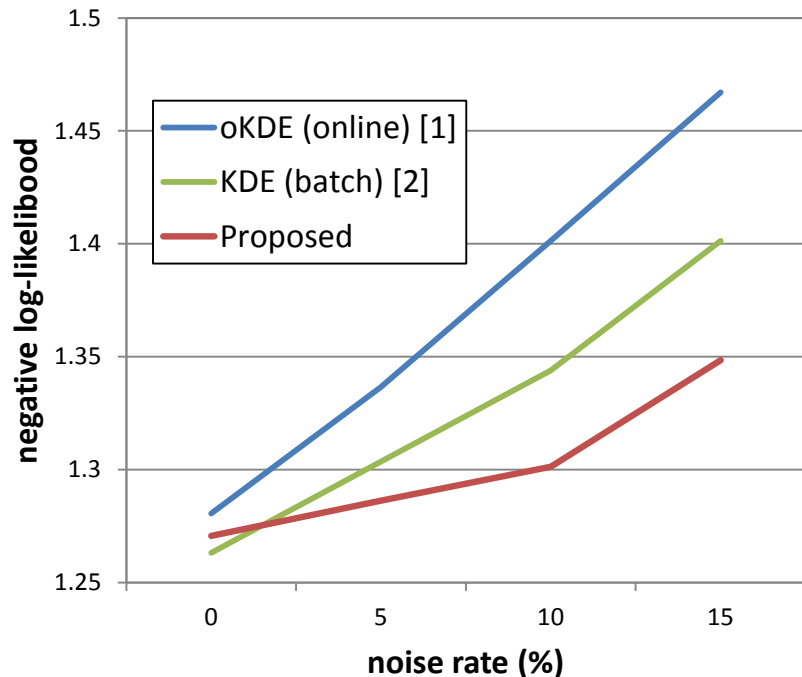
SOINN

+

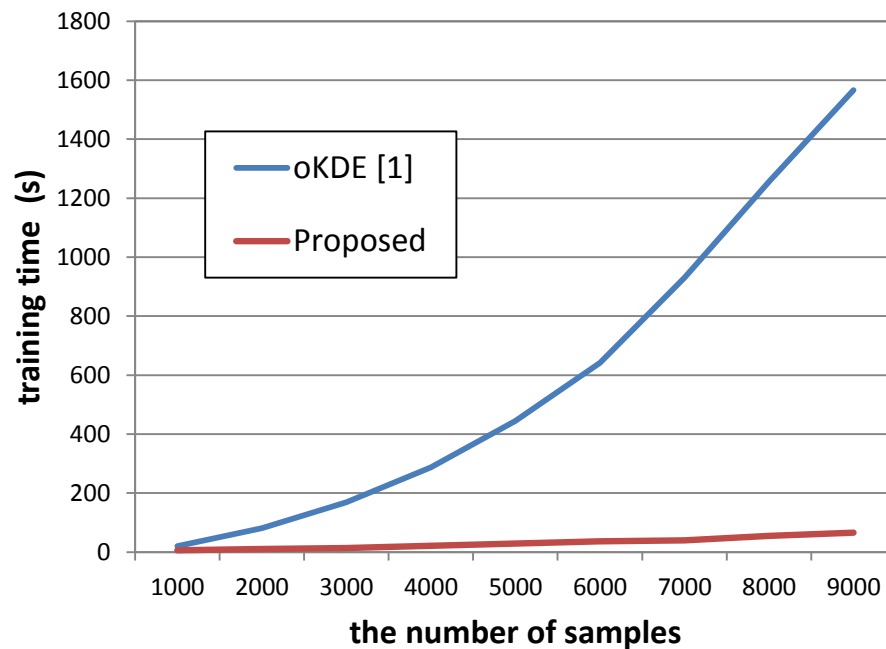
カーネル密度推定

評価実験

学習データにノイズを混合させた場合の精度の変化



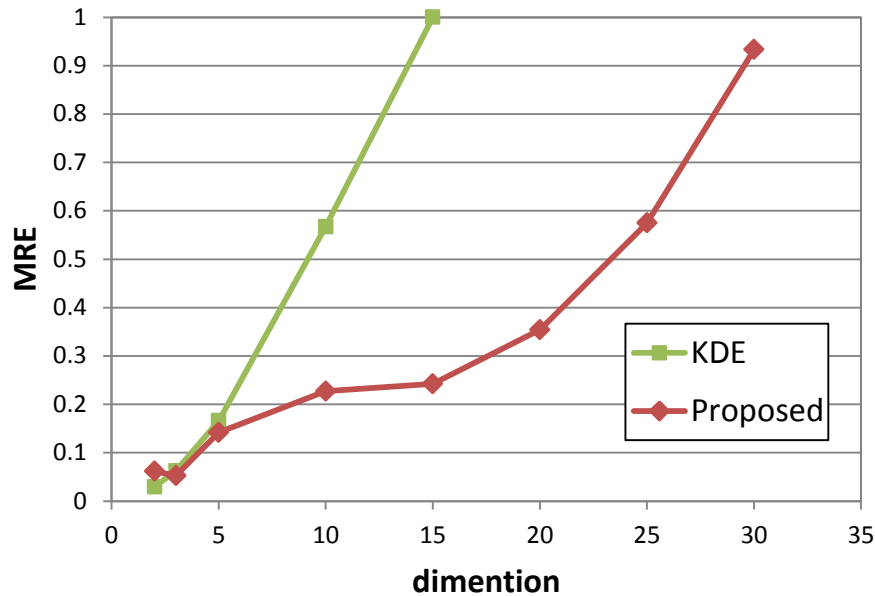
学習サンプル数に対する学習時間の変化



[1] M. Kristan, et al., Pattern Recognition, 2011.

[2] Murillo et al., Neural Information Processing Systems, 2008.

サンプル数固定の下で次元を増加させた場合の推定精度の変化

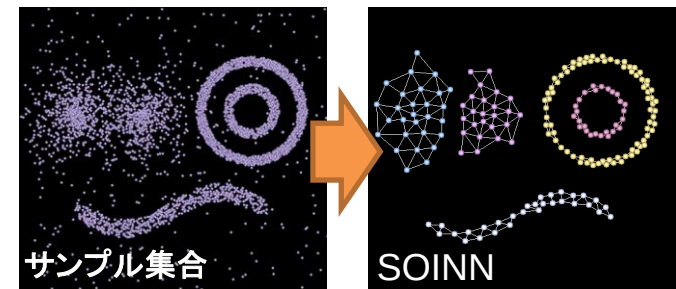


カーネル密度推定法

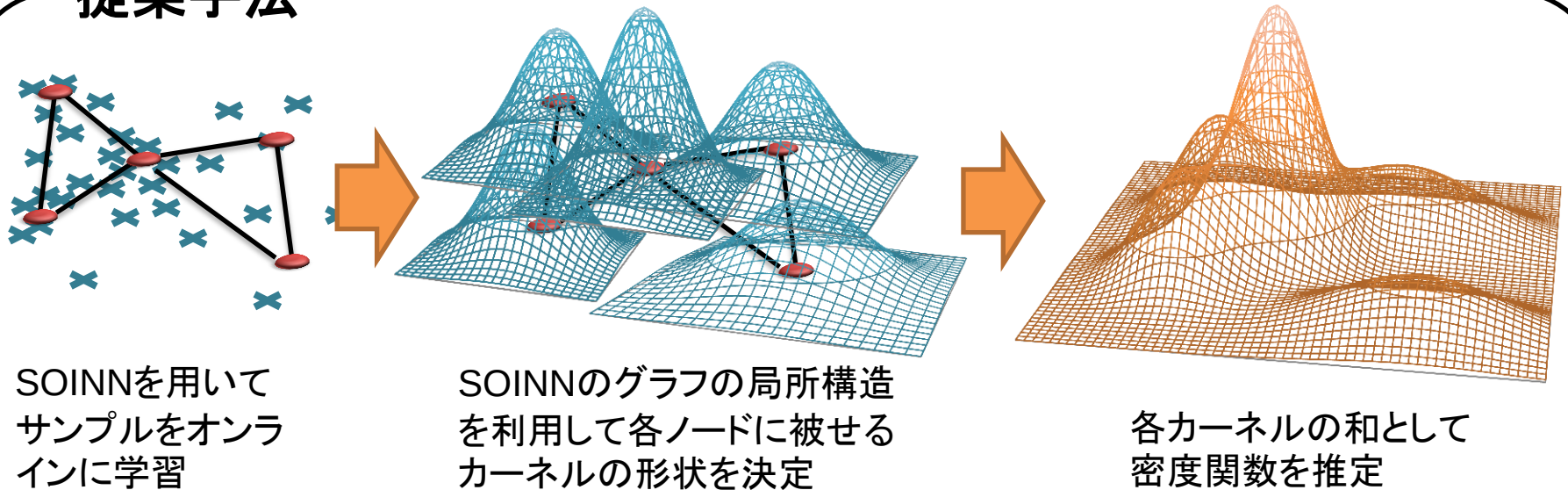
- 耐ノイズ性がない
- 高次元になると精度が低下

SOINN

- 教師なしオンライン追加学習手法
- 耐ノイズ性あり
- 競合学習によりデータをネットワーク構造として近似



提案手法



D-30

複数クラスラベル付きバイナリーデータ群からの相関抽出方法と 投薬データ解析への応用

梶大介(コニカミノルタエムジー), 山崎啓介(東工大)

【背景】

診療行為と傷病名の関係を抽出することは、医療情報分野の重要な研究課題の1つである

【目的】

複数の傷病名とそれらに対応する処方が記述されたカルテから各傷病に対する処方の分布を推定するための基礎モデルを作る

【カルテ 1】

病名 A
病名 C

投薬1
投薬2
投薬4
投薬5
投薬7

【カルテ 2】

病名 D
投薬4

...

【カルテ N】

病名 C
病名 Z

投薬 1
投薬 3
投薬 7
投薬 11
投薬 15



	病気A	病気B	...	病気Z
投薬 1	20.2%	0.3%		92.1%
投薬 2	2.1%	0.0%		45.7%
...				
投薬 N	10.5%	67.2%		0.6%

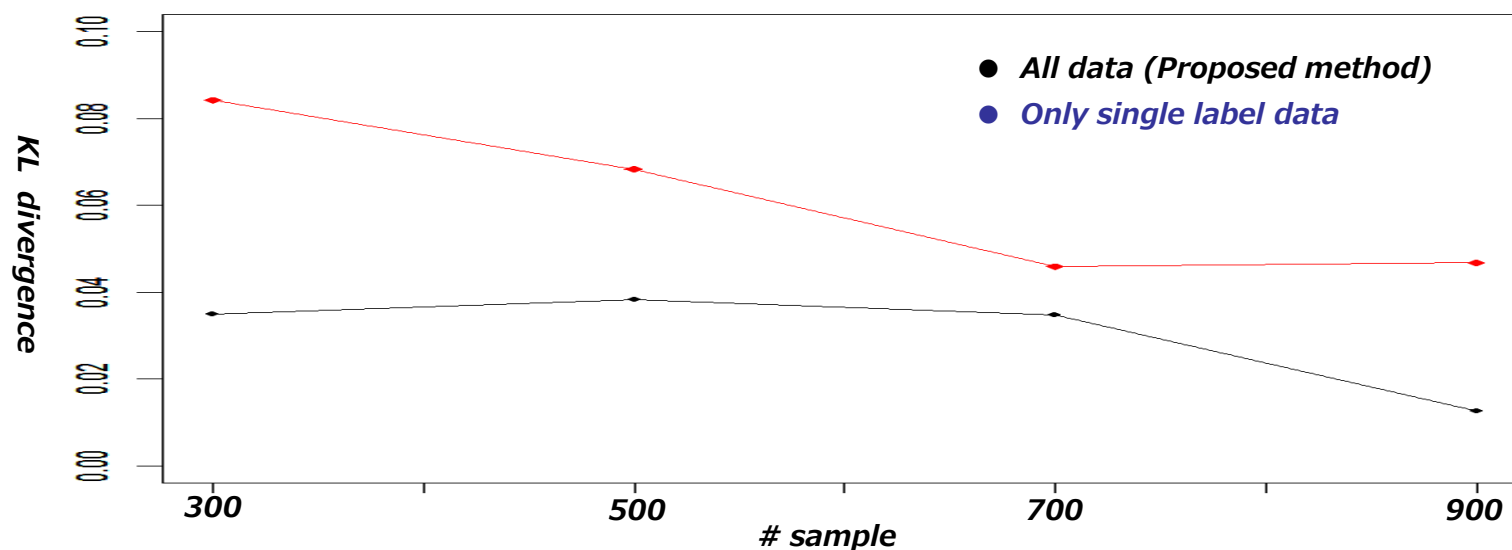
D-30

複数クラスラベル付きバイナリーデータ群からの相関抽出方法と投薬データ解析への応用

梶大介(コニカミルタエムジー), 山崎啓介(東工大)

【結果】

- ✓ 複数ラベル(傷病名)をもつデータ(処方)群から各ラベル別のデータ分布の推定するための確率モデルを提案
- ✓ 提案手法はシングルラベルのデータのみを用いた場合より優れることを理論的に証明
- ✓ 上記優位性をテストデータを用いた実験により確認



D-31

線型判別分析のための教師付き スパース共分散推定

日野英逸（早大），Reyhani Nima(アールト大)

概要

Sparse Inverse Covariance Selection(SICS)において、
目的が線型判別分析の経験共分散行列の置き換えである
とき，教師情報を利用して判別性の高い共分散構造の学
習が可能であるか検討する．

- * $p \gg n$ の状況では、経験共分散行列は一致性を持たない：

$$p/n \xrightarrow{n \rightarrow \infty} \gamma < \infty \quad l_1 \rightarrow (1 + \sqrt{\gamma})^2 \quad a.s.$$

- * naive Bayes, glasso といったスパース性を仮定した共分散推定量で経験共分散行列を置き換えることで、LDAの精度が大幅に向上することが知られている。
- * 経験リスク最小化の枠組みで、教師情報を利用したSICSを提案。凸計画問題として定式化可能。

Bayes 統計学への SOINN の導入に関する一考察

長谷川修（東工大）

SOINN とは、東工大長谷川研で独自に研究開発したニューラルネット技術です。

- SOINN の安定性の向上
- 変分ベイズ問題への SOINN の適用
- 階層ベイズ+MCMC 問題への SOINN の適用
- SOINN によるオンライン・インクリメンタル・ベイジアンネットの可能性

について議論します。

本発表の目的

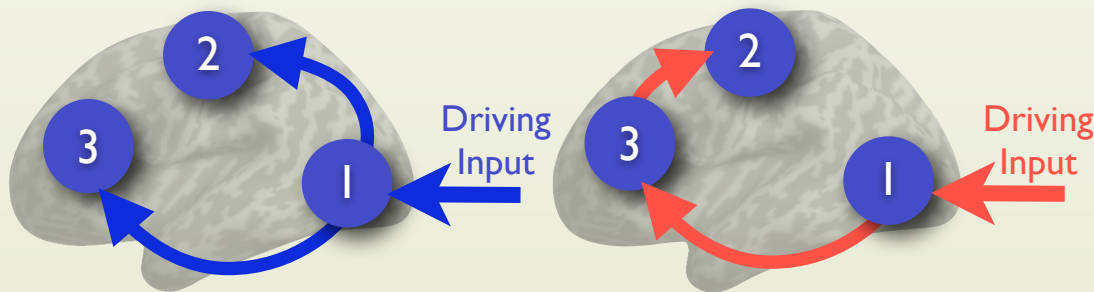
本研究は、認知神経科学上の仮説を検証するために行った。その点で、領野間ネットワーク解析について様々な手法を比較検討した研究ではないが、領野間ネットワーク解析の適用例を発表し、その方法・結果・言及できる科学的解釈について議論する。

fMRI研究における領野間ネットワーク解析の意義

活動している領野 (1, 2, 3) は同じだが、異なるネットワークで実現する認知情報処理がある (下記の図)。認知的差分法では、この異なる2つの認知情報処理を区別できない。

⇔ 認知的差分法

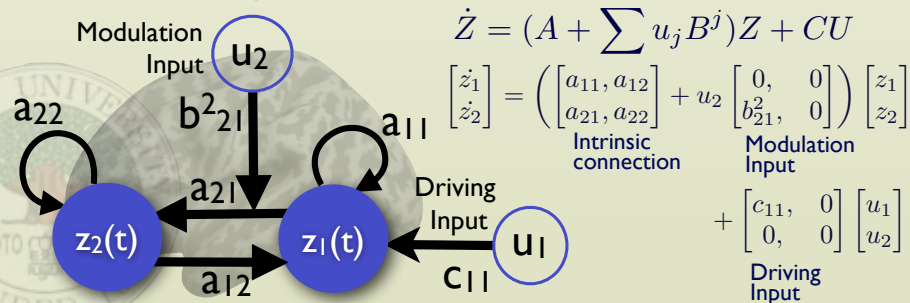
- ・ 古典的fMRI研究の解析方法
 - ・ 脳活動の大きさを異なる条件間で比較し、大きさの異なる領野を同定
→ 同定された領野が認知機能に重要
- ・ 機能局在論にもとづく解析方法
 - ・ 特定の認知情報処理は、特定の脳領野が担うという仮説



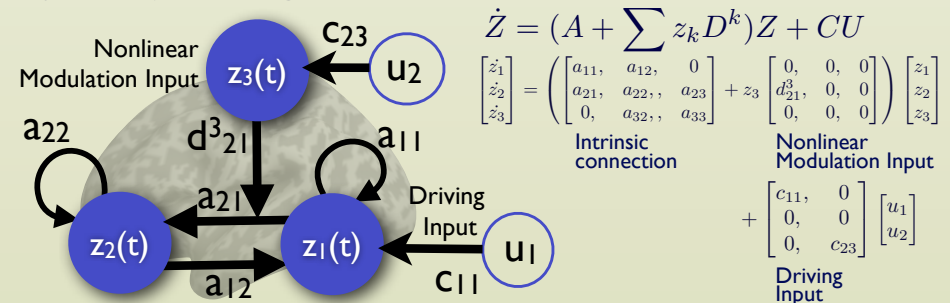
Dynamic Causal Modeling | DCM (Friston et al., 2003)

本研究で適用したネットワークのモデル。Z: 脳領野の神経活動, U: 脳外部からの刺激, 実験条件など。

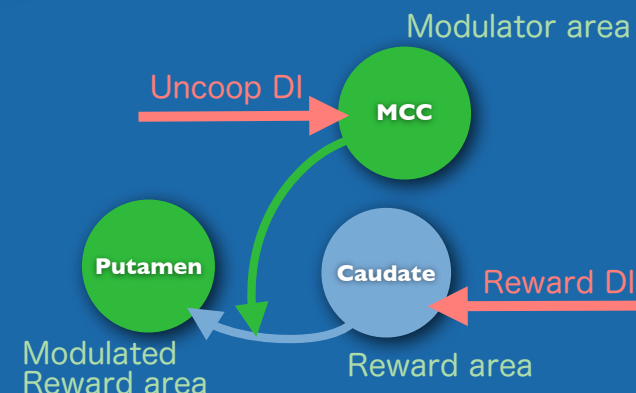
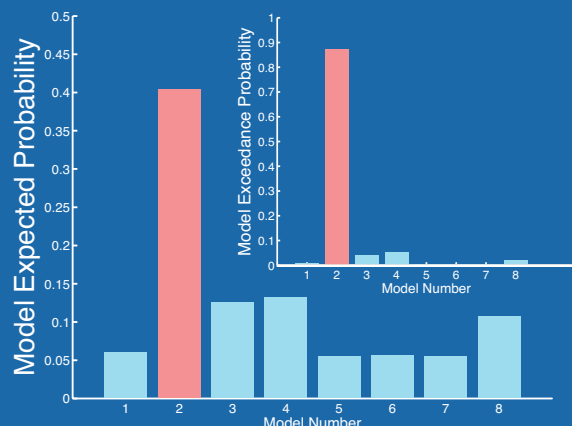
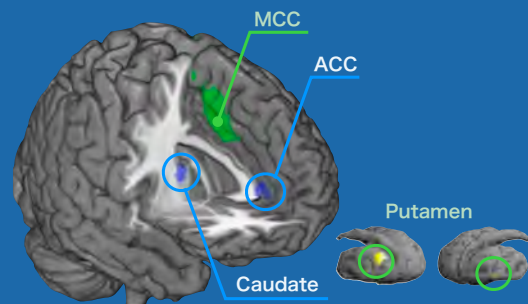
・ Bilinear DCM



・ Nonlinear DCM



解析手順 および 得られた認知神経科学上の結果



1. 関連領野の同定

ネットワーク解析の対象とする領野群を同定した。

2. パラメータ推定, モデル選択

仮説となりえる複数（ここでは8つ）のネットワークモデルについてパラメータを推定し, モデル選択をおこなった。

3. 得られた結果

相手の非協調性を判断する大脳辺縁系の領野が, 報酬関連領野の結合性を変えるという結果を得た。

fMRI研究における領野間ネットワーク検証にかかる諸問題 ～ 認知神経科学のニーズ ～

上記の研究報告を例に, 領野間ネットワーク検証にかかる下記の一般的な問題などについても議論したい。

不定性

推定すべきパラメータが多い。
ネットワークモデル以外にfMRIデータから神経活動を推定するモデルがある。これは, 生理学的モデルであるが, 目的に合わせたモデリングによりパラメータを少なくする事が可能か?

大規模なモデル選択

100-1000個のモデル群からモデル選択をしたい。ただし, モデル間の尤度差は小さい。現状では, 複数モデルで構成されるモデルファミリーを定義し, モデルファミリー間比較した後, モデルファミリー内でのモデル選択が行われる。

関連領野の決定

関連領野の同定は, 認知的差分法に依存している。全く依存しない領野同定の方法はあるか? もしくは, 活動している全領野を対象としたネットワーク解析が必要か?

...etc

0034 系列変化点推定解像度向上の試みについて

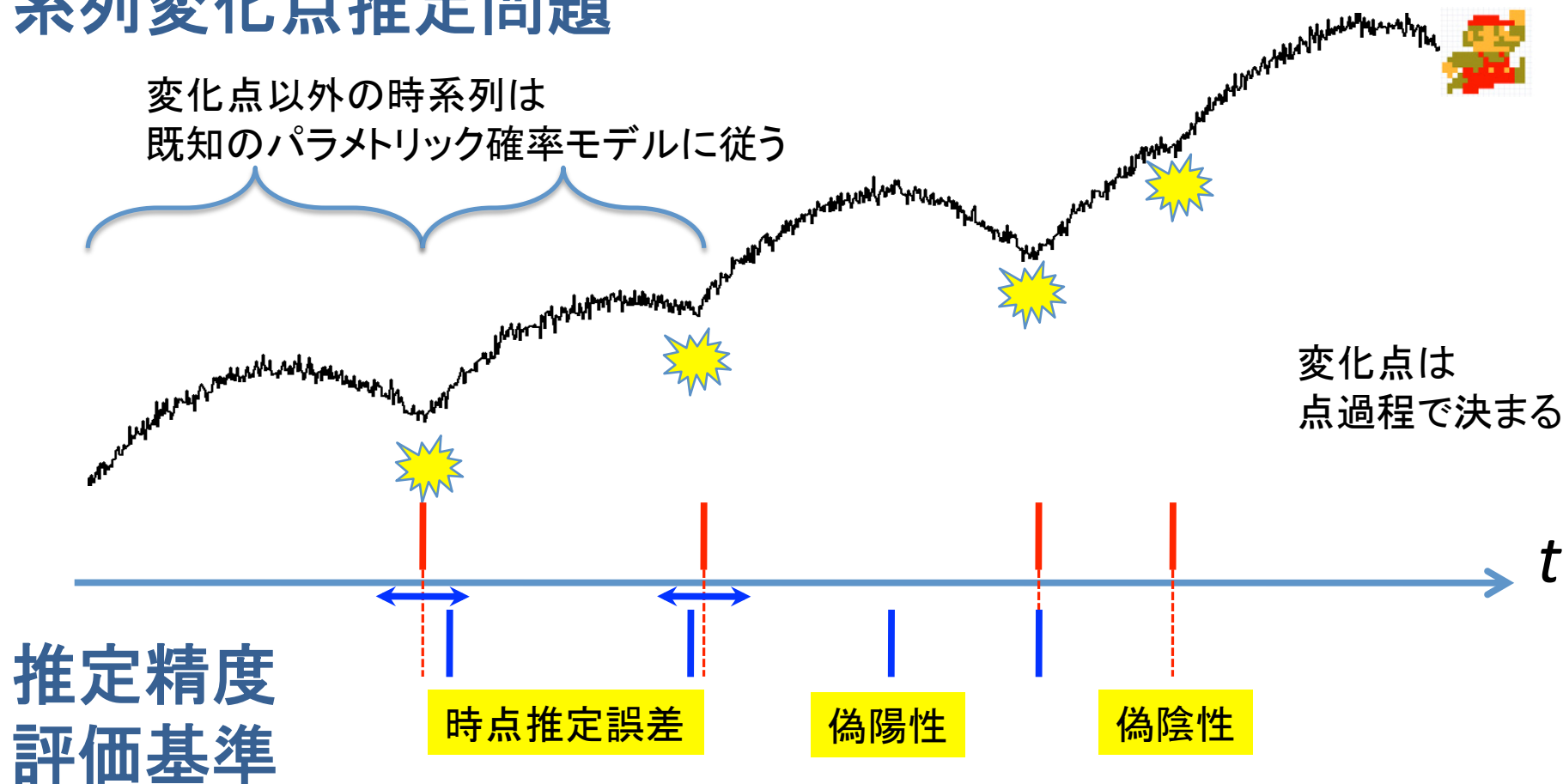
大羽成征(京都大学)



Shigeyuki Oba 大羽成征 @shigepong
早く参加登録しないと発表者でも発表時
もよ。と言われて怖くなり急いで参加登録

系列変化点推定問題

変化点以外の時系列は
既知のパラメトリック確率モデルに従う



どうすれば精度の高い推定が得られる？

基礎となる手法

Perfect simulation 法 [P. Fearnhead, 2006]

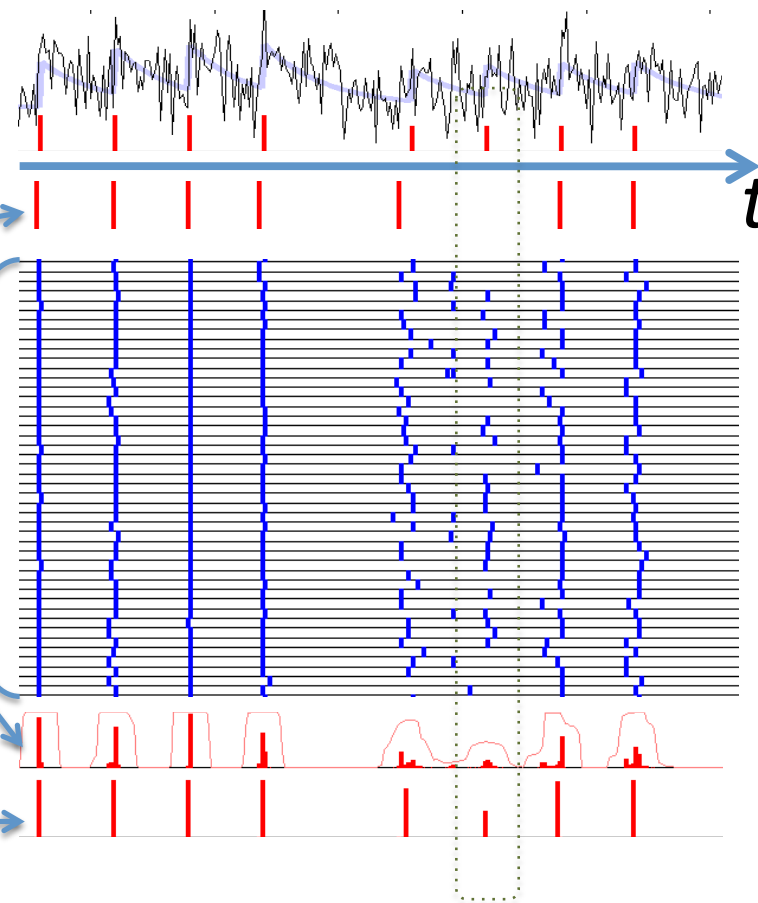
「変化点集合(個数自由)」の事後確率から

- 効率的にMAP推定できる
- 効率的にサンプリングできる
- その和として $P(\text{CP} | t)$ が推定できる

時刻 t が change point である確率

提案手法

- $P(\text{CP} | t)$ をクラスタ化して、
代表点を変化点集合の点推定とする



➡ 大きな観測ノイズのもとで MAP推定以上の「精度」が得られる

D-35

ヒルベルト–シュミット独立基準と ランダム行列理論によるノイズ変数の除去

お茶の水女子大学 川久保秀子、吉田裕亮

研究の目的

目的関数と関係を持つ説明変数の最小部分集合を求めたい。

$\{x_{ij} \mid i=1, \dots, n, j=1, \dots, p\} \sim i.i.d.$

$$\mathbf{X} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix} = [\mathbf{v}_1, \dots, \mathbf{v}_p], \mathbf{Y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

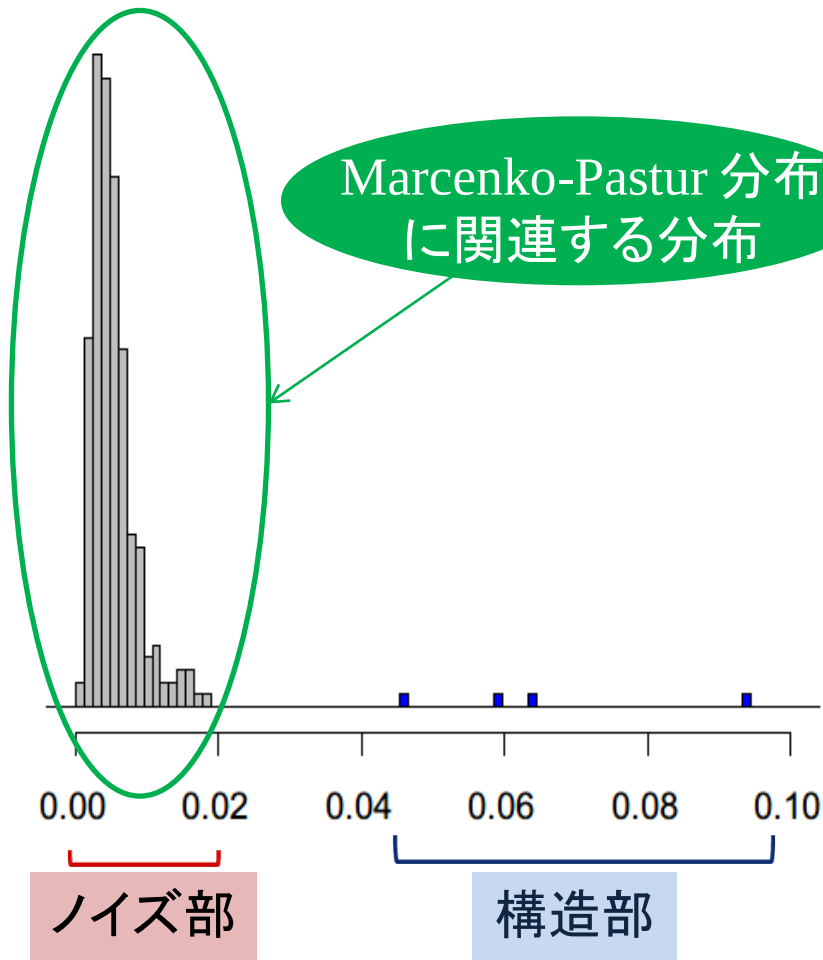
独立同一分布に従う確率変数 $v_1, \dots, v_p$ が Y と非線形な関係を持つとする。

ヒルベルト–シュミット独立基準

v_1	v_2	v_3	...	v_p
0.045	0.063	0.059	...	0.001

v_j と Y が関係を持つかどうか(独立性)をヒルベルト–シュミット独立基準によって計測し、評価値を得る。

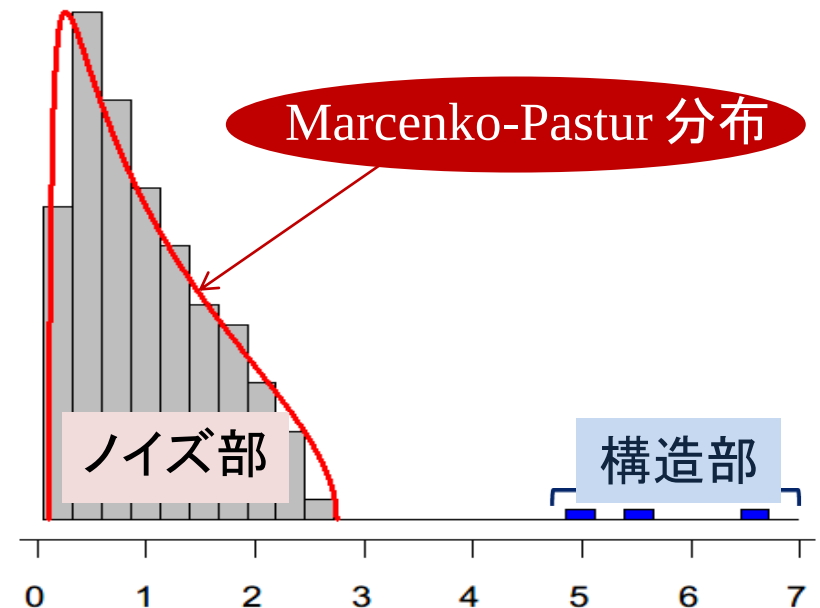
v_j の評価値のヒストグラム



Y と関係を持つ変数の
最小部分集合

Y と独立な変数 v_j の評価値は,
Marcenko-Pastur 分布に関連する分布
に従うことが, 検証から推測された.

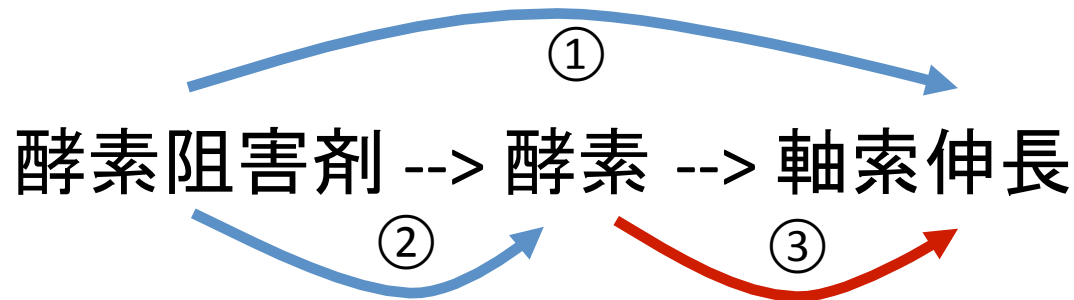
ランダム行列理論では, ランダム性と
関わりのある Marcenko-Pastur 分布を
用い, ノイズ部と構造部の推定を行う.



ランダム行列理論の考え方を応用し,
Marcenko-Pastur 分布に関連した分布
を用いてノイズ部と構造部を推定する.

D-36: 軸索伸長を制御する因子の同定に関する検討

丸野由希, 宮本敦史, 作村諭一, 池田和司



→ 既知
→ 未知

①

軸索伸長度

阻害剤A

130

阻害剤B

80

⋮

100

95

②

酵素1 酵素2 ...

阻害剤A

阻害

阻害

阻害剤B

阻害

阻害

阻害

⋮

阻害

阻害

③

酵素1

?

酵素2

?

酵素3

?

⋮

?

目的: 効率良い軸索伸長をする”酵素阻害の組合せ”を同定

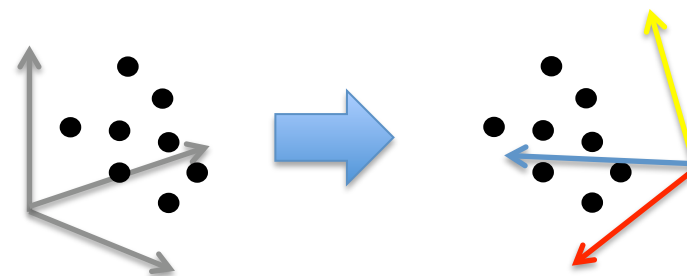
QR分解

==

*	*	*	*
0	*	*	*
0	0	*	*



Difference 2



0
0.7
1.2

0.7

+

1.2

+

CS

CS

Fisherの判別分析(FDA)を使っていて 認識率が低下したことはありませんか？

$$\Sigma_B = \frac{1}{C} \sum_{c=1}^C (\vec{\mu} - \vec{\mu}_c)(\vec{\mu} - \vec{\mu}_c)^\top$$

全てのクラス共分散行列が等しいというFisherの仮定が崩れている際には, FDA で抽出できる(クラス数-1)次元の空間は最適にはなりません.

この時には(クラス数-1)次元の空間の補空間が重要になります.

しかし, 級間分散行列 Σ_B は(クラス数-1)しかランクがありません.

提案手法

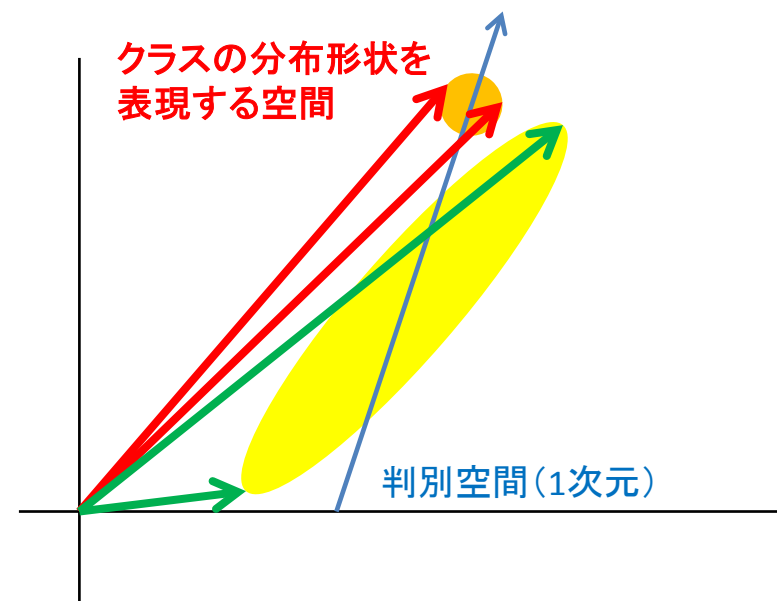
- ・ クラス毎の分布形状を表現するために、
クラス自己相関行列の固有ベクトルを利用します。
- ・ クラスの平均同士だけではなく、
固有ベクトル同士も引き離す様な写像を計算します。
- ・ この方法により、判別空間の外に分布するクラス分布を表現できます。

数値実験

- ・ MNIST, UCI-MLRから選んだ13個のDBのうち
 - ー 全てのDBでFDA後の特徴空間からの
 - ー 7個で次元削減前の特徴空間からの識別率向上を確認しました。
- ・ 計算量は高々FDAの(クラス数+1)倍です。

発表に際して

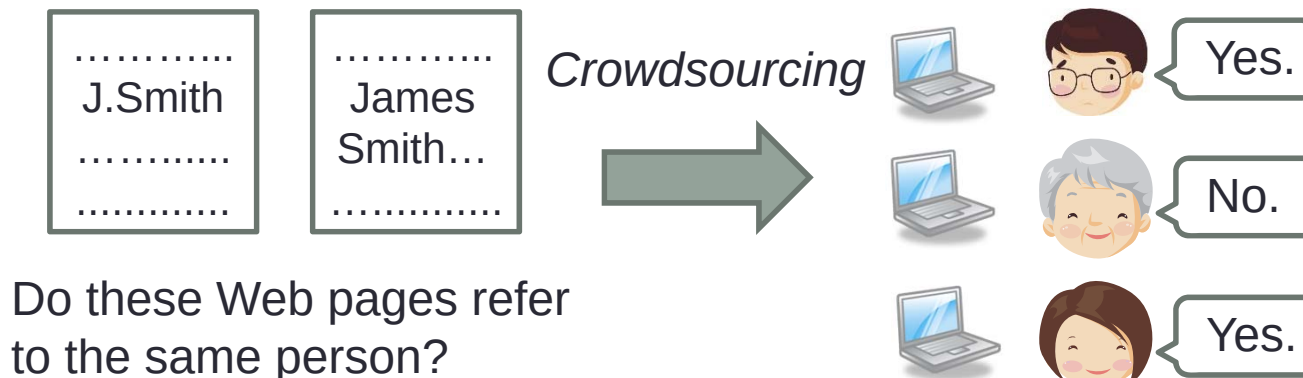
提案法は、強力で計算量が少なく、実装が簡単です。
しかし、現時点では、なぜうまくいくのか、理論的な背景がわかりません。



クラウドソーシングを用いた 同一性判定のための機械学習方式 A Machine Learning Method for Entity Resolution using Crowdsourcing

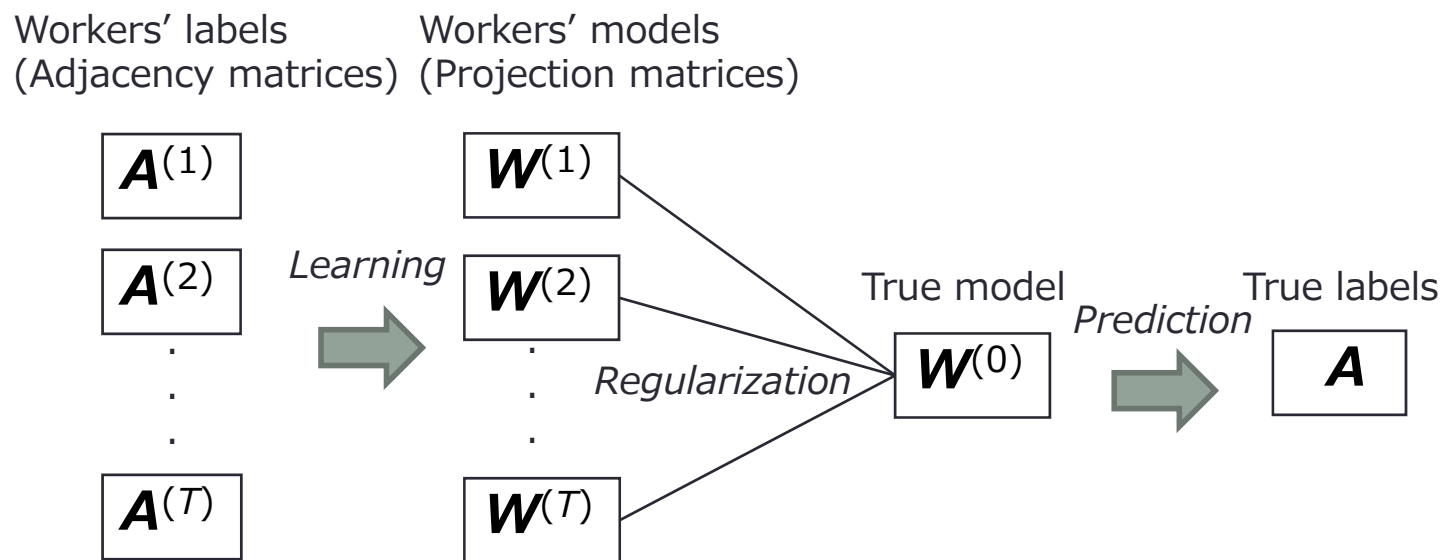
Jingjing Wang, Satoshi Oyama, Masahito Kurihara (Hokkaido University)
Hisashi Kashima (The University of Tokyo)

- ◆ We propose a supervised learning method for entity resolution using labeled data obtained by crowdsourcing.
- ◆ Our method learns an accurate entity resolution model from a set of inaccurate identification results given by crowd workers.



A Machine Learning Method for Entity Resolution using Crowdsourcing

- ◆ We extend Locality Preserving Projections (LPPs) to learn a low-dimensional projection that projects data objects for the same entity close to each other in a latent space.
- ◆ We introduce regularization to simultaneously learn the true projection matrix and workers' projection matrices from the identification labels given by crowd workers.



- ◆ The machine learning problem reduces to a single generalized eigenvalue problem.
- ◆ Once the true model is obtained, we can use it to predict the identity relations among unknown data.

偏相関係数の性質

D-39 大木 仁史

1 相関係数および偏相関係数の定義

ピアソンの積率相関係数 (Pearson's Product Moment Correlation Coefficient) あるいは単に相関係数 (Simple Correlation Coefficient) r は, p 個の多変数 $X_i (i = 1, 2, \dots, k, l, \dots, p)$ における二変数間 X_k と X_l との間の線形関係を表すもので, 次式 (1) で定義され, 相関係数 r の範囲は, $-1 \leq r \leq 1$ である. [1]

$$r_{kl} = \frac{\text{Cov}(\mathbf{X}_k, \mathbf{X}_l)}{\sqrt{\text{Var}(\mathbf{X}_k)} \sqrt{\text{Var}(\mathbf{X}_l)}} \quad -1 \leq r \leq +1 \quad (1)$$

偏相関係数 (Partial Correlation Coefficient) は, 二変数間 X_k, X_l とその他の変数群 $X_q (= X_{1,2,\dots,k-1,l+1,\dots,p})$ の中から 1 個あるいは $(p-2)$ 個の変数までにより制御された, つまりその他の変数群の値を一定に保つことにより影響を除去したときの二変数 X_k と X_l との線形関係を表すものである. その他の変数群 X_q として, 残りのすべてを取らないとき, 高次偏相関係数 (High Order Partial Correlation Coefficient) $r_{kl \cdot q}$ という. 一般に, 偏相関係数というとき, 二変数 X_k と X_l 以外の残りのすべての変数を制御変数とするときの関係を表すことが多い. その場合に限って, 偏相関係数を一つの行列の形とした偏相関係数行列で表すことができる. そして偏相関係数の範囲は, 相関係数と同じ $-1 \leq r_{kl \cdot q} \leq +1$ と定義されている. **しかしながら, 本稿では偏相関係数が 1 以上となる場合があることを示し, 多重共線性との関係を示す.**

2 偏相関係数の導出

今, 三変数 X_1, X_2, X_3 の場合を考える. 変数 X_1 の推定式および変数 X_2 の推定式が, 変数 X_3 によって, 次の線形関係にあり, 各変数は, 平均 0, 分散 1 に標準化されているものとする.

$$X_1 = a_1 + b_1 X_3 + e_1 \quad (2)$$

$$X_2 = a_2 + b_2 X_3 + e_2 \quad (3)$$

$$E[\mu_i] = 0 \quad \text{Var}(X_i^2) = 1 \quad \therefore \text{Var}(X_i^2) = r_{ii} = \sigma_{ii}^2 = 1 \quad (4)$$

このときの e_1 と e_2 との相関係数 $\text{cor}(e_1, e_2)$ は, 変数 X_3 を一定と考えた場合の変数 X_1 と X_2 との間の偏相関係数 $r_{12 \cdot 3}$ である. すなわち, 変数 X_3 の影響を除去するとは, 変数 X_1 を変数 X_3 から推測したときの残差ベクトル $\mathbf{e}_1$ と変数 X_2 を変数 X_3 から推測したときの残差ベクトル $\mathbf{e}_2$ との共分散である残差共分散であり, 変数 X_3 の平均からの平均偏差に対する分散共分散を考えたもので,

$$\text{cor}(e_1, e_2) = r_{12 \cdot 3} = \frac{\text{Cov}(X_1, X_2 | X_3)}{\sqrt{\text{Var}(X_1 | X_3)} \sqrt{\text{Var}(X_2 | X_3)}} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}} \quad (5)$$

と書くことができる. なぜならば, X_1 と X_2 との共分散 $\text{Cov}(X_1, X_2)$ は, 各変数 $X_i, (i = 1, 2, 3)$ は, 平均 $\mu_i = 0$, 分散 $\text{Var}(X_i^2) = 1$ で標準化されているので,

$$\begin{aligned} \text{Cov}(X_1, X_2 | X_3) &= \frac{1}{n} \sum_{i=1}^n (X_1 - r_{13}X_3)(X_2 - r_{23}X_3) \\ &= \frac{1}{n} \sum (X_1 X_2) - r_{23} \frac{1}{n} \sum X_1 X_3 - r_{13} \frac{1}{n} \sum X_2 X_3 + r_{13}r_{23} \frac{1}{n} \sum X_3^2 \\ &= r_{12} - r_{13}r_{23} \\ \therefore \frac{1}{n} \sum \frac{(X_1 - \mu_1)(X_2 - \mu_2)}{\sqrt{X_1^2} \sqrt{X_2^2}} &= \frac{1}{n} \sum X_1 X_2 = r_{12}, \text{Var}(X_3^2) = \frac{1}{n} \sum (X_3 - \mu_3)^2 = \sigma_3^2 = 1 \end{aligned}$$

である. また, 変数 X_1 の分散 $\text{Var}(X_1^2)$ および変数 x_2 の分散 $\text{Var}(X_2^2)$ は,

$$\begin{aligned} \text{Var}(X_1 | X_3) &= \sigma_{1 \cdot 3}^2 = \frac{1}{n} \sum (X_1 - r_{13}X_3)^2 \\ &= \frac{1}{n} \sum X_1^2 - 2r_{13} \frac{1}{n} \sum X_1 X_3 + r_{13}^2 \frac{1}{n} \sum X_3^2 = 1.0 - r_{13}^2 \\ \text{Var}(X_2 | X_3) &= \sigma_{2 \cdot 3}^2 = \frac{1}{n} \sum (X_2 - r_{23}X_3)^2 = 1.0 - r_{23}^2 \end{aligned}$$

である. よって, (5) 式を得る.

相関係数行列 $R = (r_{ij})$ が与えられれば高次偏相関係数 $r_{12 \cdot 3 \dots p}$ は, 逐次計算が行なえる. 例えば, 第一次偏相関係数 $r_{ij \cdot l}$ は, 相関係数 r_{ij} より求められ, 第二次偏相関係数 $r_{ij \cdot lm}$ は, 第一次偏相関係数 $r_{ij \cdot l}$ より求めることができる. [1]

$$r_{12 \cdot 3 \dots p} = \frac{r_{12 \cdot 3 \dots p-1} - r_{1p \cdot 3 \dots p-1} r_{2p \cdot 3 \dots p-1}}{\sqrt{(1 - r_{1p \cdot 3 \dots p-1}^2)(1 - r_{2p \cdot 3 \dots p-1}^2)}} \quad (6)$$

3 三変数の場合の偏相関係数の例

三つの変数 X_1, X_2, X_3 における相関係数行列 $\mathbf{R}$ を

$$\mathbf{R} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} = \begin{pmatrix} 1.0 & 0.1 & 0.9 \\ 0.1 & 1.0 & 0.6 \\ 0.9 & 0.6 & 1.0 \end{pmatrix} \quad (7)$$

とすると、変数 X_3 の影響を除去したときの変数 X_1 と変数 X_2 との間の偏相関係数 $r_{12.3}$ は、 $r_{12.3} = -1.261787$ である。今、 r_{12}, r_{13}, r_{23} を a, b, c と記し、更に、 $A = a, B = bc, C = b^2 + c^2$ とすると、偏相関係数 $r_{12.3}$ は

$$r_{12.3} = \frac{a - bc}{\sqrt{(1 - b^2)(1 - c^2)}} = \frac{a - bc}{\sqrt{1 - (b^2 + c^2) + b^2c^2}} = \frac{A - B}{\sqrt{1 - C + B^2}} \quad (8)$$

と記述できる。今、偏相関係数 $r_{12.3}$ の絶対値が 1 以上であるときを考えると、

$$(A - B)^2 \geq 1 - C + B^2$$

$$r_{12}^2 + r_{13}^2 + r_{23}^2 - 2r_{12}r_{23}r_{31} \geq 1 \quad (9)$$

となる。ところで、この相関係数行列 $\mathbf{R}$ の行列式 $\det(\mathbf{R})$ は、次のようである。

$$\det(\mathbf{R}) = \begin{vmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{vmatrix} \leq 0 \quad (10)$$

$$1 + 2r_{12}r_{23}r_{31} - r_{12}^2 - r_{23}^2 - r_{13}^2 \leq 0 \quad (11)$$

$$\therefore r_{12}^2 + r_{23}^2 + r_{13}^2 - 2r_{12}r_{23}r_{31} \geq 1 \quad (12)$$

である。この行列が正則であるためには、 $\det(\mathbf{R}) \neq 0$ であり、このとき逆行列 $\mathbf{R}^{-1}$ が存在する。そして、 p 個の変数間に完全なる一次の従属関係があるとき、つまり逆行列 $\mathbf{R}^{-1}$ が存在しないとき、あるいはそれに近いとき、多重共線性 (Multicollinearity) の問題が生じることが知られている。

さて、個体の数 n 、変数の数 p である行列 $\mathbf{A}(= a_{ij}) i = 1, \dots, n; j = 1, \dots, p$ において多重共線性が生じる場合の区分は、次のようである。[2]

1. 正確多重共線性

変数間に完全なる一次従属関係があり、行列のランク $\text{rank}(\mathbf{A})$ が落ちる場合 ($n \leq p$)。

2. 準多重共線性

行列のランク $\text{rank}(\mathbf{A})$ は落ちない場合 ($\text{rank}(\mathbf{A}) = p$) であるが、変数間に極めて近い従属関係にあり、最小固有値が正の値ではあるが、極めて 0 に近い場合。

4 多重共線性が生じるときの偏相関係数と重相関係数の性質

以上のことを考え合わせると、多重共線性が生じるときには、偏相関係数は 1 以上の値をとり、多重共線性が生じないのは偏相関係数が 1 未満のときである。このことは従来、重相関係数 R は、 $0 < R \leq 1$ の範囲の値をとると定義されているが、相関係数行列 $\mathbf{R}$ において多重共線性が生じる場合には重相関係数 R が $R \geq 1$ の値をとることを意味する。そして多重共線性は、必ず三変数の問題となる高次偏相関係数 r_{klq} の関係から推測できるといえる ((6) 式参照)。

正確多重共線性が生じるとき (ランク落ちする場合)

$$r_{12}^2 + r_{23}^2 + r_{31}^2 - 2r_{12}r_{23}r_{31} = 1 \quad (13)$$

準多重共線性が生じるとき (ランク落ちしない場合)

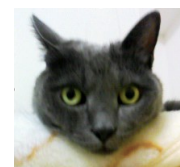
$$r_{12}^2 + r_{23}^2 + r_{31}^2 - 2r_{12}r_{23}r_{31} > 1 \quad (14)$$

多重共線性が生じないとき (正則の場合)

$$r_{12}^2 + r_{23}^2 + r_{31}^2 - 2r_{12}r_{23}r_{31} < 1 \quad (0 < \text{正則の場合} < 1) \quad (15)$$

参考文献

- [1] 塩谷実, 多変量解析概論, 朝倉書店, 1990.
- [2] 竹内啓, 統計学事典, 東洋経済新報社, 1989.



非会員 大木仁史 Ooki Hitoshi

gmacnt2012@gmail.com

無所属。統計に興味あり、R 勉強会@東京に参加。

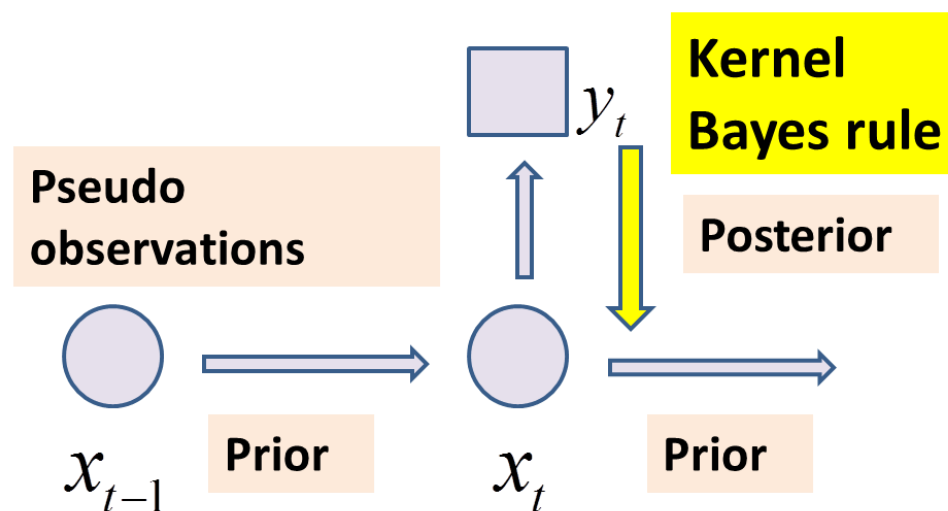
—第 11 回電子情報通信学会情報論的学習理論と機械学習 (IBISML) 研究会第 15 回情報論的学習理論ワークショップ (IBIS2012) ポスターセッション—

D-40 カーネル法によるパーティクルフィルタ

金川元信 (NAIST) 西山悠 (統数研)
Arthur Gretton (UCL) 福水健次 (統数研)

概要

- カーネルベイズ則を用いたパーティクルフィルタを提案します。
- 人工データに対する実験結果を報告します。



提案手法

1. 各時点における事前分布のカーネル平均をサンプリングによって推定.

$$\hat{m}_{X_t|Y_{1:t-1}} = \sum_{i=1}^n w_{t-1}^{(i)} \sum_{j=1}^{l_n} \frac{1}{l_n} k(\cdot, u_t^{(ij)}), \quad u_t^{(ij)} \overset{i.i.d.}{\sim} p(X_t | X_{t-1} = x_{t-1}^{(i)}), \quad i = 1, \dots, n, \quad j = 1, \dots, l_n.$$

2. 遷移関数・尤度関数からのサンプリングによって共分散作用素を推定.

$$x_t^{(i)} \sim p(X_t | X_{t-1} = x_{t-1}^{(i)}), \quad y_t^{(i)} \sim p(Y_t | X_{t-1} = x_{t-1}^{(i)}), \quad i = 1, \dots, n.$$

3. カーネルベイズ則を適用することによって事後分布のカーネル平均を推定.

$$\mathbf{w}_t = \Lambda_t G_{Y_t} ((\Lambda_t G_{Y_t})^2 + \delta_n I_n)^{-1} \Lambda_t \mathbf{k}_Y(y_t), \quad \Lambda_t = \text{diag}(n(G_{X_t} + \varepsilon_n I_n)^{-1} G_{XU} \gamma_t).$$

$$\hat{m}_{X_t|Y_{1:t}} := \sum_{i=1}^n w_t^{(i)} k(\cdot, x_t^{(i)}).$$

表： 観測値が高次元なモデルに対する
適用結果(平均自乗誤差)

n	提案手法	比較手法
100	2.78 ± 0.36	3.43 ± 0.22
500	2.38 ± 0.37	3.59 ± 0.11
1000	2.03 ± 0.23	3.58 ± 0.14
5000		3.57 ± 0.07

提案手法の特徴

- 1. 尤度関数の評価ができなくても適用可能.
⇒ 複雑な観測モデルにも対応.
- 2. 収束レートがデータの次元に非依存.
⇒ 高次元データに対しても高い性能.

モチーフ配列探索

```
seq1: GGGGCGCGATTCCAGGGGGGCGCGGGAGGGG  
seq2: CCGGATGGCACCCCCGGCCGGTGTGCCCGGC  
seq3: GGCCGGTGTGGGGGCGCGGGAATGCCACCCG  
seq4: GGGGGCGCCGGATGCTACCGGCCGGTGCCGG
```

複数の配列中に共通する部分配列を探索する

基本モデル(単一のモチーフを持つモデル)

$$f(S|u, \theta, \theta_0) \quad \theta \sim g(\theta), \theta_0 \sim h(\theta_0)$$

S : 長さ L のDNA配列

ハンサイト(θ_0 から生成)

モチーフ(開始位置は u で, θ から生成)

配列: GGGGCGCGGGGGGCGCGGGAGGGGGGGCGATTCCACGGGA

複数の入力配列を*i. i. d*データとして, パラメータ u, θ, θ_0 を推定

提案モデル(マルチモチーフモデル)

$f(S|u, \theta_1, \dots, \theta_M, \theta_0) \quad (\theta_1, \dots, \theta_M) \sim g(\theta_1, \dots, \theta_M), \theta_0 \sim h(\theta_0)$

タスク分割(事前分布に反発作用を組み込む)

$g(\theta_1, \dots, \theta_M) = \exp\left(w \sum_{j=1}^M \sum_{p=1}^M \|\theta_j - \theta_p\|_2\right) \quad w: \text{反発係数}$

探索手法(ギブス・サンプリング)

入力
配列

```
1 >seq_0↓  
2 CATTTAAGTCTCAACCCATTGTGAGTTGATTTTACAGATGGTGAAGGAAGGGGTCCAG  
3 >seq_1↓  
4 ATGTGTGTTTTCCGTTTCCCTCCCTTTTATGAGTGAGAAAAAAGCGCCTAAATTCCTC  
5 >seq_2↓  
6 GTCATCACTGTATACCTACCTACCCACCCCTCCCCAGCTCAGTGCCTGGCTCACAGTAG  
7 >seq_3↓  
8 TCTTAGGCAATTCCTGTCCAATCACAGATGGTCACATGCTGCTTTCCTGAGTTAACCTAT  
9 >seq_4↓  
10 ATTTGCAATCCTTAAAGCTGAATTGTGCAGTGCATCGGATTTGGAAGCTACTATATCACT
```



ギブスサンプリング
目標分布: $\pi(U, \theta_1, \dots, \theta_M, \theta_0 | S)$

CTGGAG

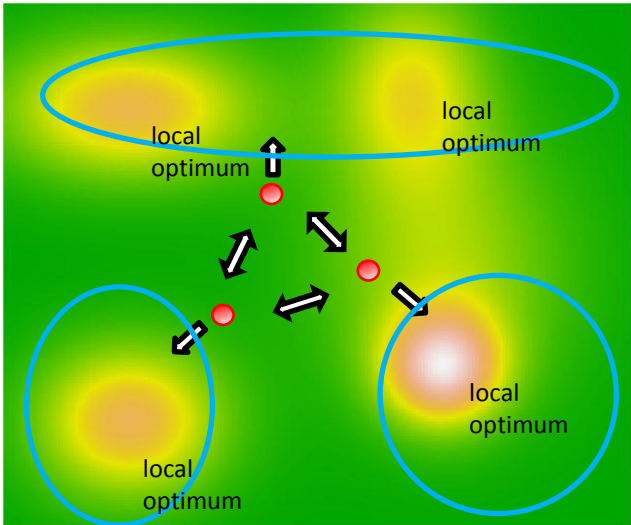
AGTGcCAG

cTCTcTgG

cTGCtTcCA

得られたモチーフ配列

探索空間イメージ



反発させることで推定範囲を分割する

複数のネットワーク構造情報に基づく ノードラベルの半教師あり学習

志賀 元紀 (豊橋技術科学大学)

馬見塚 拓 (京都大学)

➤ 様々な情報源

ネットワーク や カーネル行列 (類似度行列) で与えられる。

例) Webページ解析、ゲノム情報解析、学術文献解析

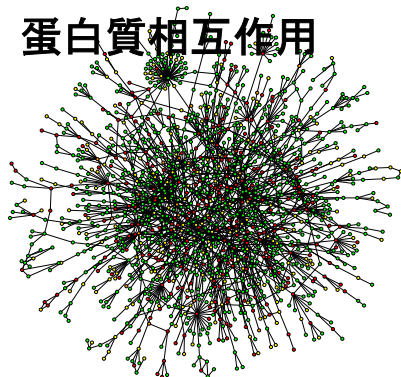
➤ 多様な情報源を上手く統合する

➤ 半教師あり学習

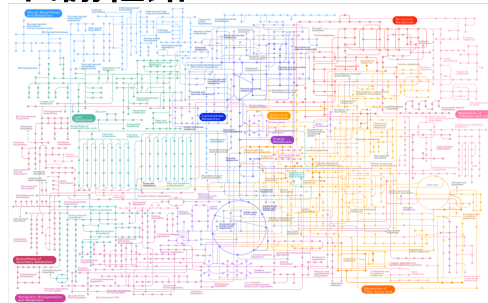
教師ラベル付きデータ と ラベル無しデータ

例) ゲノム情報解析に関するネットワーク

蛋白質相互作用



代謝経路

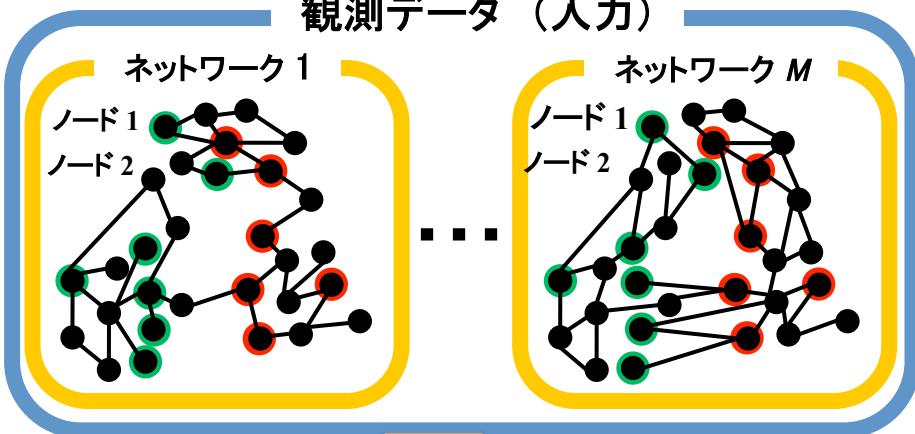


..., etc.

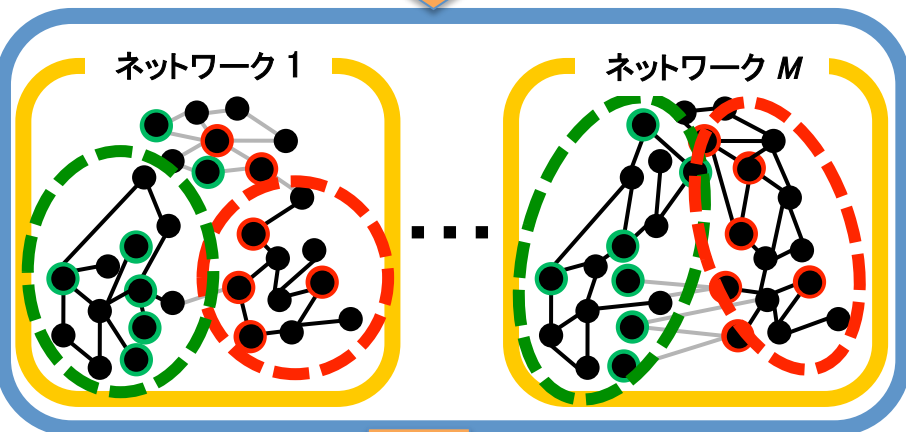
D-43 複数のネットワーク構造情報に基づくノードラベルの半教師あり学習

志賀 元紀(豊橋技術科学大学) 馬見塚 拓(京都大学)

観測データ (入力)



正確な予測に寄与できる
部分構造を取り出す

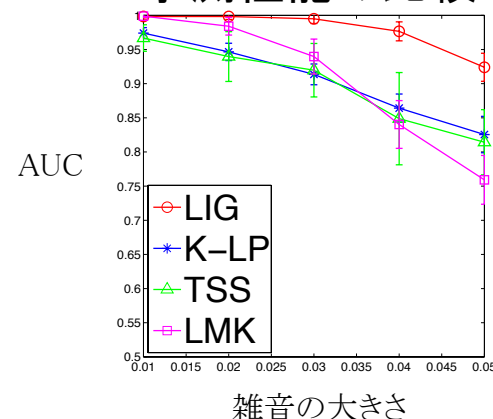


部分構造の組み合わせより、予測器を構築する。

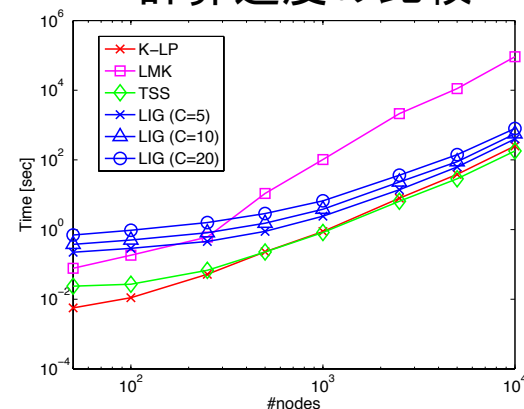
提案法(LIG)

重要な部分構造取り出し、
より正確な予測器を学習する。

予測性能の比較



計算速度の比較



詳細は、ポスター討論 と 以下の論文を参照してください！

“Efficient Semi-Supervised Learning on Locally Informative Multiple Graphs,” M. Shiga and H. Mamitsuka, *Pattern Recognition*, 45(3), 1035-1049, 2012.

D-44

ポイントマス法を用いたガウス和フィルタの改良

福永修一, 中島優次

東京都立産業技術高等専門学校

非線形・非ガウス状態空間モデルの状態推定問題に対して
ガウス和ポイントマスフィルタを提案

└ ガウス和フィルタ + **ポイントマス法** (Simandl et al. 2006)



状態の事後分布を混合
ガウス分布で近似



- ・確率分布をグリッドに分割して求める
- ・粒子フィルタよりも推定精度が高い

ガウス分布の平均と共分散行列を
ポイントマス法を用いて計算

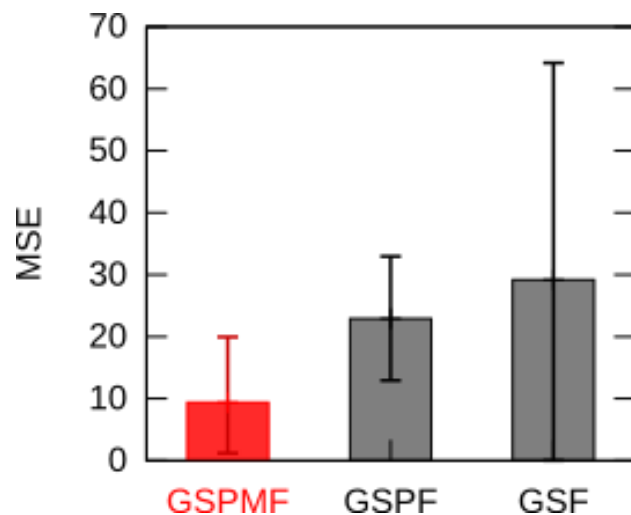
※従来手法では、拡張カルマンフィルタや
粒子フィルタが用いられている

D-44: ポイントマス法を用いたガウス和フィルタの改良

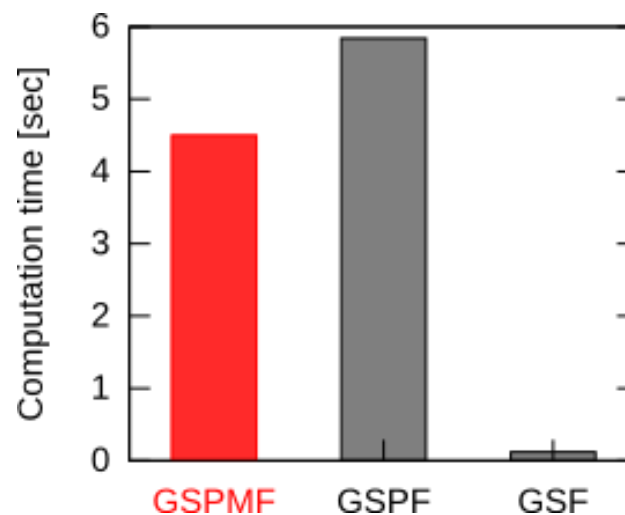
福永修一, 中島優次(東京都立産業技術高等専門学校)

シミュレーション結果

提案手法は従来手法よりも**推定精度が高く計算時間が短い**



平均二乗誤差の比較



計算時間の比較

ガウス和ポイントマスフィルタ(GSPMF)提案手法
ガウス和粒子フィルタ(GSPF)
ガウス和フィルタ(GF)

D45

局所変分法を用いたTotal Variation の 画像修復への応用

庄野 逸 

電気通信大学 大学院情報理工学研究科

岡田 真人  東京大学
THE UNIVERSITY OF TOKYO

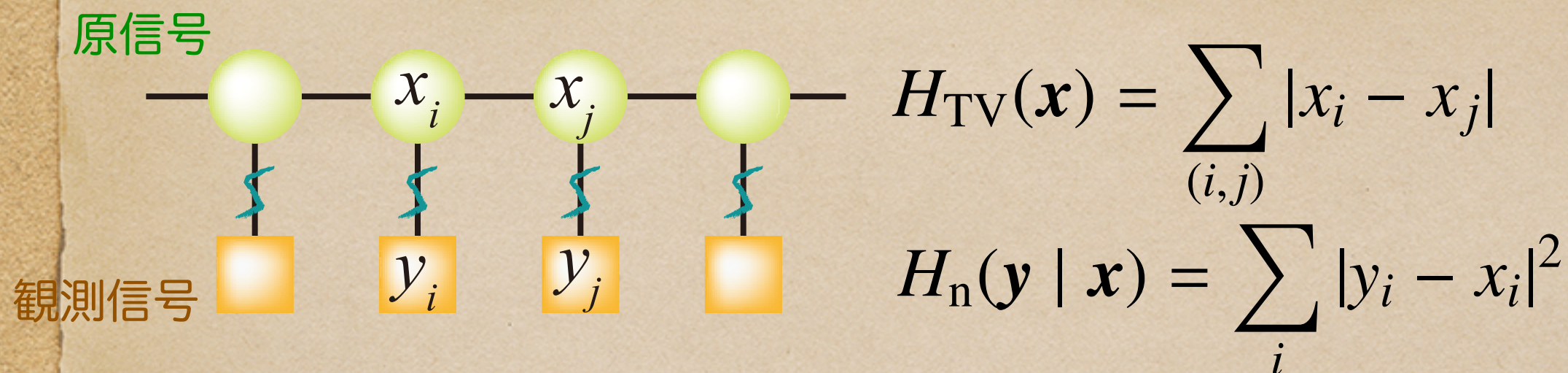
東京大学 大学院新領域創成科学研究科,
理研 BSI

- ◆ Total Variation (Rudin 1992) (TV)とは？

観測信号 y に含まれるノイズを除去の為の拘束条件

- ◆ 現在のところの研究目的
 - ◆ Bayes アプローチによる解釈と近似解の導出

D45: 局所変分法を用いたTotal Variation の画像修復への応用



最適化問題 $\min_{\mathbf{x}} \beta H_n(\mathbf{y} | \mathbf{x}) + \alpha H_{\text{TV}}(\mathbf{x})$

- TV拘束条件 → Laplace 分布 → 局所変分法で近似

$$p_{\text{TV}}(\mathbf{x}) = \prod_{(i,j)} \frac{\alpha}{2} \exp(-\alpha |x_i - x_j|)$$

$$> \prod_{(i,j)} \frac{\alpha}{2} \exp\left(-\frac{\xi_{ij}}{2} (x_i - x_j)^2 - \frac{\alpha^2}{2\xi_{ij}}\right) \rightarrow \text{ガウス関数}$$

$\{\alpha, \beta, \xi\}$ は EM アルゴリズムで決定

パラメトリックカーネル平均を用いた 状態空間フィルタリングアルゴリズム

西山悠¹, 金川元信², Arthur Gretton³, 福水健次¹

1. 統数研 2. 奈良先端大, 3. UCL

背景

- カーネル平均(kernel mean)を使ったノンパラメトリック推論の研究が行われている(e.g., Kernel HMM, Kernel Bayes' rule, Kernel MDP, Kernel POMDP). 確率モデルを仮定せずに, サンプルによるカーネル法の枠組みで推論ができる. 特に, 確率変数間の関係が複雑な場合や事前知識がない場合に有効である.

目的

- 状態遷移モデルを非線形関数+加法的ガウスノイズの既知のモデルとし, 観測モデルを訓練サンプルからノンパラメトリックに推論する場合のフィルタリングアルゴリズムを提案する.
- 正定値カーネルによるノンパラメトリック推論にパラメトリックモデルを組み合わせるための枠組みを導入する.

提案手法

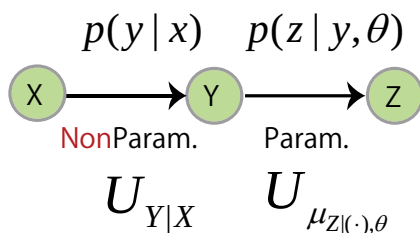
$p(y|x, \theta)$ のカーネル平均

$$\mu_{Y|x, \theta} = E_{Y|X, \theta} [k_Y(\cdot, Y) | X = x] \quad \forall x$$

が解析的に求まるクラスを対象とする。

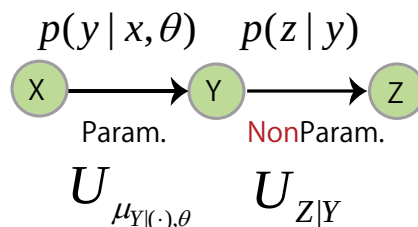
解析的なカーネル平均 $\{\mu_{Y|x, \theta}\}_{\forall x}$ を用いて, [Le Song, et al., ICML2009] の条件付き埋め込み作用素 $U_{Y|X}$ に対応するパラメトリックな場合の作用素 $U_{\mu_{Y|(\cdot), \theta}}$ を導入する。

例:(周辺化操作)



$$\mu_{Z|\theta} = U_{\mu_{Z|(\cdot), \theta}} U_{Y|X} \mu_X$$

$\mu_{Z|\theta}$ は解析的カーネル平均 $\mu_{Y|x, \theta}$ の線形結合で書かれる。

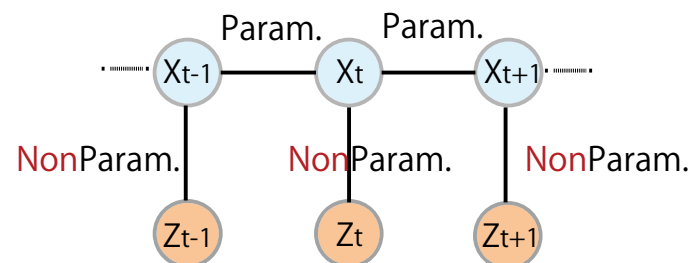


$$\mu_{Z|\theta} = U_{Z|Y} U_{\mu_{Y|(\cdot), \theta}} \mu_X$$

$\mu_{Z|\theta}$ は特徴量 $k(x, \cdot)$ の線形結合で書かれる。パラメータは重みに含まれる。

$U_{Y|X}$ と $U_{\mu_{Y|(\cdot), \theta}}$ を自由に組み合わせることができる。

状態空間モデルへの適用



- ・状態遷移モデル(既知): $x_{t+1} = f(x_t) + GaussNoise$
- ・観測モデル(未知): ノンパラメトリック推論

